

THÈSE de DOCTORAT



de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par Monsieur Ronan PONS

Le 14/10/2024

TITRE : La preuve en droit des systèmes d'intelligence artificielle

École doctorale de droit et science politique - Spécialité Droit

Unité de recherche : Institut de droit privé

Thèse réalisée en cotutelle : Faculté de droit de l'Université d'Ottawa

Thèse dirigée par Céline CASTETS-RENARD et Jessica EYNARD

Composition du jury

Céline CASTETS-RENARD, professeure à l'Université d'Ottawa, directrice de thèse

Jessica EYNARD, maîtresse de conférences à l'Université Toulouse Capitole, habilitée à diriger des recherches, directrice de thèse

Vincent GAUTRAIS, professeur à l'Université de Montréal, examinateur

Céline MANGEMATIN, professeure à l'Université Toulouse Capitole, examinatrice

Juliette SENECHAL, professeure à l'Université de Lille, rapportrice

Étienne VERGES, professeur à l'Université Grenoble Alpes, rapporteur

Nathalie DE GROVE-VALDEYRON, professeure à l'Université Toulouse Capitole, présidente du jury



*« L'université n'entend ni approuver ni désapprouver les
opinions particulières de l'auteur. »*

*A toutes les personnes qui m'ont accompagné dans cette aventure,
celles qui étaient là depuis le départ,
celles qui ont été rencontrées en chemin.*

A ma famille et mes proches.

REMERCIEMENTS

Un doctorat, dit-on généralement, est un marathon : long, difficile et solitaire. Cinq années de thèse plus tard, je peux confirmer sans aucun doute qu'il s'agit d'une expérience longue et ardue. Cependant, il serait mentir que de dire que ma thèse a été un travail solitaire. J'ai eu la chance de bénéficier à chaque instant d'un immense soutien sans lequel cette thèse ne serait pas arrivée à son terme. Par ces quelques lignes, je souhaite remercier tous ceux qui m'ont permis de débiter, persévéer et mettre un point final à cette aventure.

Les deux premières personnes vers lesquelles se tournent mes remerciements sont mes directrices de thèse Céline Castets-Renard et Jessica Eynard. Vous m'avez accordé votre confiance en acceptant de m'encadrer et vous l'avez renouvelé tout au long de ces cinq années. Pour le temps consacré, les conseils prodigués et les relectures réalisées, merci.

Je remercie également les membres de mon jury de thèse, Mesdames et Messieurs les professeur.e.s Juliette Sénéchal, Étienne Vergès, Vincent Gautrais, Céline Mangematin et Nathalie De Grove-Valdeyron pour votre expertise et les échanges que nous avons eu lors de la soutenance.

Je souhaite remercier l'Institut de droit privé de l'Université Toulouse Capitole et le Centre de recherche en droit, technologie de société de l'Université d'Ottawa de m'avoir accueilli et formé au cours de cette expérience enrichissante qu'est la thèse en cotutelle.

Je tiens à remercier l'Institut de recherche en informatique de Toulouse, en particulier les enseignants-chercheurs travaillant à la Manufacture des tabacs. Je me suis toujours senti le bienvenu dans le bureau des doctorants tout au long de ces années. Vous avez contribué à mon amour pour la recherche interdisciplinaire et vous avez participé à ma formation en tant que scientifique. De la même manière, je remercie Laurent et Clément, de l'institut de mathématiques de Toulouse. Merci pour ces longues heures d'échanges qui ont largement dépassé le seul cadre de ma thèse.

Cette thèse a pu être terminée grâce au soutien quotidien des collègues doctorants qui sont devenus au fil du temps des amis : Bastien, Chen Hui, Clément B., Clément C., Colin, Élodie,

Emmanuel, Hao Miao, Jordan, Landy, Robin, Théo, Yan, Yu Zhao et Vincent-Nam. Merci pour toutes nos discussions, soirées et pour cette solidarité dans les moments difficiles.

Puisqu'il est question d'amitié, je me dois de remercier Pierre Popoli et Romain Moreau. Nos études et métiers nous ont éloigné mais je suis fier de toujours pouvoir vous compter parmi mes amis. Maintenant que la thèse se termine, j'ai hâte de pouvoir vous rendre visite.

Je tiens à remercier particulièrement Adrien Dorise. Cher ami et associé, merci de m'avoir embarqué dans tous tes projets dont certains sont désormais aussi les miens. De l'entretien d'embauche à la soutenance, tu as été présent et tu m'as montré que la vie ne se résumait pas à la thèse. Les moments passés ensemble comptent parmi mes plus beaux souvenirs de ce doctorat.

Et finalement, merci à ma famille. En particulier mes parents. Papa, Maman, merci pour avoir écouté et m'avoir partagé toutes les émissions qui parlaient d'intelligence artificielle. Je sais à quel point le doctorat peut sembler flou vu de l'extérieur, mais vous n'avez jamais cessé de m'écouter et de me soutenir durant toutes ces années. Merci pour tout.

Enfin, merci à Marianne pour ton amour et le bonheur que tu continues à m'apporter chaque jour. Tu as été présente dans les moments de joie comme les moments difficiles et tu as toujours été ma plus fidèle supportrice.

SOMMAIRE

INTRODUCTION.....	1
I. L'explication de l'IA.....	4
II. Les enjeux normatifs de l'encadrement des systèmes d'IA.....	11
III. La preuve juridique comme outil d'analyse du droit des systèmes d'IA	17
PARTIE 1 : LES PREUVES RELATIVES AUX EXIGENCES DE CONCEPTION DES SYSTEMES D'IA.....	29
Titre I : L'apparition d'un droit spécifique aux systèmes d'IA.....	31
Chapitre 1 : Les normes applicables aux technologies d'IA	33
Chapitre 2 : L'adoption de normes juridiques spécifiques aux systèmes d'IA	97
Titre II : La place de la preuve dans les exigences de conception d'un système d'IA	157
Chapitre 1 : L'élaboration des preuves de conception	159
Chapitre 2 : Le contrôle des preuves de conception.....	223
PARTIE 2 : LES CONSEQUENCES DES PREUVES DE CONCEPTION LORS DE L'UTILISATION DU SYSTEME D'IA.....	287
Titre I : La preuve réutilisée contre un système d'IA : la responsabilité ex post du fait de l'IA	289
Chapitre 1 : L'articulation entre le règlement européen sur l'IA et la responsabilité ex post du fournisseur de systèmes d'IA.....	291
Chapitre 2 : La réutilisation des preuves de conception lors d'un dommage causé par un système d'IA	331
Titre II : La preuve produite par un système d'IA	375
Chapitre 1 : L'admissibilité de la preuve produite par un système d'IA.....	377
Chapitre 2 : L'appréciation des preuves produites par un système d'IA	427
CONCLUSION GENERALE	495

LISTE DES ABREVIATIONS

Abréviation	Terme développé
A.	Arrêté
Aff.	Affaire
Ass. Plén.	Assemblée plénière de la Cour de cassation
c.	Contre
CA	Cour d'appel
CAA	Cour administrative d'appel
CADA	Commission d'accès aux documents administratifs
C. civ.	Code civil
C. cons	Code la consommation
CE	Conseil d'État
C. éduc	Code de l'éducation
CEDH	Cour européenne des droits de l'Homme
CEN	Comité européen de normalisation
CENELEC	Comité européen de normalisation en électronique et en électrotechnique
CEPD	Comité européen à la protection des données
Ch. Mixte	Chambre mixte de la Cour de cassation
Civ.	Chambre civile de la Cour de cassation
CJCE	Cour de justice des Communautés européennes
CJUE	Cour de justice de l'Union européenne
CNIL	Commission nationale informatique et libertés
Com.	Chambre commerciale de la Cour de cassation
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
CPC	Code de procédure civile
CSP	Code la santé publique
CRPA	Code des relations entre le public et l'administration
DGFIP	Direction générale des Finances publiques
D.C.	District Center
DSA	Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques)
ETSI	Institut européen des normes de télécommunications
G29	Groupe de travail Article 29 sur la protection des données
JO	Journal Officiel
LGDJ	Librairie générale de droit et de jurisprudence
LIL	Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés
n°	Numéro
op.cit	<i>opus citato</i> (dans l'ouvrage cité)
P.	Page
PUF	Presses Universitaires de France
RGPD	Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre

TAJ
UE

circulation de ces données, et abrogeant la directive 95/46/CE
(règlement général sur la protection des données)
Traitement des antécédents judiciaires
Union européenne

INTRODUCTION

« *Models are just opinions embedded in mathematics* » - Cathy O'Neil¹

En 2007, Adrian Fenty, alors maire de Washington D.C., décida de s'attaquer aux problèmes de performance des élèves dans certaines écoles de la ville². La responsabilité des mauvaises notes des élèves fut attribuée aux enseignants. Nommée en tant que rectrice de Washington pour gérer ce problème, Michelle Rhee décida de concevoir un algorithme d'évaluation des professeurs, appelé IMPACT. En 2009, IMPACT fut mis en place et les professeurs furent alors évalués. À la fin de l'année scolaire 2010-2011, les professeurs qui obtinrent les moins bonnes évaluations par le logiciel furent renvoyés. Au total, 206 professeurs perdirent leur emploi³.

En seulement deux ans, les mauvais éléments furent écartés et le système scolaire de Washington D.C. fut optimisé. Malgré des années de protestation afin d'obtenir des explications sur le fonctionnement d'IMPACT, le modèle utilisé pour évaluer les enseignants n'a toujours pas été révélé. Aujourd'hui, l'hypothèse la plus probable est que le niveau des élèves était augmenté artificiellement par les précédents professeurs lors des tests que les élèves devaient passer. En effet, le *Washington Post* et *USA Today* publièrent les résultats d'une enquête révélant qu'un grand nombre de réponses des tests d'évaluation des élèves présentaient de nombreuses ratures, laissant présager une triche importante. Par peur d'obtenir un mauvais score, certains professeurs auraient donc cédé à la pression. Les données sur lesquelles s'appuyaient le logiciel IMPACT étaient donc très probablement erronées. En dépit de cette révélation, aucune contestation ne fut possible pour les professeurs licenciés.

La détermination de la fin et des moyens. Outre les conséquences dramatiques pour les enseignants, l'affaire du logiciel IMPACT illustre bien le fait que les systèmes algorithmiques représentent avant tout des choix de valeurs imposées à des individus par le biais d'un système informatique. Tout d'abord, la finalité du système informatique utilisé témoigne de la vision de son concepteur. Dans le cas du logiciel IMPACT, attribuer la responsabilité des mauvais

¹ O'NEIL Cathy, *Weapons of Math Destruction*, Crown, 2017.

² L'ensemble de cette affaire est issue de l'ouvrage suivant : *Ibid.* PP. 16-26.

³ Il s'agissait des enseignants ayant obtenu des évaluations les classant dans les 5% les plus faibles de l'ensemble des évalués.

résultats aux professeurs ne relève en rien de l'évidence. Une autre possibilité aurait pu être d'optimiser les budgets de ces écoles afin d'offrir de meilleures conditions de travail aux élèves et aux enseignants. Également, le problème aurait pu être la fatigue trop importante des professeurs du fait d'une charge disproportionnée qu'il aurait été possible d'alléger à l'aide d'un logiciel informatique. Or, une seule de ces solutions a été prise en compte dans le logiciel IMPACT.

En outre, le recours à une logique d'optimisation est aussi un choix fort en termes de valeurs. L'optimisation d'une situation, à l'aide d'un système informatique ou non, s'inscrit dans ce que le philosophe André Gorz appelle une « réforme réformiste »⁴. Une réforme réformiste est une réforme « *qui subordonne ses objectifs aux critères de rationalité et de possibilité d'un système et d'une politique donnés* »⁵. En cela, elle s'oppose à la réforme non-réformiste qui est revendiquée en fonction « *de ce qui doit être rendu possible en fonction des besoins et des exigences humaines* »⁶. Une fois le prisme de résolution du problème décidé, il devient très difficile de le remettre en question. Ainsi, lorsque le système d'IA ne parvient pas à résoudre le problème donné, c'est parce qu'il n'est pas assez performant et non pas parce que la solution choisie n'est pas la bonne⁷. Un système d'IA représente donc à la fois la fin et les moyens choisis initialement pour résoudre un problème⁸.

Le droit de l'IA. Les systèmes d'IA n'ont jamais évolué dans un espace de non-droit. En l'absence de règles spécifiques, le droit commun régit ces outils technologiques et il continuera de s'appliquer après l'entrée en vigueur de telles règles. Depuis plusieurs années, les dispositions et les projets de textes législatifs visant à encadrer spécifiquement les systèmes d'IA se multiplient. Or, le droit matérialise lui aussi des finalités, qui représentent alors des valeurs⁹. Plurielles et variables, leur présence est constante en droit¹⁰. Le droit de l'IA fait donc coexister plusieurs valeurs en son sein.

⁴ GORZ Andre, *Stratégie ouvrière et néocapitalisme*, Editions du seuil, 1964.

⁵ *Ibid.*

⁶ *Ibid.*

⁷ GREEN Ben, "Good" isn't good enough, 2019.

⁸ Voir notamment : MENECEUR Yannick, *L'intelligence artificielle en procès*, Bruylant, 2020. P. 399-403.

⁹ ALBIGÈS Christophe, *Introduction au droit*, Bruylant, 2024. BARBIER Hugo et GHESTIN Jacques, *Introduction générale*, LGDJ. LARROUMET Christian, *Traité de droit civil*, Economica, 2013. MALINVAUD Philippe et BALAT Nicolas, *Introduction à l'étude du droit*, LexisNexis, 2024. MOLFESSIS Nicolas et TERÉ François, *Introduction générale au droit*, Dalloz, 2023.

¹⁰ Selon Jean-Louis Bergel, les finalités du droit peuvent être réduites à trois grandes alternatives : « *la justice ou l'utilité d'une part, l'individualisme ou le collectivisme d'autre part, la justice ou la sécurité enfin.* » - BERGEL Jean-Louis, *Théorie Générale du Droit*, Dalloz, 2012. P. 32.

Toutefois, le sujet n'est pas de supprimer toute valeur associée à l'IA, ou qui pourrait être intégrée dans un système d'IA, car la recherche de neutralité de l'IA est un objectif voué à l'échec. En effet, l'IA et le droit sont empreints des valeurs du système social et économique au sein desquels ils sont produits¹¹. L'intérêt du sujet est davantage dans la compréhension de cette coexistence de valeurs dans le but de comprendre comment le droit peut imposer efficacement les siennes. Pour cela, il faut donc préalablement connaître les valeurs en présence. Évidemment, les finalités poursuivies par le droit de l'IA, en particulier par le récent règlement européen sur l'IA¹², seront étudiées. Toutefois, il sera surtout nécessaire de s'intéresser à l'IA et, par extension, au système d'IA. Bien que l'intérêt porté aux technologies d'IA soit particulièrement fort ces dernières années, il ne s'agit pas du premier âge d'or de l'IA. La compréhension des valeurs de l'IA et des systèmes d'IA ne peut donc se passer de l'étude de son histoire¹³ et des évolutions qui ont façonné l'IA que nous connaissons aujourd'hui **(I)**. Du point de vue juridique, l'existence de valeurs inhérentes à l'IA n'a d'intérêt que parce que les systèmes d'IA peuvent entraîner, ou risquent d'entraîner, des conséquences sur le monde. En d'autres termes, les systèmes d'IA possèdent une certaine force normative. La force normative des systèmes d'IA trouve son origine à la fois dans les statistiques et dans l'informatique. Dès lors, le phénomène de coexistence de valeurs peut s'analyser comme une coexistence de normes. La normativité des systèmes d'IA peut alors converger avec les normes juridiques mais également en diverger¹⁴. L'analyse des dynamiques d'articulation entre les normes techniques, issues des systèmes d'IA, et les normes juridiques devient alors un sujet majeur dans la compréhension du droit de l'IA **(II)**. Pour étudier cette interaction normative, il faut alors parvenir à identifier le contexte au sein duquel elle prend place. En d'autres termes, il s'agit de savoir où et quand la norme juridique rencontre la norme d'origine technique. Dans notre cadre, la norme juridique est matérialisée par les règles juridiques du droit de l'IA. Quant à la norme technique, elle se concrétise au travers du système d'IA (et non de l'IA). Tant que le droit n'est pas rédigé de façon automatisée, les systèmes d'IA n'interagissent pas directement avec la norme juridique. La rencontre entre ces normes de nature différente a lieu lors de la mise en application du droit de l'IA. En effet, la manière dont le débiteur d'une obligation juridique exécute ladite obligation représente le résultat de l'articulation entre la norme juridique et la norme technique. Par exemple, un concepteur d'un système d'IA doit exécuter une obligation

¹¹ WINNER Langdon, « Do Artifacts Have Politics ? », *MIT Press*, 109, 1980.

¹² « Règlement sur l'intelligence artificielle », 2024.

¹³ MENECEUR Yannick, « Ce que nous enseignent les trois âges d'or de l'intelligence artificielle » dans MENECEUR Yannick, *L'intelligence artificielle en procès*, Bruylant, 2020. P. 45-53.

¹⁴ CARBONNIER Jean, *Essais sur les lois*, Dalloz, 1995.

qui va à l'encontre des finalités qu'il souhaite intégrer dans son logiciel. Si l'exigence juridique n'est pas satisfaite (ou mal satisfaite), cela signifie que le concepteur a privilégié l'intégration de ses objectifs dans le système d'IA plutôt que ceux poursuivis par le droit, et inversement. Dans cette situation, la preuve de la réalisation de l'obligation ne sera pas suffisante pour convaincre le juge, ou tout autre destinataire, de la conformité du concepteur au droit. Ainsi, lorsque la preuve juridique utilisée convainc son destinataire, cela signifie en théorie que la norme technique est cohérente avec la norme juridique. La preuve représente donc l'articulation entre la norme juridique et la norme technique. Par conséquent, notre étude du droit de l'IA sera réalisée sous le prisme de l'étude de la preuve juridique **(III)**.

I. L'EXPLICATION DE L'IA

Les prémices de l'IA. Les premières recherches pouvant s'inscrire dans ce qui sera appelé ensuite « l'intelligence artificielle » remontent aux travaux des professeurs Warren McCulloch et Walter Pitts en 1943¹⁵. Dans leur papier, McCulloch et Pitts présentent un réseau de neurones artificiels interagissant entre eux. Chaque neurone possède deux états possibles : marche et arrêt. Un neurone s'active en état « marche » lorsqu'il reçoit une stimulation de suffisamment de neurones voisins¹⁶. Les deux chercheurs démontrent alors que ce réseau de neurones artificiels peut à la fois représenter l'ensemble des connecteurs logiques existants (ET, OU, NON, etc.) et calculer toute fonction mathématique calculable. Suite à cet article, plusieurs chercheurs vont produire des travaux précurseurs dans le futur domaine de l'IA dont le plus célèbre est celui publié en 1950 par Alan Turing et intitulé « *Computing Machinery and Intelligence* »¹⁷. Au sein de cet article, Turing présente le désormais célèbre test de Turing¹⁸

¹⁵ MCCULLOCH Warren S. et PITTS Walter, « A logical calculus of the ideas immanent in nervous activity », *Bulletin of Mathematical Biophysics*, 5, 1943. P. 16.

¹⁶ NORVIG Peter et RUSSELL Stuart, *Intelligence artificielle - Une approche moderne*, PEARSON, 2021.

¹⁷ TURING Alan, « Computing Machinery and Intelligence », *Mind*, LIX, 1950.

¹⁸ Le test de Turing peut être résumé comme suivant : « Un évaluateur humain est chargé de juger une conversation textuelle entre un humain et une machine. L'évaluateur sait que l'un des deux participants est une machine, mais ne sait pas lequel. S'il n'est pas en mesure de discerner l'homme de la machine après 5 minutes de conversation, la machine a passé le test avec succès. » - L Bastien, « Test de Turing - Un test pour mesurer l'intelligence artificielle », sur *INTELLIGENCE-ARTIFICIELLE.COM*, publié le 5 juin 2024.

mais également l'apprentissage automatique¹⁹, l'apprentissage par renforcement²⁰ et même les algorithmes génétiques²¹.

La naissance de l'IA comme discipline scientifique. Au cours de l'été 1956, John McCarthy, alors professeur assistant à l'Université de Dartmouth, organise un séminaire de deux mois à l'occasion duquel il réunit dix chercheurs spécialisés en réseaux de neurones, théorie des automates et en étude de l'intelligence²². Le comité de scientifiques s'organise alors autour d'un objectif à atteindre lors de cet événement :

« L'étude reposera sur la conjecture que chaque aspect de l'apprentissage ou de toute autre facette de l'intelligence peut être décrit en principe si précisément qu'une machine peut être construite pour le simuler. Nous tenterons de proposer des solutions pour que les machines puissent utiliser le langage, former des abstractions et des concepts, résoudre des types de problèmes réservés pour l'instant aux humains, et puissent se perfectionner. »²³

La conférence de Dartmouth représente la naissance de l'intelligence artificielle en tant que discipline scientifique. Inventé par John McCarthy, le terme « intelligence artificielle » est adopté officiellement par les participants²⁴. Au cours de ces deux mois de travail collectif, le groupe de chercheurs ne produit pas les « avancées significatives »²⁵ promises au début de l'été. La proposition la plus notable du séminaire est le programme *Logic Theorist*, élaboré par les chercheurs en informatique, Allen Newell et Herbert Simon, capable de démontrer des

¹⁹ « L'apprentissage automatique (machine learning en anglais) est un champ d'étude de l'intelligence artificielle qui vise à donner aux machines la capacité d' "apprendre" à partir de données, via des modèles mathématiques. Plus précisément, il s'agit du procédé par lequel les informations pertinentes sont tirées d'un ensemble de données d'entraînement » - « Apprentissage automatique », sur Cnil.

²⁰ « L'apprentissage par renforcement est un procédé d'apprentissage automatique consistant, pour un système autonome, à apprendre les actions à réaliser, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps. Le système est plongé au sein d'un environnement, et prend ses décisions en fonction de son état courant. En retour, l'environnement procure une récompense, qui peut être positive ou négative. Au fil des expériences, le système cherche un comportement décisionnel optimal, en ce sens qu'il maximise la somme des récompenses au cours du temps. » - « Apprentissage par renforcement », sur CNIL.

²¹ « L'algorithme génétique est un algorithme d'optimisation, reposant le principe de la sélection naturelle [...]. Entrant dans la catégorie des algorithmes évolutionnistes, il fait appel à des algorithmes générés aléatoirement, dans un univers défini, qui interagissent entre eux et avec l'univers. Ils ont la capacité d'incorporer des informations présentes au sein de l'univers, et donc d'évoluer (de "muter") en fonction de ces inputs. » - « Algorithme génétique », sur Journal du net, publié le 19 juillet 2022.

²² NORVIG Peter et RUSSELL Stuart, *Intelligence artificielle - Une approche moderne*, PEARSON, 2021. P. 16.

²³ MCCARTHY John, MINSKY Marvin, ROCHESTER Nathaniel et al., « A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence », 1955.

²⁴ PICKOVER Clifford A., *La fabuleuse histoire de l'intelligence artificielle : Des automates aux robots humanoïdes*, Dunod, 2021.

²⁵ « Nous pensons que des avancées significatives sont possibles si une équipe judicieusement sélectionnée de scientifiques travaille sur ce sujet pendant un été » - MCCARTHY John, MINSKY Marvin, ROCHESTER Nathaniel et al., op. cit.

théorèmes mathématiques²⁶. En dépit de ces résultats en-deçà des attentes de départ, les participants du séminaire de Dartmouth en 1956 demeurent très confiants sur les capacités prometteuses des technologies²⁷ et poursuivent leurs avancées pendant une décennie.

Les limites des premiers modèles. Au cours de cette période, les systèmes d'IA fonctionnent par la combinaison d'opérations. La logique est alors d'essayer toutes les combinaisons d'opérations possibles jusqu'à en trouver une correspondant à la solution souhaitée. Si cette approche se révèle efficace pour la résolution de problèmes simples, à savoir avec peu de possibilité d'actions possibles pour le système, elle s'avère particulièrement limitée lorsque le problème implique de nombreuses entités. Connu sous le nom d' « explosion combinatoire », ce problème désigne le fait qu'un changement du nombre de données à traiter suffit à rendre la solution très difficile, voire impossible à calculer par un ordinateur. En d'autres termes, l'approche de l'IA à cette époque fonctionne théoriquement mais la puissance de calcul des ordinateurs est largement insuffisante pour l'exécuter en pratique. Cette incapacité à surmonter la problématique d'explosion combinatoire est la critique majeure du rapport rédigé par James Lighthill en 1973²⁸. Particulièrement critique du potentiel de l'IA face aux problèmes réels, l'article contribue significativement à la décision du gouvernement de Grande-Bretagne d'arrêter les subventions au domaine de l'IA. Outre cette limite, les technologies d'IA font également l'objet de critiques plus fondamentales concernant le fonctionnement des réseaux de neurones. Marvin Minsky, père fondateur de l'IA, et Seymour Papert publie en 1969 « *Perceptrons* »²⁹. L'article met en avant les limites des capacités des perceptrons, une forme simpliste de réseaux de neurones, dans la représentation des problèmes. Ajouté au rapport Lighthill, l'article renforce la perte de confiance dans l'IA et la réduction drastique des fonds accordés à la recherche en IA.

L'apparition des systèmes experts. À partir de 1970, une approche différente de l'IA est alors explorée afin de résoudre des problèmes sans pour autant se heurter à l'obstacle de l'explosion combinatoire. Jusqu'alors, les méthodes d'IA ont reposé sur une exploration de combinaisons d'opérations jusqu'à trouver une solution au problème. Pour faire face à cela, plusieurs

²⁶ MCCORDUCK Pamela, *Machines Who Think*, Routledge, 2004.

²⁷ « La manière la plus simple de résumer les faits consiste à dire qu'il existe désormais des machines capables de penser, d'apprendre et de créer. En outre, leur capacité à accomplir des choses va rapidement s'accroître jusqu'à ce que, dans un futur proche, le champ des problèmes qu'elles pourront aborder soit coextensif à celui auquel s'applique l'esprit humain. » - Déclaration de Herbert Simon en 1957. Cité dans NORVIG Peter et RUSSELL Stuart, *Intelligence artificielle - Une approche moderne*, PEARSON, 2021.

²⁸ LIGHTHILL James, « Artificial Intelligence : A General Survey », 1972.

²⁹ MINSKY Marvin et PAPERT Seymour A., *Perceptrons*, The MIT Press, 1969.

chercheurs en informatique³⁰ décident de se reposer davantage sur les connaissances spécifiques des experts du domaine dans lequel s'inscrit le problème à résoudre. De ce fait, les programmes sont appelés des systèmes-experts. Le premier système expert commercial, appelé R1, fait son apparition en 1982. Il a pour objectif de configurer les commandes des systèmes informatiques de l'entreprise Digital Equipment Corporation, alors deuxième plus grande entreprise du secteur. L'utilisation de R1 permet de faire économiser à la société 40 millions de dollars par an³¹. Contrairement aux approches exploratoires des années 60, les systèmes experts parviennent à résoudre des problèmes du monde réel. Ce succès relance l'engouement pour l'IA et, en 1990, toutes les grandes entreprises américaines possédaient un département de recherche dédié aux systèmes experts. Les gouvernements nationaux³² réinvestissent à nouveau massivement afin de garantir la compétitivité de leurs industries dans un secteur de l'IA alors évalué à plusieurs milliards d'euros³³. Cependant, la force des systèmes experts est également leur plus grande limite. Parce qu'ils sont programmés avec des règles précises, ils ne sont pas capables d'apprendre et ne parviennent donc pas à résoudre des problèmes contenant de l'incertitude.

Le retour en force de l'apprentissage automatique. Pour faire face à l'incertitude, une nouvelle approche, le raisonnement probabiliste, est explorée par les scientifiques. Alors que les systèmes experts raisonnaient de façon binaire (un élément est vrai ou faux), cette approche privilégie le recours aux probabilités qui sont plus adaptées à représenter le fonctionnement désordonné du monde réel. Ensuite, là où les systèmes experts étaient composés de règles rigides écrites manuellement, l'approche du raisonnement probabiliste préfère utiliser l'apprentissage automatique. Les modèles d'apprentissage automatique sont dits « connexionnistes » tandis que les systèmes experts sont dits « symboliques ». L'augmentation rapide de la capacité de calcul des ordinateurs et la production massive de données disponibles permet contribuent à améliorer significativement les performances des modèles d'apprentissage automatique³⁴. Enfin, en 2011, l'invention des méthodes d'apprentissage profond a permis aux méthodes d'apprentissage automatique de passer un cap dans de nombreux domaines tels que

³⁰ BUCHANAN Bruce et FEIGENBAUM Edward, « Heuristic Dendral : a Program for Generating Explanatory Hypotheses in Organic Chemistry », *Edinburg University Press*, 1969.

³¹ NORVIG Peter et RUSSELL Stuart, *Intelligence artificielle - Une approche moderne*, PEARSON, 2021.

³² A titre d'exemple, le Japon lance en 1981 le projet « Ordinateurs de cinquième génération » avec un budget de 1,3 milliards de dollars. RIEU Alain-Marc et MORIZET-MAHOUDEAUX Pierre, « Introduction », *Persée - Portail des revues scientifiques en SHS*, 1994.

³³ NORVIG Peter et RUSSELL Stuart, *op. cit.*

³⁴ CARDON Dominique, *A quoi rêvent les algorithmes*, SEUIL, 2015.

la reconnaissance d'image, la reconnaissance de parole ou encore la traduction automatique³⁵. Ces nouvelles prouesses des technologies d'IA ont apporté avec elles un vif intérêt, toujours présent à ce jour, sur cette approche aussi bien de la part des entreprises et des investisseurs que de la part du grand public et des médias.

La conception originelle de Turing. Au fil des soixante-dix années qui ont constitué son histoire, la conception de l'IA a grandement évolué. Le test de Turing représente l'une de ces conceptions de l'IA. Le test de Turing consiste à demander à un humain de poser des questions par écrit et deviner si leurs réponses ont été produites par un ordinateur ou par un individu humain. Si l'interrogateur humain ne parvient pas à identifier les réponses produites par le système informatique, ce dernier réussit le test de Turing. La conception de ce test présente deux caractéristiques essentielles représentatives de la conception de l'IA choisie. Premièrement, le référentiel de test est l'humain. Le système informatique ne doit pas pouvoir être distingué de l'humain pour réussir le test. L'intelligence artificielle est alors définie par sa similarité à l'intelligence humaine. Secondement, le test évalue le fait que les résultats produits par le système informatique, à savoir les réponses aux questions posées, soient indistinguables de ceux produits par un humain. Le test ne se fonde donc pas sur une similarité de raisonnement mais sur une similarité de comportement. En résumé, le test de Turing définit, par sa procédure, le système d'IA comme un système informatique se comportant de façon identique à un être humain³⁶.

La conception dominante de l'agent rationnel. Aujourd'hui, le paradigme dominant dans la conception de l'IA est celui de l'agent rationnel. L'IA est alors une entité qui se comporte de façon optimale afin d'atteindre le meilleur résultat possible. De la même manière que le test de Turing, la conception de l'IA comme modèle rationnel se désintéresse du raisonnement du système informatique, seul le résultat produit importe. En revanche, la comparaison du comportement n'est plus avec l'humain mais avec un idéal de rationalité. Informatiquement, cela peut se traduire par un système d'IA qui a pour objectif de maximiser une fonction de récompense ou minimiser une fonction de coût. En d'autres termes, le paradigme de l'agent rationnel définit le système d'IA comme un système faisant les bons choix³⁷.

³⁵ NORVIG Peter et RUSSELL Stuart, *op. cit.*

³⁶ Pour plus de détails sur la double distinction « raisonnement contre comportement » et « humain contre rationalité », voir *Ibid.*

³⁷ *Ibid.*

Les limites actuelles du paradigme de l'agent rationnel. Bien que dominante, cette conception de l'IA présente certaines limites. En effet, le modèle de l'agent rationnel nécessite de fournir à l'ordinateur un objectif clair et précisément spécifié à atteindre. Dans de nombreuses situations, l'objectif que le concepteur du système d'IA souhaite intégrer est difficile à formuler clairement pour le système informatique tout en restant fidèle à l'objectif de départ. Par exemple, identifier les bons candidats pour un emploi à pourvoir est un objectif qui peut difficilement être résumé de façon simpliste. L'adéquation entre l'objectif souhaité par le concepteur et l'objectif intégré au sein du système informatique est appelé « problème d'alignement des valeurs »³⁸. L'ensemble des règles juridiques applicables aux technologies d'IA, ou droit de l'IA ci-après, s'inscrit exactement dans cette problématique. Les obligations juridiques concernant la conception et/ou l'utilisation des logiciels d'IA imposent la prise en compte de nouveaux objectifs qui viennent s'ajouter à celui du concepteur du système d'IA et à celui intégré dans le système informatique. Par exemple, le système d'IA destiné à supprimer les contenus illicites postés sur une plateforme en ligne devra également respecter la liberté d'expression des utilisateurs. Ainsi, le droit impose des contraintes, ici sous la forme d'une mise en balance d'intérêts, au système d'IA dans la poursuite de son objectif initial. Le problème d'alignement des valeurs doit alors être résolu entre l'objectif intégré dans le système informatique, l'objectif poursuivi par son fournisseur³⁹ et/ou déployeur⁴⁰ et l'objectif imposé par le droit. Il n'est pas possible d'accepter que le logiciel produise des conséquences illégales au seul motif que cela lui permet d'atteindre de façon optimale son objectif.

Le choix de définition du système d'IA. La diversité de conceptions de l'IA entraîne l'existence d'une très grande diversité de propositions de définitions de l'IA, sans qu'aucune d'entre elles ne parviennent à faire consensus. Définir l'IA de façon universelle est donc un défi extrêmement difficile qui n'a toujours pas été résolu et qui ne saurait être répondu ici. Par conséquent, le droit européen a fait le choix d'encadrer non pas l'IA mais les systèmes d'IA. Ainsi, le règlement européen sur l'IA de 2024 définit le système d'IA comme :

« un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son

³⁸ *Ibid.*

³⁹ « une personne physique ou morale, une autorité publique, une agence ou tout autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit ; » - « Règlement sur l'intelligence artificielle », 2024. Article 3.3).

⁴⁰ « une personne physique ou morale, une autorité publique, une agence ou un autre organisme utilisant sous sa propre autorité un système d'IA sauf lorsque ce système est utilisé dans le cadre d'une activité personnelle à caractère non professionnel ; » - « Règlement sur l'intelligence artificielle », 2024. Article 3.4).

déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels »⁴¹

Le choix de définir la notion de système d'IA pour établir le périmètre de notre étude présente plusieurs avantages. Premièrement, le droit européen (mais aussi canadien et états-unien) ne réglemente pas l'IA à proprement parler mais les systèmes d'IA. Dès lors, la coexistence de valeurs entre le droit et la technique apparaît lors de l'encadrement des systèmes d'IA, plus précisément lors de la mise en application des textes juridiques. Par conséquent, la notion de système d'IA se révèle davantage pertinente pour étudier ce phénomène que celle d'IA.

Ensuite, le choix de définir la notion de système d'IA conformément au règlement européen sur l'IA présente un second avantage. La définition du système d'IA choisie par l'Union européenne est partagée par d'autres textes juridiques européens⁴², non européens ainsi que par des institutions internationales⁴³. Par exemple, l'organisation de coopération de développement économiques (OCDE) a adopté une définition quasi-identique à celle du règlement européen :

« un système d'IA est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels. Les systèmes d'IA peuvent présenter différents niveaux d'autonomie et de capacité d'adaptation après leur déploiement. »⁴⁴

Également, le Canada ainsi que les Etats-Unis ont adopté dans leur texte une définition du système d'IA proche de celle de l'Union européenne :

« système technologique qui, de manière autonome ou partiellement autonome, traite des données liées à l'activité humaine par l'utilisation d'algorithmes génétiques, de réseaux neuronaux, d'apprentissage automatique ou d'autres techniques pour générer du contenu, faire des prédictions ou des recommandations ou prendre des décisions. »⁴⁵

« un système automatisé qui peut, pour un ensemble donné d'objectifs définis par l'homme, faire des prédictions, des recommandations ou prendre des

⁴¹ « Règlement sur l'intelligence artificielle », 2024.

⁴² « Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit », 2024.

⁴³ CASTETS-RENARD Céline, « Proposition de règlement sur l'intelligence artificielle (derniers développements) », *Recueil Dalloz*, 2023.

⁴⁴ « Updates to the OECD's definition of an AI system explained ». Traduction réalisée par l'auteur.

⁴⁵ *Loi sur l'intelligence artificielle et les données*, 16 juin 2022.

décisions influençant des environnements réels ou virtuels. Les systèmes d'intelligence artificielle utilisent des données d'entrée d'origine humaine ou non humaine pour percevoir des environnements réels et virtuels; ils transforment ces perceptions en modèles par le biais d'une analyse automatisée et utilisent l'inférence des modèles pour formuler des options d'information ou d'action »⁴⁶

Notre analyse s'inscrit principalement en droit européen et en droit français. Or, le cadre théorique de notre analyse étant défini en grande partie par la notion de système d'IA, il est indispensable que celle-ci soit cohérente avec la définition choisie en droit européen. De plus, la similarité de la définition européenne avec celles élaborées par d'autres législateurs et institutions internationale confère un certain intérêt international à notre étude. En effet, la réutilisation des réflexions et des apports réalisés ci-après sera facilitée au-delà du droit européen en raison de cette proximité théorique de la notion de système d'IA. Par conséquent, pour toutes ces raisons, la notion de système d'IA telle que définie par l'Union européenne se révèle être la plus pertinente pour les besoins de notre analyse des valeurs de l'IA.

II. LES ENJEUX NORMATIFS DE L'ENCADREMENT DES SYSTEMES D'IA

L'absence de neutralité de l'IA. *« La technologie n'est ni bonne, ni mauvaise, ni neutre »⁴⁷* énonçait le professeur d'histoire des sciences Melvin Kranzberg. Par le biais de sa première loi sur la technologie, l'historien des sciences désigne le fait que les technologies ont des conséquences dépassant les effets directs liés à leur utilisation :

« L'interaction de la technologie avec le contexte social est telle que les développements techniques ont souvent des conséquences environnementales, sociales et humaines qui vont bien au-delà des objectifs immédiats des dispositifs et pratiques techniques eux-mêmes, et qu'une même technologie peut avoir des résultats très différents lorsqu'elle est introduite dans des contextes différents ou dans des circonstances différentes. »⁴⁸

⁴⁶ « Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence », 2023. Section 3(e). Traduction par l'auteur.

⁴⁷ « Technology is neither good nor bad; nor is it neutral. » - KRANZBERG Melvin, « Technology and History », *Technology and Culture*, 27, [The Johns Hopkins University Press, Society for the History of Technology], 1986.

⁴⁸ *Ibid.*

L'IA ne fait pas exception à la règle. Comme toute technologie, il est indispensable de ne pas s'arrêter à l'étude de ses conséquences directes⁴⁹. Le développement d'une technologie n'est jamais le produit d'une dynamique interne hermétique à son contexte. L'analyse du système économique et social dans lequel s'inscrit une technologie est aussi important à prendre en compte que l'analyse de la technologie elle-même⁵⁰. Par conséquent, il est indispensable de rappeler le contexte dans lequel s'inscrit le recours contemporain à l'IA afin de comprendre les effets globaux entraînés par l'utilisation des systèmes d'IA et, plus particulièrement, leurs conséquences normatives.

Le contexte d'apparition des approches modernes de l'IA. Actuellement, les approches connexionnistes, à savoir appartenant aux méthodes d'apprentissage automatique (dont l'apprentissage profond), sont particulièrement mises en avant lorsqu'il est question d'IA. Le retour en grâce de ces approches s'explique, d'une part, par l'augmentation de la puissance de calcul des ordinateurs et, d'autre part, par l'accès à des quantités phénoménales de données. Ce second point se révèle particulièrement intéressant car l'accès à d'importants jeux de données nécessite au préalable que ces jeux de données existent. La mise en données croissante de la société est un phénomène central dans la compréhension du contexte dans lequel s'inscrit le recours à l'IA⁵¹.

Pendant longtemps, la mesure et le calcul sont restées une affaire de spécialistes⁵². D'un côté, les scientifiques les utilisaient comme des outils au service de la connaissance. Ainsi, le biologiste place des bracelets GPS sur les pattes d'un oiseau afin de découvrir sa trajectoire migratoire et, ensuite, de la comprendre. De l'autre côté, pour les États et les entreprises, ces indicateurs sont parfois utilisés de façon simplifiés pour justifier la prise de mesures. À partir des années 1980, les politiques néolibérales menées conduisent à une généralisation du recours aux indicateurs⁵³. Les classements, des universités⁵⁴ ou des villes⁵⁵, et les notations, de travailleurs ou encore de restaurants, se multiplient. Les indicateurs se sont alors répandus dans

⁴⁹ MENECEUR Yannick, « Les trois grands défis posés par la gouvernance de l'intelligence artificielle et de la transformation numérique », *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale*, 23, Éditions Nota bene, 2021.

⁵⁰ WINNER Langdon, « Do Artifacts Have Politics ? », *MIT Press*, 109, 1980.

⁵¹ Pour Antoinette Rouvroy, la récolte massive de données est le premier « temps » de la gouvernementalité algorithmique - ROUVROY Antoinette et BERNIS Thomas, « Gouvernementalité algorithmique et perspectives d'émancipation », *Rezeaux*, n° 177, La Découverte, 2013. P. 168.

⁵² CARDON Dominique, *A quoi rêvent les algorithmes*, SEUIL, 2015.

⁵³ *Ibid.*

⁵⁴ « ShanghaiRanking's Academic Ranking of World Universities ».

⁵⁵ « TOP 10 des villes où il fait bon vivre en France », sur *Paradisimmio* « Classement des meilleures villes étudiantes 2023-2024 - L'Etudiant ».

notre quotidien afin de quantifier toujours plus de nouvelles activités. Les données récoltées et les calculs servent de moins en moins à comprendre les comportements et de plus en plus à les conduire. La statistique n'est plus seulement une information, elle est aussi une référence, un objectif à atteindre ou à maintenir.

Pour reprendre l'expression de Dominique Cardon, on entre alors dans une société de calcul⁵⁶. Cette transformation passe, dans un premier temps, par la mise en donnée massive de la société. La quantité de données générée par les activités humaines explose. En 2024, 29.000 gigaoctets d'informations sont publiées dans le monde par seconde⁵⁷. Une fois la phase de collecte effectuée, la seconde étape est de valoriser ces données au travers d'un traitement. Au regard du volume de données dont il est question, l'analyse humaine seule paraît difficilement envisageable. Pour traiter ces données, il est décidé d'utiliser des procédés techniques : les algorithmes informatiques. Il s'agit d'un ensemble d'instructions qui, une fois intégré dans un logiciel, permet de classer, séparer, visualiser automatiquement les données⁵⁸. Grâce au recours aux algorithmes informatiques, la donnée devient une information, une connaissance voire un savoir⁵⁹. C'est dans ce contexte de mutation sociale que les technologies d'IA révèlent tout leur potentiel. En effet, leur efficacité s'explique autant par les prouesses mathématiques et informatiques des chercheurs que par le milieu propice dans lequel elles sont employées⁶⁰. Les systèmes informatiques employant des technologies d'IA contribuent à parfaire cette société du calcul. Ils en sont autant une conséquence qu'une cause.

Les enjeux normatifs liés au fonctionnement de l'IA. Si elle est une étape indispensable pour comprendre notre sujet d'étude, la présentation du contexte actuel de l'IA ne suffit pas à saisir l'entière des enjeux, particulièrement normatifs, de ces technologies. À l'image du contexte dans lequel est utilisée une technologie, le fonctionnement de cette dernière a un effet sur les conséquences indirectes entraînées par son usage. Pour reprendre les propos du professeur Langdon Winner, certaines technologies présentent de façon inhérente un caractère politique de par leur fonctionnement⁶¹. Ainsi, de nombreuses technologies de l'information appartiennent à cette catégorie⁶². Par exemple⁶³, le professeur Lawrence Lessig a mis en avant la force

⁵⁶ CARDON Dominique, *op. cit.*

⁵⁷ « Planetoscope - Statistiques : Informations publiées dans le monde sur le net (en Gigaoctets) ».

⁵⁸ CARDON Dominique, *op. cit.*

⁵⁹ ACKOFF Russell, « From Data to Wisdom », 16, 1989.

⁶⁰ SIMONDON Gilbert, *Du mode d'existence des objets techniques*, AUBIER, 2012.

⁶¹ WINNER Langdon, « Do Artifacts Have Politics ? », *MIT Press*, 109, 1980.

⁶² WALSH Philip, « Do Algorithms Have Politics? », sur *Blog Thinking Tech*, publié le 2 février 2021.

⁶³ La puce RFID est un autre exemple de technologie dont le fonctionnement même facilite le traçage de l'individu. LAVELLE Sylvain, « Politiques des artefacts. », *Cités*, 39, Presses Universitaires de France, 2009.

normative du code informatique dans le cyberspace et les conséquences que cela entraîne⁶⁴. Dès lors, la description du contexte économique et social de l'IA doit être complétée par une analyse de son fonctionnement.

La force normative de la quantification. Le système d'IA est un outil statistique qui ne peut traiter que des valeurs numériques. La catégorisation, l'évaluation ou encore la prédiction produite par un tel logiciel se fonde uniquement sur des données chiffrées. Le processus de quantification auquel contribue les technologies d'IA ne peut être réduit à des enjeux purement mathématiques. Selon le professeur Alain Supiot, le processus de quantification est intimement normatif⁶⁵. La gouvernance par les nombres met en place de nouveaux rapports de pouvoirs entre les individus en faveur de ceux qui établissent les nombres. *A fortiori*, la gouvernance algorithmique ne déroge pas à ce constat. Quelle que soit la finalité poursuivie, le système d'IA s'appuie et produit une représentation du réel dont les modalités de conception ont été déterminées, ou du moins orientées lors de la conception. Plus qu'une machine, le système d'IA est aussi un instrument de pouvoir. Ce pouvoir est d'autant plus grand qu'il est généralement ignoré, volontairement ou non, par le droit⁶⁶. Réglementer les systèmes d'IA, c'est donc aussi encadrer des instruments normatifs et s'emparer des enjeux de pouvoir qu'ils représentent.

La force normative du code informatique. En dépit de l'existence de certaines séparations formelles, la technologie et le droit partagent donc un point commun important : ils sont tous deux une source de normes. Ce constat est particulièrement prégnant en matière informatique, et donc *a fortiori* pour les technologies d'IA. Parce qu'ils sont avant tout des systèmes informatiques, les systèmes d'IA contribuent à produire ce que le professeur Lawrence Lessig appelle des contraintes d'architecture. Au sein de sa théorie de la Nouvelle école de Chicago (aussi appelée théorie des points pathétiques), Lawrence Lessig liste quatre types de sources normatives : le droit, le marché, les normes sociales et l'architecture⁶⁷. Chacune d'entre elles a la capacité de réguler le comportement des individus. Pour illustrer, l'auteur prend l'exemple de la réduction de la consommation de cigarettes. Pour parvenir à cet objectif, le gouvernement dispose de plusieurs moyens. Premièrement, il peut faire voter une loi interdisant la cigarette. Deuxièmement, il peut taxer les cigarettes dans le but d'entraîner une augmentation des prix

⁶⁴ LESSIG Lawrence, *Code*, Basic Books, 2006. Pour plus d'informations, voir *supra*.

⁶⁵ La quantification peut servir quatre fonctions normatives : rendre compte, administrer, juger et légiférer. SUPIOT Alain, *La Gouvernance par les nombres*, Fayard, 2015.

⁶⁶ *Ibid.*

⁶⁷ « *The government must weigh the costs against the benefits and select the mode that regulates most effectively.* » - LESSIG Lawrence, « The New Chicago School », *The Journal of Legal Studies*, 27, [The University of Chicago Press, The University of Chicago Law School], 1998. P. 669.

afin de réduire la demande des consommateurs. Troisièmement, le gouvernement peut imposer des campagnes de sensibilisation contre la consommation de cigarettes comme c'est le cas en France avec l'imposition d'images sur les paquets. Quatrièmement, le gouvernement peut imposer aux fabricants de réduire progressivement le taux de nicotine dans les cigarettes. De cette manière, il espère diminuer l'intensité de l'addiction à la cigarette et donc sa consommation par les individus. Ces quatre sources normatives ne sont pas exclusives et peuvent être utilisées, notamment au travers du droit, de concert, comme c'est le cas en France pour le tabac (interdiction de vente aux mineurs, augmentation du prix du paquet et campagne publicitaire anti-tabac).

La concurrence normative avec le droit. La reconnaissance de la force normative du code informatique permet de le catégoriser comme une norme non-juridique au sens de la distinction établie par Carbonnier⁶⁸. À l'image de la morale, de la religion ou encore des règles d'hygiène, le code informatique peut donc interagir normativement avec la règle de droit. Ces phénomènes d'internormativité peuvent illustrer une cohérence entre les différents systèmes normatifs, à savoir lorsque les contraintes s'exercent dans une même direction⁶⁹. Dans le cas du droit et du code informatique, cela signifie que le concepteur du logiciel (ici d'IA) et le législateur poursuivent un objectif similaire. L'existence d'une cohérence entre le droit et les autres sources normatives en présence facilite alors son application et accroît son effectivité⁷⁰. À l'inverse, lorsque les normes en présence sont dissonantes, l'effectivité de la norme juridique s'en trouve réduite.

Dans le cas des systèmes d'IA, la force normative de l'IA présente différents points de tensions avec la norme juridique. Premièrement, le système d'IA peut être utilisé pour fouiller dans les données. Son objectif est alors d'identifier des schémas récurrents dans ces données par le biais de corrélations. Pour la professeure Antoinette Rouvroy, ces corrélations constituent des normes immanentes, à savoir des normes :

« qui ne s'imposent pas de l'extérieur, mais surgissent spontanément, pourrait-on dire, de la vie elle-même, du monde lui-même, de façon indépendante de toute qualification, de toute évaluation et de toute délibération. »⁷¹

⁶⁸ CARBONNIER Jean, *Essais sur les lois*, Dalloz, 1995. P. 147.

⁶⁹ *Ibid.*

⁷⁰ *Ibid.* LESSIG Lawrence, *op. cit.*

⁷¹ ROUVROY Antoinette et BERNIS Thomas, « Gouvernamentalité algorithmique et perspectives d'émancipation », *Rezeaux*, n° 177, La Découverte, 2013.

Constitutives de ce que l'auteur nomme la gouvernamentalité algorithmique⁷², ces normes immanentes peuvent diverger profondément des normes juridiques. En effet, les normes immanentes représentent, plus ou moins exhaustivement, le monde tel qu'il était au moment de la collecte des données à partir desquelles les normes ont été découvertes. *A contrario*, la norme juridique décrit le monde tel que la société pense qu'il devrait être, et non pas tel qu'il est (ou était)⁷³. Par exemple, le droit à la non-discrimination ne signifie pas que les actes discriminatoires n'existent pas dans la société actuelle. L'absence de discrimination est un objectif qui doit être atteint selon la société, d'où l'existence de la règle juridique. La distance qui sépare le monde réel du monde du « devoir être » représenté par la règle de droit est un élément central pour Alain Supiot⁷⁴. En effet, lorsque cette distance devient trop importante, la réalité disqualifie l'ordre normatif du droit et le décrédibilise. Dès lors, les normes immanentes produites par les systèmes d'IA peuvent donc entrer en tension avec la norme juridique sur ce point.

Secondement, un système d'IA est perçu au regard de sa capacité à faire les « bons » choix⁷⁵. La notion de « bon choix » est déterminée par le concepteur du système d'IA. En distinguant le bon du mauvais choix, le concepteur élabore et applique sa norme par le biais du système informatique. Cette norme peut alors entrer en tension avec la norme juridique, soit parce que celle-ci pose un objectif différent à atteindre, soit parce que l'atteinte de l'objectif du concepteur enfreint une règle de droit. Par exemple, la description du bon candidat à un emploi par le concepteur ne peut pas se fonder sur des caractéristiques discriminantes telles que le genre, l'origine ethnique ou encore les opinions politiques. Dans une logique similaire, une loi qui imposerait une parité homme / femme dans un milieu très majoritairement peuplé d'hommes est un objectif supplémentaire pouvant entrer en tension avec l'objectif initial du concepteur. Soulevé par les professeurs Russell et Norvig, le problème d'alignement des valeurs peut donc être abordé comme une situation de dissonance normative.

La constatation de la concurrence normative en droit de l'IA. Outre la présentation théorique des interactions normatives au sein du droit de l'IA, il est également possible d'aborder ce phénomène de façon plus concrète. En effet, la rencontre entre les normes juridiques et les normes d'origine technique se matérialise lors de la mise en application des

⁷² *Ibid.*

⁷³ SUPIOT Alain, *La Gouvernance par les nombres*, Fayard, 2015.

⁷⁴ *Ibid.*

⁷⁵ Voir I. §1. Les principes-objectifsL'explication de l'IA ; NORVIG Peter et RUSSELL Stuart, *Intelligence artificielle - Une approche moderne*, PEARSON, 2021.

obligations par le débiteur. Par exemple, en encadrant la conception des systèmes d'IA, le règlement européen sur l'IA⁷⁶ impose donc une conformité de la norme technique à la norme juridique. Ainsi, le respect de l'obligation par le débiteur entraîne une cohérence normative tandis que le manquement au droit entraîne une dissonance normative. L'état de l'articulation normative peut donc être constaté par l'évaluation de la preuve fournie par le débiteur lors d'un contrôle de la conformité du système d'IA. La preuve juridique d'une exigence relative à un système d'IA offre donc un prisme d'analyse privilégié de la cohérence ou la dissonance normative au sein du droit de l'IA.

III. LA PREUVE JURIDIQUE COMME OUTIL D'ANALYSE DU DROIT DES SYSTEMES D'IA

La preuve en droit. Dans le langage courant, la notion de preuve est définie par le biais de sa fonction, à savoir : « *ce qui sert à établir qu'une chose est vraie* »⁷⁷. Il convient néanmoins de préciser cette définition pour l'intégrer dans le champ juridique. De manière générale, la preuve juridique possède une double signification. Premièrement, elle désigne l'acte d'établir l'existence d'un fait ou d'un acte conformément aux formes admises et/ou requises par la loi⁷⁸. Secondement, il s'agit aussi du moyen qui permet d'établir la preuve du fait ou de l'acte⁷⁹. En d'autres termes, la preuve juridique désigne à la fois la fonction d'établissement d'une situation juridique (fait juridique ou acte juridique) et l'élément utilisé pour remplir cette fonction.

La fonction de la preuve en droit est sujet à débat au sein des juristes. En effet, l'établissement d'un acte juridique ou d'un fait juridique fait l'objet de deux courants distincts d'interprétation. D'un côté, l'objectif de la preuve serait l'établissement de la vérité. Ainsi, pour les professeurs Colin et Capitant : « *prouver, c'est faire reconnaître en justice la vérité d'une allégation par laquelle on affirme un fait d'où découlent des conséquences juridiques.* »⁸⁰. De la même manière, Domat définit la preuve comme « *ce qui persuade l'esprit d'une vérité* »⁸¹. La position

⁷⁶ « Règlement sur l'intelligence artificielle », 2024.

⁷⁷ « Preuve - Définitions, synonymes, prononciation, exemples | Dico en ligne Le Robert ».

⁷⁸ « *Démonstration de l'existence d'un fait ou d'un acte dans les formes admises et requises par la loi* » - CORNU Gérard, *Vocabulaire juridique*, 2022. « *Établissement de la réalité d'un fait ou de l'existence d'un acte juridique.* » - DEBARD Thierry et GUINCHARD Serge, *Lexique des termes juridiques 2022-2023*, Dalloz, 2022. « *Établissement, selon les formes prescrites par la loi, de l'existence d'un fait ou d'un acte* » - CABRILLAC Rémy, *Dictionnaire du vocabulaire juridique 2022*, LEXISNEXIS, 2021.

⁷⁹ « *Moyen employé pour faire la preuve* » - CORNU Gérard, *op. cit.* « *procédé utilisé à cette fin [de prouver] (écrit, témoignage ...)* » - DEBARD Thierry et GUINCHARD Serge, *op. cit.* « *Procédé utilisé pour établir la preuve d'un fait ou d'un acte* » - CABRILLAC Rémy, *op. cit.*

⁸⁰ COLIN Ambroise et CAPITANT Henri, *Cours élémentaire de droit civil français.*, 1930. P.197.

⁸¹ DOMAT Jean et DE BOUCHEVRET Benoît l'Usurier, *Les loix civiles dans leur ordre naturel, le droit public et Legum delectus.*, 1756. P. 245.

selon laquelle la preuve juridique a pour fonction de révéler la vérité doit cependant être nuancée au regard de la notion de « conviction ». En effet, si A. Colin et H. Capitant définissent la fonction de la preuve en justice comme l'établissement de la vérité, ils ajoutent ensuite que les preuves « *sont des éléments de conviction que les parties sont autorisées à soumettre à la justice* »⁸². Ainsi, de l'autre côté de la doctrine, de nombreux auteurs s'accordent à dire que la fonction de la preuve (en justice) est de convaincre le destinataire de l'existence des faits, qu'il s'agisse du juge ou des jurés. La fonction de la preuve est alors davantage de convaincre que d'aider à révéler une vérité objective sur les faits juridiques ou actes juridiques du litige⁸³. Dans le deuxième tome de son traité élémentaire de droit civil, le professeur Planiol débute ainsi son ouvrage par la définition suivante : « *On appelle "preuves" les divers procédés employés pour convaincre le juge* »⁸⁴. Pour le professeur Jean-Louis Bergel, le contexte du procès ne permet pas la recherche de la vérité. Ainsi, la preuve « *s'administre dans les limites d'un délai raisonnable nécessaire à l'intervention d'une décision et doit être considérée comme suffisante ou insuffisante, puisque le juge ne peut, sous peine de déni de justice, s'abstenir de juger* »⁸⁵. Certains auteurs font également le choix de mêler les deux courants au sein d'une définition⁸⁶. Par exemple, les professeurs Etienne Vergès, Géraldine Vial et Olivier Leclerc considère la preuve en justice comme « *les opérations visant à emporter la conviction du juge sur l'exactitude ou l'inexactitude d'une affirmation de fait dans une procédure juridictionnelle* »⁸⁷. L'état des lieux de ces conceptions doctrinales de la preuve met alors en avant différents critères de définition de la preuve en droit qu'il convient d'aborder avant d'énoncer la définition retenue pour la présente analyse.

La forme de la preuve en droit. La spécificité de la preuve en droit est un sujet généralement associé à celui de la définition de la notion. La conception de la preuve comme moyen d'emporter la conviction du juge permet notamment de comparer la preuve en droit à la preuve en science. Ainsi, la fonction de la preuve juridique serait un critère distinctif avec la preuve en science dont l'objectif serait, *a contrario*, la recherche de la vérité⁸⁸. L'existence de règles de procédure limitant l'admissibilité de certaines preuves est une autre différence, la distinction

⁸² COLIN Ambroise et CAPITANT Henri, *op. cit.*

⁸³ « *En matière judiciaire, la preuve a plus pour objet de convaincre le juge que d'établir la réalité objective* » - BERGEL Jean-Louis, *Théorie Générale du Droit*, Dalloz, 2012. P. 329.

⁸⁴ PLANIOL Marcel (1853-1931) Auteur du texte, *Traité élémentaire de droit civil*, 1922.

⁸⁵ BERGEL Jean-Louis, *op. cit.*

⁸⁶ Voir notamment : GLASSON Ernest, *Précis théorique et pratique de procédure civile*, 1908.

⁸⁷ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 8.

⁸⁸ « *[...]le contexte et le but de la preuve judiciaire la différencient de la preuve scientifique* » - BERGEL Jean-Louis, *op. cit.*

étant alors réduite au périmètre du système de preuve légale⁸⁹. Pour d'autres, la preuve en droit n'est pas différente de la preuve prise dans d'autres contextes. La preuve n'est alors qu'une information grâce à laquelle les faits sont prouvés⁹⁰. Ainsi, pour Bentham :

*« toute preuve comprend au moins deux faits distincts : l'un que l'on peut appeler le fait principal, celui dont il s'agit de prouver qu'il existe ou qu'il n'existe pas ; l'autre le fait probatoire, celui qui est employé à prouver le oui ou le non du fait principal. »*⁹¹

Dès lors, la preuve n'est qu'un « *moyen pour une fin* »⁹². Cette absence de spécificité peut alors être vue comme une invitation à l'interdisciplinarité⁹³.

La présente thèse ne vise pas à offrir une réponse à l'existence ou non d'une spécificité de la preuve en droit. La notion de preuve sera donc toujours entendue au sens juridique. Cependant, le terme de preuve sera parfois accompagné de qualificatifs, en particulier l'expression de « preuve technique ». À nouveau, il n'est pas question de s'éloigner du périmètre de la preuve en droit. Le recours à des qualificatifs désigne non pas la nature de la preuve mais sa forme. En tant qu'élément, la preuve peut revêtir différentes formes. À titre d'exemple, une preuve statistique est une preuve juridique qui prend la forme d'une statistique. Il ne s'agit en aucun cas d'une preuve au sens mathématique du terme. Le qualificatif « statistique » désigne ici la forme de la preuve et non sa nature⁹⁴. Par conséquent, il est central de garder à l'esprit la différence entre la forme de la preuve, qui est variable, et sa nature, qui demeurera toujours juridique.

La preuve juridique. Le second point indispensable à aborder est le contexte d'utilisation de la preuve. De manière générale, le juriste restreint son intérêt pour la preuve en droit à la seule preuve en justice⁹⁵. L'étude de la preuve en droit est alors circonscrite au périmètre des règles de droit relatives à la preuve⁹⁶. Dans le contexte du droit de l'IA, la limitation à la seule preuve

⁸⁹ LECLERC Olivier, « La distinction entre la “preuve en droit” et la “preuve en science” est-elle pertinente ? », in *Preuve scientifique, preuve juridique*, Larcier, 2011.

⁹⁰ KEANE Adrian et MCKEOWN Paul, *The Modern Law of Evidence*, Oxford University Press, 2012. P. 1.

⁹¹ DUMONT Étienne et BENTHAM Jeremy, *Traité des preuves judiciaires*, 1823. P. 16.

⁹² *Ibid.*

⁹³ PARK Roger C. et SAKS Michael J., « Interdisciplinary Trends in Evidence Scholarship », 2005. TWINING William, « Evidence as a Multi-disciplinary Subject », *Law, Probability and Risk*, 2, 2008.

⁹⁴ Dans les cas où la notion de preuve sera entendue dans un sens autre que juridique, l'auteur le mentionnera et l'expliquera explicitement.

⁹⁵ TRUILHÉ-MARENGO Eve, *Preuve scientifique, preuve juridique*, Larcier, 2011. P. 8. « Dans l'acceptation commune l'art de la preuve semble particulièrement appliqué à la pratique des tribunaux ; c'est là son point saillant » - DUMONT Étienne et BENTHAM Jeremy, *op. cit.*

⁹⁶ LECLERC Olivier, « La distinction entre la “preuve en droit” et la “preuve en science” est-elle pertinente ? », in *Preuve scientifique, preuve juridique*, Larcier, 2011.

en justice nous apparaît particulièrement limitante et dommageable dans la réalisation de notre étude. En effet, la preuve en justice est celle destinée à emporter la conviction du juge, ou des jurés. Les destinataires de la preuve en droit sont alors grandement réduits. Par exemple, le règlement européen sur l'IA du 13 juin 2024⁹⁷ met en avant une grande diversité d'acteurs en charge de vérifier l'application du règlement à différents niveaux (autorités administratives, organismes privés, etc.) qui se retrouveraient exclus du périmètre de l'analyse si la preuve venait à être limitée à la seule preuve judiciaire. Par conséquent, il est nécessaire de s'émanciper du seul contexte contentieux pour aborder la preuve juridique dans sa globalité et intégrer l'ensemble des destinataires des preuves juridiques utilisées en droit de l'IA.

Au regard de l'ensemble de ces éléments, la preuve juridique désigne ci-après les opérations et les moyens visant à emporter la conviction de son destinataire quant à l'exactitude d'un fait juridique ou d'un acte juridique.

La preuve comme indicateur d'effectivité de la réglementation sur l'IA. En 1923, le professeur Planiol écrivait à propos de la preuve en justice :

*« Un droit n'est rien sans la preuve de l'acte juridique ou du fait matériel d'où il dérive. Dès que cet acte ou ce fait est contesté, on a besoin de le prouver pour convaincre le juge de l'existence du droit, faute de quoi on ne peut en obtenir le respect. C'est donc la preuve qui vivifie le droit et le rend utile. »*⁹⁸

De la même manière, Bentham fait un lien direct entre l'effectivité des lois et les règles de procédure⁹⁹. Or, pour reprendre ces mots : *« l'art de la procédure n'est essentiellement que l'art d'administrer la preuve »*¹⁰⁰. Par conséquent, la preuve est un élément indispensable de l'effectivité d'une norme juridique¹⁰¹.

Les systèmes d'IA entretiennent différentes interactions avec l'effectivité de la règle de droit. Premièrement, l'utilisation d'un système d'IA entraîne de nombreux défis probatoires affectant l'effectivité de la norme juridique. Un exemple particulièrement représentatif de ce problème

⁹⁷ « Règlement sur l'intelligence artificielle », 2024.

⁹⁸ PLANIOL Marcel (1853-1931) Auteur du texte, *Traité élémentaire de droit civil*, 1922.

⁹⁹ « Ces lois ne seraient d'aucun effet si le législateur ne créait en même temps d'autres lois, qui ont pour objet de faire accomplir les premières : ce sont les lois de la procédure » - DUMONT Étienne et BENTHAM Jeremy, *Traité des preuves judiciaires*, 1823. P. 2.

¹⁰⁰ *Ibid.* P. 3.

¹⁰¹ « Le droit existe indépendamment de toute preuve mais il ne peut prévaloir, en cas de contestation, que si l'on parvient à mettre en lumière les faits et les actes de nature à en confirmer la véracité » - BERGEL Jean-Louis, *Théorie Générale du Droit*, Dalloz, 2012.

est la preuve d'une discrimination algorithmique. Ainsi, l'opacité particulièrement importante des systèmes d'IA complexifie grandement l'accès aux preuves par la prétendue victime¹⁰². Le risque est alors que les exigences probatoires incombant à la partie deviennent en pratique impossible à satisfaire¹⁰³, en particulier dans le cas d'une discrimination indirecte¹⁰⁴. Dans le cas d'un système d'IA biaisé de façon discriminatoire, la norme technique, à savoir le fonctionnement du système d'IA, n'est pas cohérente avec la norme juridique. Au niveau européen, la difficulté de l'accès à la preuve, du fait de l'opacité algorithmique, a d'ores et déjà été identifiée comme un obstacle majeur à l'effectivité de la règle de droit appliquée aux systèmes d'IA. En septembre 2022, la Commission européenne a publié une proposition de directive relative à la responsabilité en matière d'IA dont le seul objectif est de faciliter l'accès aux preuves par les personnes lésées et l'allègement de leur charge probatoire¹⁰⁵. Également, la proposition de directive européenne relative à la responsabilité du fait des produits défectueux a été élaborée dans un objectif similaire¹⁰⁶. Le texte contient notamment des dispositions facilitant l'accès à la preuve afin de résoudre les problèmes liés à la charge probatoire des personnes lésées¹⁰⁷. La preuve juridique est donc à la fois une condition majeure de l'effectivité de la réglementation sur les systèmes d'IA, et par conséquent un prisme d'analyse particulièrement pertinent pour étudier ce phénomène.

Deuxièmement, le législateur européen a fait le choix d'encadrer spécifiquement les systèmes d'IA, en particulier les phases préalables à leur mise sur le marché. En effet, le récent règlement (UE) 2024/1689 sur l'intelligence artificielle¹⁰⁸ élabore des obligations relatives au fonctionnement de certains systèmes d'IA¹⁰⁹. La réglementation de certains aspects techniques

¹⁰² GROZDANOVSKI Ljupcho, « In search of effectiveness and fairness in proving algorithmic discrimination in EU law », *Common Market Law Review*, 58, 2021. P. 112.

¹⁰³ *Ibid.*

¹⁰⁴ La charge de la preuve est beaucoup plus difficile dans le contexte d'une discrimination indirecte que dans le cas d'une discrimination directe. Voir notamment : Cour de justice des Communautés européennes, 31 mai 1989, n° 109/88. Par. 30.

¹⁰⁵ « La présente directive établit des règles communes concernant : (a) la divulgation d'éléments de preuve sur les systèmes d'intelligence artificielle (IA) à haut risque afin de permettre à un demandeur d'étayer une action civile en réparation de dommages causés par une faute extracontractuelle ; (b) la charge de la preuve en cas d'actions civiles fondées sur une faute extracontractuelle introduites devant les juridictions nationales pour des dommages causés par un système d'IA. » - Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 1.1.

¹⁰⁶ « En particulier, la présente proposition vise à [...] alléger la charge de la preuve dans les cas complexes » - COMMISSION EUROPÉENNE, Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux, 28 septembre 2022.

¹⁰⁷ « Toutefois, la directive présentait aussi plusieurs lacunes : [...] la charge de la preuve (c'est-à-dire la nécessité, pour obtenir réparation, de prouver que le produit était défectueux et que ce défaut avait causé le dommage subi) posait des difficultés aux personnes lésées dans des cas complexes » - *Ibid.*

¹⁰⁸ « Règlement sur l'intelligence artificielle », 2024.

¹⁰⁹ Voir Section II : Une conformité juridique ex ante pour les systèmes d'IA à haut risque.

des systèmes d'IA fait interagir inévitablement les normes juridiques avec les normes issues de la technique. Dans ce contexte, le recours à la preuve juridique dans notre analyse permet d'étudier de façon prospective les potentielles difficultés d'application du règlement européen sur l'IA. La preuve est donc un éclairage original sur l'effectivité de ce nouveau texte central dans la réglementation sur les systèmes d'IA.

La place de la preuve dans l'effectivité des textes juridiques sur l'IA n'est toutefois pas un phénomène spécifique au droit européen. Ainsi, bien que la présente thèse se concentre sur le droit de l'Union, le raisonnement et les apports qui y seront développés pourront être réutilisés dans d'autres contextes. Tout d'abord, si l'Union européenne apparaît comme une pionnière de la réglementation de l'IA, de nombreux États sont également en cours d'élaboration de lois spécifiques aux systèmes d'IA. À titre d'exemples, les États-unis¹¹⁰, le Canada¹¹¹ ou encore le Brésil¹¹² et la Corée du sud¹¹³ sont en cours d'adoption de projets de loi relatifs à l'IA en cours d'adoption. Une fois ces législations votées et mises en application, la question de la preuve juridique deviendra, à l'image de la réglementation européenne, un sujet décisif pour garantir leur effectivité dans tous les systèmes juridiques. En outre, dans une société globalisée, le marché européen de l'IA interagit naturellement avec les autres marchés du monde et joue un rôle d'influence important¹¹⁴. Afin de faciliter les échanges commerciaux de systèmes d'IA, il sera indispensable que les différentes réglementations sur l'IA offrent des garanties juridiques de niveau similaire, notamment concernant la protection des droits et libertés des individus, afin de mettre éventuellement en œuvre une reconnaissance mutuelle. Pour évaluer cette équivalence juridique, les preuves demandées par chacun des textes seront très probablement un facteur majeur à prendre en compte. Par conséquent, les développements de notre thèse

¹¹⁰ *Algorithmic Accountability Act*, 3 février 2022, H.R.6580. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence », 2023.

¹¹¹ *Loi sur l'intelligence artificielle et les données*, 16 juin 2022, C-27.

¹¹² *Projeto de lei nº2338 dispõe sobre o uso da Inteligência Artificial (Projet de loi relatif à l'utilisation de l'IA)*, 3 mai 2023. Traduction réalisée par l'auteur. Le projet de loi vient s'ajouter au projet de loi 21/2020 déposé le 11 février 2020 : *Projeto de lei nº 21/2020 estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências (Projet de loi nº 21/2020 établissant des principes, des droits et des devoirs pour l'utilisation de l'intelligence artificielle au Brésil et prévoit d'autres dispositions)*, 11 février 2020. Voir également : AFFONSO SUZA Carlos, PARRA CANO Flavia et BRANCO Sergio, « La réglementation de l'intelligence artificielle au Brésil : la décortication des propositions législatives actuelles », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle. Entre règles sectorielles et régime général. Perspectives comparées.*, Bruylant, 2023.

¹¹³ *Loi sur l'intelligence artificielle (Corée du sud)*, 14 février 2023. Traduction par l'auteur.

¹¹⁴ BRADFORD Anu, *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020. SIEGMANN Charlotte et ANDERLJUNG Markus, « The Brussels Effect and Artificial Intelligence », arXiv, 2022.

pourront être exploités au sein de travaux s'inscrivant en dehors du droit de l'Union européenne¹¹⁵.

La preuve comme lien entre le fait informatique et le droit. Le choix d'intégrer la preuve juridique au sein de notre analyse sur le droit des systèmes d'IA présente également un second avantage du fait de la relation qu'elle entretient avec le fait. Au premier abord, faits et droit peuvent apparaître comme deux notions immiscibles, voire opposées¹¹⁶. « *Un fait ne peut produire d'effets juridiques par ses qualités intrinsèques* »¹¹⁷ écrit Bergel. Dans le contexte du procès, les parties n'ont la charge que de la preuve des questions de faits¹¹⁸. D'ailleurs, certains faits sont explicitement écartés en ce qu'ils n'intéressent pas le droit. C'est le cas des faits certains qui, par définition, sont présumés comme vrais¹¹⁹. Toutefois, la distinction entre le fait et le droit se brouille lors de la mise en application des règles de droit¹²⁰. En effet, celle-ci nécessite une qualification juridique des faits, les faisant alors pénétrer dans le droit¹²¹. Grâce aux preuves, le juge peut déterminer les conséquences juridiques applicables aux faits qualifiés en présence.

Les systèmes d'IA, et de manière générale les éléments impliqués dans leur conception et leur utilisation, appartiennent aux faits. En raison du cadre juridique appliqué aux systèmes d'IA, certains de ces faits seront amenés à être utilisés comme preuves et donc à être qualifiés juridiquement. Ce passage du monde des faits au monde du droit est un espace privilégié pour étudier les enjeux normatifs autour du droit des systèmes d'IA. Les systèmes d'IA ne sont pas des outils technologiques neutres, ils ont une force normative qui affecte le comportement des individus dans une direction donnée, conformément à certaines valeurs politiques et sociales.

¹¹⁵ Néanmoins, il convient de mentionner le travail d'adaptation non négligeable qu'il sera nécessaire d'effectuer pour une personne souhaitant appliquer les raisonnements et les apports développés ci-après dans un cadre juridique autre que le droit européen et le droit français.

¹¹⁶ « *The opposition of fact and law is widely accepted in the current legal writings [...] The opposition in question, however, is far from clear* » - WRÓBLEWSKI Jerzy, « Facts in Law », *ARSP: Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, 59, Franz Steiner Verlag, 1973.

¹¹⁷ BERGEL Jean-Louis, *Théorie Générale du Droit*, Dalloz, 2012. P. 334.

¹¹⁸ DEUMIER Pascale, *Introduction générale au droit*, 5^e éd., LGDJ, 400 p.. CABRILLAC Rémy, *Introduction générale au droit*, Dalloz, 2023 CARBONNIER Jean, *Droit civil. Introduction*, Presses Universitaires de France - PUF, 2002 CORNU Gérard, *Droit civil. Introduction au droit*, LGDJ, 2007 MORVAN Patrick et MALAURIE Philippe, *Introduction au droit*, LGDJ, 2022 THOMAT-RAYNAUD Anne-Laure et BEIGNIER Bernard, *Introduction au droit*, LGDJ, 2024.

¹¹⁹ « *L'usage des preuves ne regarde pas les faits qui sont naturellement certains, et dont la vérité est toujours présumé, si le contraire n'est prouvé ; mais il regarde seulement les faits incertains, et dont la vérité n'est pas présumée si elle n'est prouvée* » - DOMAT Jean et DE BOUCHEVRET Benoît l'Usurier, *Les loix civiles dans leur ordre naturel, le droit public et Legum delectus.*, 1756.P. 247.

¹²⁰ PERELMAN Ch., « La Distinction Du Fait Et Du Droit Le Point De Vue Du Logicien », *Dialectica*, 15, Wiley, 1961 WRÓBLEWSKI Jerzy, « Facts in Law », *ARSP: Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, 59, Franz Steiner Verlag, 1973.

¹²¹ BERGEL Jean-Louis, *Théorie Générale du Droit*, Dalloz, 2012.

Parfois, les faits juridiques recherchés sont aussi empreints de valeurs¹²². Le recours à des preuves non neutres pour convaincre le juge de leur existence peut faire apparaître un alignement ou une opposition entre ces valeurs. La preuve juridique est donc à la fois un outil d'effectivité du droit des systèmes d'IA et la représentation du phénomène d'internormativité qui se déroule au sein du système d'IA.

La place de la preuve en droit des systèmes d'IA. Problématique de la thèse. L'analyse de la preuve permet donc de mettre en lumière les interactions entre la norme technique et la norme juridique mais aussi les conséquences de ces interactions sur le droit des systèmes d'IA. Ainsi, la présente recherche vise à répondre à la problématique suivante : quels rôles occupent la preuve au sein de la réglementation sur les systèmes d'IA ? Quelles en sont les conséquences au sein de ce cadre juridique ? Tout d'abord, l'étude du rôle de la preuve nécessite d'analyser comment les conflits de normes et de valeurs, présentés précédemment, se matérialisent au sein de la preuve juridique dans le contexte du droit des systèmes d'IA. Egalement, la présente recherche s'étend à l'étude de la place de la preuve, à savoir l'attention et l'implication accordée à la preuve par les différents acteurs du fait de leur rôle dans la réglementation, qu'il s'agisse du législateur, des débiteurs des obligations ou des personnes en charge de l'évaluation desdites preuves. La place accordée sera alors mise en perspective avec le rôle réel occupé par la preuve dans le droit des systèmes d'IA. Ce dernier sera alors analysé au regard de l'importance des conséquences de la preuve considérée comme pertinente pour attester de la conformité au droit. La présente étude souhaite donc mettre en lumière l'intensité et l'ampleur des conséquences de la preuve, et donc indirectement du phénomène d'internormativité, sur l'ensemble de la réglementation sur les systèmes d'IA.

Les interactions entre la preuve juridique et les systèmes d'IA. Afin d'exploiter au mieux le prisme de la preuve, il est nécessaire d'identifier les relations qu'elle entretient avec les systèmes d'IA. Cette articulation peut être réduite à trois liens majeurs. Tout d'abord, la preuve juridique et le système d'IA présentent des liens au sein du règlement européen sur l'IA¹²³. Le texte impose des obligations relatives à la conception de certains systèmes d'IA pour les fournisseurs desdits logiciels. Ces derniers doivent se mettre en conformité, d'une part, et de constituer des preuves pour attester de cette mise en conformité, d'autre part. Ce travail doit notamment se faire grâce à des essais techniques réalisés pendant la conception du système

¹²² « *The facts which are demonstrated in law are not value-free and, hence, for the proof of their existence one has to evaluate* » - WRÓBLEWSKI Jerzy, *op. cit.* P. 163. PERELMAN Ch., *op. cit.*

¹²³ « Règlement sur l'intelligence artificielle », 2024.

d'IA. L'élaboration et l'application de ces procédures d'évaluations ainsi que les résultats qui en découlent sont autant de preuves de conception pertinentes pour convaincre de la conformité du logiciel¹²⁴, et ainsi obtenir son autorisation d'entrer sur le marché.

Ensuite, les résultats produits par un système d'IA sont aussi soumis à des exigences juridiques. Ainsi, une décision prise par un système d'IA de désavantager ou retirer un contenu présent sur une plateforme en ligne doit respecter le droit à la liberté d'expression. De la même manière, le système d'IA chargé de recruter les candidats pour un emploi doit produire des décisions conformes au droit à la non-discrimination. De façon générale, le résultat d'un système d'IA qui causerait un préjudice pourrait engager la responsabilité de son fournisseur ou de son déployeur. En effet, le contenu des propositions des directives européennes relatives à la responsabilité en matière d'IA¹²⁵ et à la responsabilité du fait des produits défectueux¹²⁶ mettent ainsi en lumière l'importance de la preuve dans ce contexte. Dès lors, la personne lésée doit prouver le fait générateur, le lien de causalité et le préjudice subi afin d'espérer obtenir réparation. Que ce soit au travers des régimes de responsabilité civile ou du régime de protection des droits fondamentaux, le demandeur et le défendeur doivent donc démontrer par la preuve l'existence ou l'absence des faits constitutifs de responsabilité du fait d'un système d'IA.

Enfin, la preuve juridique peut être produite par un système d'IA. En effet, le système d'IA, à l'image d'autres technologies, peut être utilisé comme procédé de preuve pour convaincre le juge de l'exactitude d'un fait juridique. Si la reconnaissance faciale en est l'exemple le plus populaire, les recours aux systèmes d'IA comme instrument probatoire se multiplient¹²⁷. À titre d'exemple, la direction générale des finances publiques (DGFIP ci-après) a généralisé l'utilisation d'un système d'IA à l'ensemble du territoire français dans le but de détecter et de sanctionner les piscines de particuliers non déclarés à l'administration fiscale¹²⁸. La preuve juridique n'est donc pas seulement au service de la conformité réglementaire du système d'IA

¹²⁴ « Les essais garantissent que les systèmes d'IA à haut risque fonctionnent de manière conforme à leur destination et qu'ils sont conformes aux exigences énoncées dans la présente section. » - Ibid. Article 8.6.

¹²⁵ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022.

¹²⁶ Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 28 septembre 2022.

¹²⁷ De nombreux projets de recherche européens sont financés dans l'objectif de concevoir des systèmes d'IA pour aider les autorités répressives dans leurs activités. À titre d'exemple, voir : « Home | Starlight ».

¹²⁸ « Le fisc généralise la détection des piscines non déclarées par intelligence artificielle », sur *Les Echos*, publié le 29 août 2022.

mais peut être aussi générée par ce dernier. Par conséquent, l'utilisation d'un système d'IA s'articule également avec le droit de la preuve en justice.

La présentation de ces liens les uns à la suite des autres présente un avantage indéniable de clarté et de pédagogie. Cependant, elle laisse aussi à penser séparément ces différents phénomènes, dont le seul point commun serait le fait d'être des formes d'articulations entre le système d'IA et la preuve. Il n'en est rien. D'une part, les différentes branches du droit auxquelles appartiennent chacun de ces phénomènes interagissent ensemble, bien au-delà des seules règles applicables aux systèmes d'IA. D'autre part, l'articulation entre le système d'IA et la preuve n'est pas une simple caractéristique commune entre ces phénomènes mais un lien entre eux. En d'autres termes, les différentes articulations entre le système d'IA et la preuve juridique ne doivent pas être considérées comme trois phénomènes distincts mais comme un ensemble dont les composantes exercent des effets les uns sur les autres. Par exemple, il sera démontré que la manière avec laquelle la preuve judiciaire produite par un système d'IA est traitée pourra entraîner des conséquences sur les preuves exigées lors de la conception dudit système, et inversement. Dès lors, l'étude du droit des systèmes d'IA sous le prisme de la preuve nécessite la prise en compte de ces trois phénomènes. En effet, l'absence de l'un d'entre eux nuirait grandement à la compréhension globale du sujet.

L'étude du droit des systèmes d'IA par le prisme de la preuve. Annonce de plan. La preuve matérialise donc la concurrence normative présente lors de la mise en application du droit des systèmes d'IA. L'étude de la preuve juridique est d'autant plus signifiante que le législateur se concentre précisément sur la norme technique. En d'autres termes, plus les règles de droit encadrent des aspects techniques du système d'IA, plus l'interaction entre le droit et la technique sera forte, plus la preuve du respect des obligations juridiques représentera la cohérence ou la dissonance normative en présence. En imposant des exigences spécifiques à la conception des systèmes d'IA, le règlement européen sur l'IA est la meilleure illustration de ce phénomène. Par le biais dudit texte, le législateur européen exige que des systèmes d'IA particuliers, les systèmes d'IA à haut risque, présentent certaine qualité de fonctionnement. À titre d'exemples, les fournisseurs de ces systèmes d'IA doivent utiliser des jeux de données de conception « *suffisamment représentatifs* »¹²⁹ ou encore « *complets* »¹³⁰ ainsi que garantir « *un niveau approprié d'exactitude* »¹³¹ de leur logiciel. Pour vérifier la conformité des systèmes

¹²⁹ « Règlement sur l'intelligence artificielle », 2024. Article 10.3.

¹³⁰ *Ibid.*

¹³¹ *Ibid.* Article 15.1.

d'IA à ces exigences, le règlement européen impose le recours à des tests réalisés sur le système d'IA. Les résultats de ces tests, représentant alors le fonctionnement du système d'IA, sont donc des preuves qui seront utilisées pour convaincre l'auteur d'un contrôle de la conformité du logiciel. Il apparaît alors clairement que la preuve proposée par le fournisseur du système d'IA concerné est une représentation de la norme technique intégrée dans le logiciel. Un contrôle réussi renvoie donc l'existence d'une cohérence normative alors qu'un contrôle constatant un manquement signifie la présence d'une dissonance normative. Ces preuves de l'exécution des obligations de conception prévues par le règlement sur l'IA, que nous appellerons des preuves de conception pour la suite de l'étude, représentent donc de façon particulièrement directe et visible le résultat de la concurrence entre la norme technique et la norme juridique. Par conséquent, ces preuves de conception sont un élément central de notre étude et constitueront donc le point de départ de notre recherche (**Partie 1**). En attestant de la conformité du système d'IA à haut risque, les preuves de conception permettent l'obtention d'un certificat européen nécessaire à l'entrée sur le marché dudit logiciel¹³². Elles sont donc des éléments justifiant le bon fonctionnement du système d'IA et favorisent la confiance des individus lors de la phase d'utilisation du logiciel. Par exemple, en cas de préjudice causé par un système d'IA, la caractérisation de la faute, ou de l'absence de faute, du fournisseur peut être réalisée à l'aide des preuves de conception. Cette hypothèse est notamment renforcée par la proposition de directive européenne sur la responsabilité en matière d'IA qui réalise un lien direct entre la conformité du système d'IA et la faute du fournisseur¹³³. De la même manière, le juge qui est amené à apprécier le résultat d'un système d'IA, en tant que preuve judiciaire, pourrait utiliser des preuves de conception dudit logiciel pour éclairer son évaluation. Les liens discutés précédemment entre les différentes composantes de notre recherche sur la preuve dans le droit des systèmes d'IA sont ici manifestes. Pour traiter le sujet dans son ensemble, il sera donc nécessaire d'examiner les conséquences des preuves de conception lors de l'utilisation des systèmes d'IA (**Partie 2**).

¹³² *Ibid.* Article 99.

¹³³ « la condition énoncée au paragraphe 1, point a) [la faute du fournisseur à haut risque] n'est remplie que si le plaignant a démontré que le fournisseur ou, le cas échéant, la personne soumise aux obligations qui incombent au fournisseur n'a pas respecté l'une des exigences suivantes, énoncées dans lesdits chapitres, compte tenu des mesures prises dans le cadre du système de gestion des risques et des résultats de ce système conformément à [l'article 9 et à l'article 16, point a), de la législation sur l'IA] » - Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 4.2.

PARTIE 1 : LES PREUVES RELATIVES AUX EXIGENCES DE CONCEPTION DES SYSTEMES D'IA

Introduction. Les bénéfices prometteurs associés aux technologies d'IA a conduit l'Union européenne à préparer son marché unique numérique afin d'accueillir une production et une utilisation de plus en plus importante de ces outils¹³⁴. Cette adaptation du marché de l'IA s'accompagne d'une évolution du droit européen. Avec le vote très récent de son ambitieux règlement européen sur l'IA, l'Union européenne fait figure de pionnière dans le droit de l'IA. Le droit de l'IA est entendu ici comme l'ensemble des règles juridiques applicables aux technologies d'IA. Extrêmement large, le droit de l'IA intègre tous les secteurs d'activité et comprend de nombreuses branches du droit¹³⁵. Le nouveau règlement européen sur l'IA s'intègre dans ce cadre juridique, sans pour autant en constituer la totalité¹³⁶. Proposé pour la première fois par la Commission européenne le 21 avril 2021, le texte est finalement adopté presque trois ans plus tard le 13 juin 2024. Au regard des nombreux problèmes juridiques posés par le recours aux technologies d'IA, le règlement concentre des attentes et des enjeux considérables. De plus, le caractère extrêmement récent de ce texte européen en fait un sujet d'analyse particulièrement intéressant. Le règlement européen illustre alors la manière dont le législateur européen a choisi d'aborder cette innovation technologique qu'est l'intelligence artificielle.

Annonce de plan. Au regard de la nouveauté du règlement, il convient, dans un premier temps, de s'attarder sur ce-dernier afin de comprendre et évaluer la manière avec laquelle le législateur européen est venu encadrer spécifiquement les systèmes d'IA (**Titre 1**). Une fois l'étude du règlement terminée, il sera temps d'explorer le rôle joué par la preuve dans ce nouveau régime juridique (**Titre 2**).

¹³⁴ « Les priorités de la Commission ».

¹³⁵ MERABET Samir, *Vers un droit de l'intelligence artificielle*, 2020.

¹³⁶ Sur les nombreuses initiatives législatives en droit européen, voir notamment : *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, (2022) ; et COMMISSION EUROPÉENNE, *Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux*, 28 septembre 2022.

TITRE I : L'APPARITION D'UN DROIT SPECIFIQUE AUX SYSTEMES D'IA

Introduction. Droit de la cybersécurité, droit des données à caractère personnel, droit des contrats informatiques, le droit du numérique est la branche du droit qui étudie les interactions entre le droit et les innovations technologiques. À ce titre, la démocratisation de l'utilisation des technologies d'IA est un de ces bouleversements et pose ainsi des défis majeurs aux cadres juridiques nationaux comme supranationaux. Le droit français et le droit européen font actuellement l'objet d'une grande émulation sur le sujet de l'IA. Les législateurs disposent de plusieurs stratégies pour conserver la pertinence des règles de droit. Ils peuvent faire évoluer les dispositions existantes ou créer de toute pièce un nouveau cadre juridique. Le législateur européen a choisi de ne pas céder à ce dilemme et déploie deux approches complémentaires avec à la fois une évolution des dispositions existantes mais surtout avec l'élaboration d'une réglementation spécifique à l'IA dont le règlement européen sur l'IA. L'IA, au sens de l'ensemble des disciplines scientifiques qui la compose, et le droit interagissent alors l'un avec l'autre au sein du texte. Les preuves apparaissent alors comme un lieu extrêmement propice à l'articulation entre ces deux champs disciplinaires. Au premier abord, l'hypothèse selon laquelle les preuves juridiques, en particulier celles revêtant une forme technique, occuperaient une place particulière au sein du règlement sur l'IA est difficile à saisir. Afin d'étudier dans quelle mesure ces preuves jouent un rôle clé dans la future réglementation européenne sur l'IA, il faut au préalable comprendre le processus qui a amené le législateur à leur confier, consciemment ou non, cette importance.

Annonce de plan. Pour cela, il convient d'étudier tout d'abord le paysage normatif existant autour de l'IA dans le but d'identifier ce qui constitue les problématiques clés de la future réglementation sur l'IA (**Chapitre 1**). Ensuite, il sera analysé dans quelles mesures les exigences contenues dans règlement européen sur l'IA répondent à ces besoins (**Chapitre 2**).

CHAPITRE 1 : LES NORMES APPLICABLES AUX TECHNOLOGIES D'IA

Introduction. Que ce soit en droit européen ou ailleurs, les initiatives visant à encadrer juridiquement les technologies d'IA sont relativement récentes. Cependant, l'absence d'exigences spécifiques envers ces technologies ne signifie pas que l'utilisation de l'IA soit dépourvue de contraintes normatives. Le droit de l'IA vient donc occuper une place dans un environnement normatif loin d'être vierge. Les initiatives législatives visant à encadrer les technologies d'IA sont motivées par la volonté de répondre à certains besoins présents dans le contexte normatif existant. Pour mettre en lumière ces besoins, il est important d'observer les objectifs exprimés par les législateurs au travers des différents documents préparatoires produits. Toutefois, pour atteindre une compréhension étendue des enjeux sous-jacents au droit de l'IA, il est indispensable d'étudier l'environnement normatif dans lequel il viendra s'insérer. L'analyse des textes existants, éthiques et juridiques, permet d'identifier les besoins auxquels la nouvelle réglementation sur l'IA pourrait satisfaire. À l'aide de ces éléments, il sera possible d'élaborer une grille de lecture de l'importance des dispositions contenues dans les futurs textes en droit de l'IA. En d'autres termes, l'importance d'une exigence sera analysée au regard de sa capacité à répondre concrètement aux besoins préalablement identifiés.

Annonce de plan. Tout d'abord, le choix d'intégrer les normes éthiques dans le périmètre d'étude s'explique par la place très importante qu'elles occupent dans le paysage normatif de l'IA. D'une part, les textes éthiques concernant l'IA sont extrêmement nombreux et ont initié en grande partie les réflexions autour des défis et risques liés à l'IA dans notre société. D'autre part, les différents acteurs du secteur de l'IA (entreprises, société civile, universitaires, États, etc.) utilisent ces textes pour formuler leurs souhaits quant au droit de l'IA. Les textes éthiques sont donc une représentation intéressante des besoins du secteur (**Section I**). Ensuite, les normes juridiques de droit commun doivent également être envisagées. Bien qu'elles ne soient pas spécifiques à l'IA, ces exigences s'appliquent tout de même aux cas d'utilisation des technologies d'IA. Parce qu'il vient se superposer à ces dispositions, le droit de l'IA doit s'articuler correctement avec elles afin de limiter au maximum les risques d'incohérences et donc d'insécurité juridique. Dans ce but, il est indispensable d'avoir conscience des objectifs poursuivis, mais aussi des lacunes existantes, au sein de ces dispositions de droit positif. L'étude de ces problèmes permettra ainsi de mettre en lumière les besoins issus des normes juridiques (**Section II**).

Section I : Les principes éthiques : sources d'inspiration du droit de l'IA

Annnonce de plan. Devant l'utilisation toujours plus importante des technologies d'IA et les dangers inhérents qui l'accompagnent, de nombreux acteurs de la société se sont investis dans la compréhension et la lutte de ces risques. Sans attendre le processus législatif au rythme plus lent, des entreprises, administrations, institutions, associations et autres entités ont élaboré et publié leurs réponses à ces problématiques sous la forme de textes éthiques non contraignants. Souvent formalisés au travers d'une liste de principes, ces documents mettent en avant les besoins et les axes de réponses des acteurs concernant la place de l'IA dans notre société. Au fil des années, le nombre de contributions s'est multiplié, produisant un corpus éthique de principes et de valeurs indispensable dans l'identification des enjeux liés à ces technologies d'IA. De façon plus ou moins intentionnelle, ces travaux éthiques structurent la réflexion autour de notions centrales dans l'élaboration d'un encadrement juridique de l'IA (**Sous-section 1**). Cependant, les textes éthiques produits contiennent des limites et des lacunes importantes, autant sur le fond que sur la forme. Par conséquent, le législateur qui souhaite s'inspirer de ces principes doit rester vigilant afin de corriger au mieux ces écueils, sous peine de les retrouver par la suite dans le cadre juridique de l'IA (**Sous-section 2**).

Sous-section 1. L'apport d'un consensus de principes de l'IA

Introduction. Malgré le pluralisme des auteurs des textes éthiques, certains principes apparaissent de façon répétée dans les différents documents. La fréquence d'apparition de certains principes a progressivement augmenté ces dernières années à tel point que l'on peut parler de convergence sur la présence indispensable de ces derniers¹³⁷. Au fur et à mesure, cette convergence donne lieu à un consensus normatif. Ce consensus sur des principes éthiques de l'IA contribue à façonner l'environnement normatif dans lequel ces technologies évoluent. Pour reprendre les termes de la CNIL dans son rapport de 2017, la norme éthique préfigure la norme juridique¹³⁸. En fonction des études, le nombre exact de principes considérés comme récurrents

¹³⁷ FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, « Principled Artificial Intelligence », 2020.

¹³⁸ COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Comment permettre à l'Homme de garder la main ? Les enjeux éthiques de l'intelligence artificielle. », CNIL, 2017.

varie de cinq principes¹³⁹ à plus d'une dizaine¹⁴⁰. Tenter de décrire exhaustivement l'ensemble des principes énoncés par la totalité des textes techniques n'est pas pertinent pour plusieurs raisons. D'une part, les différences de méthodologie et d'appellations des principes complexifient grandement ce travail sans apporter de valeur ajoutée au propos. Par exemple, certains textes vont unir les notions de responsabilité *ex post* et de responsabilité *ex ante* sous un seul principe tandis que d'autres chartes vont les énoncer de façon distincte. D'autre part, les principes éthiques sont abordés ici pour éclairer une analyse à venir sur l'encadrement des systèmes d'IA, et non sur le droit de l'IA en général. Par conséquent, tous les principes éthiques ne sont pas pertinents à traiter car ils ne pourront pas servir cet objectif. Ainsi les principes renvoyant exclusivement au contexte de déploiement desdits systèmes ou à leur utilisation ne seront pas traités¹⁴¹. Enfin, l'énonciation des principes ci-après s'inspire en grande partie des catégories de principes élaborées par J. FJELD et al. dans leur étude de 2020¹⁴².

Annonce de plan. Les principes éthiques au sein de ce consensus seront décomposés en deux catégories. Les « principes-objectifs » et les « principes-outils ». Les principes-objectifs sont des principes qui doivent être respecté en tant que tel : ils représentent leur propre finalité (§1). Les principes-outils, quant à eux, sont des principes qui doivent être respectés mais dont l'intérêt et l'importance qui leur sont accordés s'expliquent principalement en ce qu'ils permettent dans un second temps d'atteindre les principes-objectifs (§2). Ces deux catégories ne sont pas hermétiques et s'articulent entre elles. Certains principes-objectifs peuvent permettre d'améliorer le respect d'autres principes éthiques tandis que certains principes-outils sont parfois aussi considérés comme des finalités en tant que tels. Il s'agira alors de distinguer la fonction dominante accordée à chacun des principes éthiques du consensus.

¹³⁹ COMMISSION EUROPÉENNE POUR L'EFFICACITÉ DE LA JUSTICE (CEPEJ), « Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement », Conseil de l'Europe, 2018.

¹⁴⁰ FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, *op. cit.* JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019.

¹⁴¹ À titre d'exemple, le principe de « promotion des valeurs humaines », bien qu'essentiel dans la réflexion éthique globale sur l'IA ne sera pas traité car il se concentre exclusivement sur des enjeux post-déploiement du système d'IA.

¹⁴² FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, *op. cit.*

§1. Les principes-objectifs

Annonce de plan. Les principes-objectifs discutés ici sont les exigences primordiales qui sous-tendent l'élaboration des systèmes d'IA au sein des textes éthiques. Ainsi, il s'agit des principes les plus discutés au sein du corpus éthique. Indépendamment de l'identité de l'auteur ou de son secteur d'activité, les principes de *fairness* **(A)**, de vie privée **(B)** et de fiabilité technique **(C)** sont omniprésents dans la liste de principes des textes éthiques.

A– Le principe de fairness

Les besoins derrière le principe de *fairness*. La présence permanente du principe de *fairness* au sein des textes éthiques s'explique en grande partie par ses nombreuses violations médiatisées. Que ce soit dans les ressources humaines¹⁴³¹⁴⁴, la publicité¹⁴⁵, l'allocation de ressources publiques¹⁴⁶ ou encore la justice¹⁴⁷, en tant qu'aide à la décision ou seul décideur, les exemples de manquements au principe de *fairness* par des systèmes d'IA sont légion. La pluralité des secteurs touchés a conduit à établir la *fairness* comme un principe général à respecter dans tout contexte d'utilisation d'un système d'IA, dès lors que des êtres humains peuvent être affectés. Le principe de *fairness* peut être poursuivi à deux niveaux différents. En premier lieu, la *fairness* peut être respectée au niveau du logiciel. L'évaluation se concentre alors sur son fonctionnement et les résultats qu'il produit. En second lieu, le principe est envisagé à un niveau plus global. Il s'agit d'analyser la *fairness* dans le recours aux technologies d'IA, pris collectivement et non logiciel par logiciel. En effet, il s'avère que les systèmes d'IA sont utilisés majoritairement à l'encontre des mêmes catégories de population, celles-ci subissant la plupart des conséquences négatives de l'IA sans pour autant bénéficier des

¹⁴³ RAGHAVAN Manish, BAROCAS Solon, KLEINBERG Jon *et al.*, Mitigating bias in algorithmic hiring: evaluating claims and practices, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2020.

¹⁴⁴ SÁNCHEZ-MONEDERO Javier, DENCİK Lina et EDWARDS Lilian, What does it mean to « solve » the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2020.

¹⁴⁵ Amit Datta, Michael Tschantz & Anupam Datta, « Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination » (2014).

¹⁴⁶ « Pays-Bas. Scandale des allocations familiales », sur *Amnesty International*, publié le 25 octobre 2021.

¹⁴⁷ L'exemple célèbre de COMPAS, un système d'IA prédisant le risque de récidive chez les individus qui s'est avéré défavoriser significativement les populations afro-américaines et hispaniques. JULIA ANGWIN Jeff Larson, *Machine Bias*, [<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>].

avantages¹⁴⁸. Selon le niveau choisi, la conception de l'exigence change drastiquement. Au regard de notre objectif de recherche, la *fairness* sera traité principalement au niveau de la phase de conception, à savoir dans le fonctionnement du système d'IA.

Explication du principe. La *fairness* est le second principe le plus présent dans l'ensemble des textes éthiques. Régulièrement associé avec les notions d'équité et/ou de non-discrimination, ces assimilations ne sont toutefois pas des synonymes. Le principe de *fairness* est d'ailleurs bien plus présent dans l'ensemble des textes éthiques que celui de « non-discrimination »¹⁴⁹. La notion de *fairness* pourrait se traduire en français par le concept de *juste*, à savoir ce qui n'est pas injuste. Néanmoins la signification du terme, autant en langue anglaise que sa traduction française, est encore sujet à de nombreux débats. Le terme existait avant son utilisation en éthique de l'IA et déjà son caractère polysémique posait problème. Dans un article dédié à la notion de *fairness* dans le règlement (UE) 2016/679 relatif à la protection des données¹⁵⁰ (RGPD ci-après), Malgieri constate que la *fairness* a plusieurs significations dans ce seul texte. Il distingue deux significations de la notion. La première « *semble référer à une approche substantielle, dont l'objectif est d'éviter les effets néfastes dans certaines circonstances, en particulier quand il faut trouver un équilibre entre des intérêts contraires*¹⁵¹ ». Mais la *fairness* possède aussi son pendant procédural, qui dépasse l'analyse du résultat produit et pour se concentrer sur le processus de prise de décision¹⁵². Sous cet aspect, d'autres solutions sont proposées par l'éthique comme la capacité à contester une décision prise par un système d'IA, d'obtenir un contrôle et/ou une correction de la décision par un humain, etc¹⁵³. La *fairness* est un principe éthique large qui ne trouve pas d'équivalent unique au niveau juridique et son utilisation en éthique de l'IA partage cette caractéristique.

La majorité des textes éthiques abordent la *fairness* et la non-discrimination au travers de la problématique des biais. Un biais est une déformation de la perception dans le raisonnement,

¹⁴⁸ O'NEIL Cathy, *Weapons of Math Destruction*, Crown, 2017.

¹⁴⁹ JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019.

¹⁵⁰ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

¹⁵¹ Gianclaudio Malgieri, *The concept of fairness in the GDPR: a linguistic and contextual interprétation*, Barcelona, Spain, Association for Computing Machinery, 2020.

¹⁵² HARRISON Galen, HANSON Julia, JACINTO Christine et al., An empirical study on the perceived fairness of realistic, imperfect machine learning models, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2020.

¹⁵³ GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019.

celui d'un être humain aussi bien que celui d'un système d'IA. Pour le second cas on parle alors de biais algorithmiques. Il s'agit du cas où le système algorithmique va analyser incorrectement certaines données de façon systématique. En intégrant de façon consciente ou non des biais dans l'algorithme, les concepteurs ne se contentent pas de reproduire les biais préexistants¹⁵⁴ mais peuvent aussi aggraver ces biais en termes d'échelle et/ou de vitesse de propagation¹⁵⁵. Les textes éthiques énoncent donc la problématique sous la forme d'un appel à la prévention contre ces biais. L'objectif de la *fairness*, pris ici au niveau du système d'IA, renvoie donc à la réduction, voire la disparition des biais injustes présents dans le fonctionnement ainsi qu'à la réparation et la limitation des résultats injustes produits.

B- Le principe de vie privée

Les besoins derrière le principe de vie privée. Les technologies d'IA possèdent une relation de longue date avec la vie privée. Au stade de la conception des systèmes d'IA, c'est par le prisme du droit des données à caractère personnel que les technologies d'IA interagissent avec le principe de vie privée¹⁵⁶. Avant la démocratisation de l'IA, le traitement de données personnelles par un système algorithmique classique faisait déjà l'objet de préoccupations. Parmi les neuf critères permettant d'identifier la nécessité de réaliser une analyse d'impact au sens du RGPD, on retrouve « l'évaluation ou *scoring* », la « prise d'une décision automatique avec effet légal ou similaire » et le « croisement de données ». Ces trois critères peuvent être réunis lors de l'utilisation d'un algorithme, mais ils le seront d'autant plus lorsqu'il s'agit d'un logiciel recourant à l'IA.

De plus, les technologies d'IA, en particulier les méthodes d'apprentissage automatique, ont mis en lumière une nouvelle utilisation des données. Là où les algorithmes dits « traditionnels » exploitaient les données lors de leur utilisation (en tant que données d'entrées, aussi appelés *inputs*), les logiciels d'IA ont recours à des données également lors de leur conception. Cette fonction distincte des données est prise en compte par le droit des données personnelles. En

¹⁵⁴ NISSENBAUM Helen, « Accountability in a computerized society », *Sci Eng Ethics*, 2, 1996.

¹⁵⁵ VILLANI Cédric, « Donner un sens à l'intelligence artificielle : Pour une stratégie nationale et européenne », 2018. P121-122

¹⁵⁶ Au stade de l'utilisation du logiciel d'IA, le lien avec la vie privée se situe dans les risques encourus. La capacité de profilage accrue, le développement d'outils comme la reconnaissance faciale offrent aux utilisateurs de ces logiciels représentent de nombreuses possibilités de porter atteinte, proportionnée ou non, à la vie privée des individus. Bien qu'important dans les textes éthiques, cet aspect ne sera pas traité en détails ici car il est moins pertinent pour l'analyse future sur le droit des systèmes d'IA.

effet, la commission nationale de l'informatique et des libertés (CNIL ci-après) reconnaît que le traitement de données à caractère personnel lors de la conception et lors de l'utilisation poursuit deux finalités différentes au sens de la réglementation¹⁵⁷. Ensuite, cette nouvelle utilisation des données personnelles s'accompagne d'une volonté de la part des concepteurs d'exploiter des quantités massives de données. Désormais c'est la quantité de données que l'on peut obtenir qui prime¹⁵⁸, la qualité de la donnée devient, de façon critiquable, une exigence secondaire. Ce besoin de données amplifie des habitudes déjà existantes de collecter le maximum de données¹⁵⁹, sans réflexion préalable sur l'objectif poursuivi¹⁶⁰, entrant alors en contradiction claire avec le principe de minimisation¹⁶¹ imposé par la législation¹⁶².

Enfin, les systèmes d'IA présentent de nouveaux risques pour la sécurité des données¹⁶³. Les technologies d'apprentissage produisent un modèle grâce aux données d'entraînement. De ce fait, le modèle est le résultat d'un traitement de nombreuses données personnelles. Si le modèle n'est pas considéré aujourd'hui en tant que tel comme une donnée personnelle, les cyberattaques qu'il subit peuvent constituer une violation de données¹⁶⁴ au sens de la législation européenne et française¹⁶⁵.

Explication du principe. Malgré la présence très importante du principe dans les textes éthiques, celui-ci est très rarement défini¹⁶⁶. Les auteurs des textes éthiques traitent la vie privée sous deux aspects : en tant que droit et en tant que valeur éthique. Malgré ce lien fort avec le

¹⁵⁷ « Du point de vue de la protection des données, ces deux étapes ne remplissent pas le même objectif et doivent donc être séparées. » - « IA ».

¹⁵⁸ ZUBOFF Shoshana, « Big other », *J Inf Technol*, 30, 2015. ZUBOFF Shoshana, *L'Âge du capitalisme de surveillance*, Zulma, 2020.

¹⁵⁹ INFORMATION COMMISSIONER'S OFFICE, « Big data, artificial intelligence, machine learning and data protection », Information Commissioner's Office (ICO), 2017.

¹⁶⁰ BENSAMOUN Alexandra et LOISEAU Grégoire, *Droit de l'intelligence artificielle*, 2019. A. DEBET, Chapitre 10 « Intelligence artificielle et droit des données à caractère personnel ».

¹⁶¹ « Les données à caractère personnel doivent être : [...] adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées » - « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 5.c).

¹⁶² DELTORN Jean-Marc, « La protection des données personnelles face aux algorithmes prédictifs | Revue des droits et libertés fondamentaux », *Revue des Droits et Libertés Fondamentales*, n° 12, 2017.

¹⁶³ SERVICE DE RECHERCHE DU PARLEMENT EUROPÉEN, « A governance framework for algorithmic accountability and transparency - Think Tank », Parlement Européen, 2019.

¹⁶⁴ « Une violation de la sécurité entraînant, de manière accidentelle ou illicite, la destruction, la perte, l'altération, la divulgation non autorisée de données à caractère personnel transmises, conservées ou traitées d'une autre manière, ou l'accès non autorisé à de telles données ; » - RGPD, article 4.12)

¹⁶⁵ « Si un modèle d'IA fait l'objet d'une attaque en confidentialité réussie (par inférence d'appartenance, exfiltration ou inversion par exemple), cela peut constituer une violation de données. » - « IA ».

¹⁶⁶ JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019.

droit, il serait donc incorrect de résumer le principe de vie privée au droit à la vie privée, quel que soit le cadre juridique dans lequel on se place. La vie privée est parfois associée à la notion d'intimité. Il peut alors s'agir de « l'intimité de la pensée et des émotions »¹⁶⁷ qui représente un espace d'intimité qui doit être protégé de l'utilisation de systèmes d'IA¹⁶⁸. La charte de Montréal propose même la création « d'espaces d'intimité » physique, expression utilisée en opposition à la menace d'une surveillance totale et permanente des individus. Pour la plupart des auteurs, le respect de la vie privée implique de permettre aux personnes concernées d'avoir un contrôle sur leurs données exploitées par les systèmes d'IA. Cette influence sur ces données peut être indirecte. Ainsi dans ses lignes directrices, le groupe d'experts de haut niveau de la commission européenne sur l'IA proposent la création d'outils et d'institutions pour garantir ce contrôle sur les données. D'autres textes abordent le contrôle direct de l'individu en demandant l'élaboration de droit ou l'élargissement de droits existants pour la personne concernée. Par exemple, les principes éthiques élaborés par *Access Now*¹⁶⁹ proposent comme solution un élargissement des circonstances de recours au droit à l'effacement inscrit dans le RGPD.

C– Le principe de fiabilité technique

L'intelligence artificielle est avant tout un ensemble d'approches techniques. De par cette nature, les outils qui ont recours à des technologies d'IA sont considérés comme des produits technologiques. Un produit possède une finalité d'utilisation et sa capacité à accomplir cette finalité détermine son efficacité. Pour cela il doit posséder un fonctionnement digne de confiance du point de vue de cet objectif de départ. La fiabilité est un objectif pleinement intégré par les concepteurs aux logiciels d'IA. Ce principe-objectif doit être atteint avant toute exploitation du système¹⁷⁰. En l'absence de fiabilité, la capacité des systèmes à apporter les bénéfices pour les citoyens et, plus généralement, leur utilité est grandement diminuée. Le principe de fiabilité regroupe deux sous-principes éthiques qui retranscrivent cette vision technique des systèmes d'IA : la sûreté et la sécurité.

¹⁶⁷ « Déclaration de Montréal pour un développement responsable de l'IA », 2018.

¹⁶⁸ Les autrices Judith Rochfeld et Célia Zolynski discutent ainsi de l'idée de « droit à l'intégrité psychique » et d'« autodétermination cognitive » : ROCHFELD Judith et ZOLYNSKI Célia, « La valeur des émotions : Quel régime pour le "capitalisme mental" ? », dans *Entre art et technique : les dynamiques du droit. Mélanges en l'honneur de Pierre Sirinelli*, Dalloz, 2022.

¹⁶⁹ « Human Rights in the Age of Artificial Intelligence », Access Now, 2018.

¹⁷⁰ GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019. P.6

La sûreté. La sûreté est l'exigence d'un fonctionnement correct du système d'IA et sa capacité à ne pas produire des comportements imprévus envers l'environnement du système, notamment les personnes affectées¹⁷¹. La sûreté est fortement liée à la notion d'exactitude. L'exactitude est une fonction permettant au système de produire un résultat correct. Une exactitude faible implique un taux d'erreur important et donc la fourniture de résultats inadéquats, pouvant causer un préjudice à la personne affectée¹⁷². Le taux d'erreur doit bien évidemment être le plus petit possible et cela dans des situations les plus diverses possibles. L'exactitude doit demeurer constante malgré la diversité des données d'entrées que le logiciel peut recevoir, on parle alors de robustesse¹⁷³. La sûreté dépend donc aussi de l'utilisation du système dans des conditions correctes, c'est-à-dire dans les conditions prévues par les concepteurs du système. Toutefois, il n'est pas possible pour les concepteurs de prévoir toutes les utilisations imaginables des futurs utilisateurs. Une utilisation improbable par ces derniers peut produire une erreur et donc des dommages à la personne affectée. Le respect du principe de fiabilité implique la réalisation de mesures lors du développement par les concepteurs mais aussi lors de l'exécution par les déployeurs du système. Ainsi la documentation du taux d'erreur et des situations d'utilisations prévues lors de la conception est utile pour garantir le respect de ce principe. Il est aussi possible pour le concepteur d'évaluer les risques mais cela peut s'avérer difficile car le contexte joue un rôle important dans la survenance de ces dysfonctionnements. Toutefois, les textes s'accordent pour dire que ce n'est pas une raison pour confier la totalité de la charge de cette exigence au déployeur, les développeurs ayant aussi leur rôle à jour dans le respect de ce principe.

La sécurité. Le second aspect de la fiabilité est la sécurité. Elle désigne la capacité du système à se protéger contre les menaces extérieures¹⁷⁴. L'atteinte de ce principe-objectif impose de tester la résilience du système, c'est-à-dire sa capacité à continuer à fonctionner correctement

¹⁷¹ *Ibid.* P. 17.

¹⁷² Le terme désigne la ou les personnes physiques qui font l'objet d'un résultat produit pour un système ou qui sont affectées par un tel résultat. Le terme et la définition sont issues de l'article 3.8 de la version du règlement sur l'IA proposée par le Parlement européen. La notion et la définition ont été retirées dans la version finale du texte.

¹⁷³ « *Le degré auquel un système ou un composant logiciel peut fonctionner correctement en présence d'entrées non valides ou de conditions environnementales difficiles.* » - FOOD AND DRUG ADMINISTRATION, « Glossary of Computer System Software Development Terminology (8/95) », Office of regulatory affairs - FDA, 2019. Le terme de « robustesse » comprend parfois également un aspect relatif à la sécurité du logiciel : « *La robustesse technique passe par la mise au point de systèmes d'IA selon une approche de prévention des risques, et de telle manière que ces systèmes se comportent, de manière fiable, conformément aux attentes, tout en réduisant le plus possible les atteintes involontaires et inattendues, et en empêchant toute atteinte inacceptable.* » - GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, *op. cit.* P. 20.

¹⁷⁴ FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, « Principled Artificial Intelligence », 2020.

en dépit des dommages potentiels subis. Présent notamment dans la directive SRI 2¹⁷⁵ ou encore le règlement 2022/2554 sur la résilience opérationnelle numérique du secteur bancaire¹⁷⁶, l'objectif de résilience est désormais récurrent au sein des réglementations européennes relatives aux systèmes informatiques. Comme la sûreté, la sécurité ne se limite pas à des solutions techniques. En effet elle impose l'élaboration et la mise en œuvre de plan de secours lors de la survenance d'une attaque. Les attaques peuvent tout aussi bien être dirigées vers le système que vers les infrastructures ou les données (l'empoisonnement de données par exemple)¹⁷⁷. Ces différents éléments peuvent alors faire l'objet de tests réguliers afin de vérifier l'efficacité des mesures mises en place. Toutefois, la sécurité ne se limite pas une logique d'évitement des risques d'attaques, elle comprend également un volet relatif à la réaction après la survenance d'une attaque. Ainsi la communication d'informations relatives aux attaques subies et aux vulnérabilités identifiées est aussi une condition essentielle d'une sécurité satisfaisante¹⁷⁸. Enfin, la prise de mesures pour limiter les conséquences en cas de faille de sécurité est aussi essentielle. Le maintien de la confidentialité, de l'intégrité et de la disponibilité des données et du système se prépare dès la conception. Cette idée de limitation des dommages potentiels après une violation de sécurité n'est pas nouvelle et se retrouve notamment dans le RGPD sous la forme d'une exigence de mesures d'atténuation¹⁷⁹.

§2. Les principes-outils

Annnonce de plan. Les principes-outils demeurent des principes qui doivent être respectés selon les textes éthiques. Toutefois, ils sont généralement utilisés comme des moyens de surveiller, prévenir ou sanctionner le respect des principes-objectifs. Même s'ils peuvent parfois représenter un objectif en tant que tel, celui-ci laisse la majorité du temps sa place au profit de la réalisation des principes-objectifs présentés précédemment. À la différence des principes-

¹⁷⁵ « Directive (UE) 2022/2555 du Parlement européen et du Conseil du 14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement (UE) no 910/2014 et la directive (UE) 2018/1972, et abrogeant la directive (UE) 2016/1148 (directive SRI 2) (Texte présentant de l'intérêt pour l'EEE) », 2022. Le texte emploie le terme de « cyberrésilience ».

¹⁷⁶ « Règlement (UE) 2022/2554 du Parlement européen et du Conseil du 14 décembre 2022 sur la résilience opérationnelle numérique du secteur financier et modifiant les règlements (CE) no 1060/2009, (UE) no 648/2012, (UE) no 600/2014, (UE) no 909/2014 et (UE) 2016/1011 (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022.

¹⁷⁷ GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019. P. 20

¹⁷⁸ « Déclaration de Montréal pour un développement responsable de l'IA », 2018. P15

¹⁷⁹ Article 33.2.d) du RGPD

objectifs, le respect des principes-outils est donc perçu principalement au travers de leur capacité à satisfaire d'autres principes-objectifs. Les trois principes-outils éthiques les plus récurrents sont le principe de transparence **(A)**, le principe de responsabilité **(B)** et le principe de contrôle humain **(C)**.

A- Le principe de transparence

Les besoins derrière le principe de transparence. La transparence est le principe le plus partagé par les textes éthiques. La transparence est une préoccupation commune à tous les acteurs, aussi bien par les États que les entreprises et même la société civile. L'omniprésence de ce principe-outil peut s'expliquer par différentes raisons.

Tout d'abord, la transparence est une réponse directe à une spécificité bien connue de l'IA : son opacité¹⁸⁰. La représentation du système d'IA comme une « boîte noire » a ancré le problème de l'opacité, souvent avéré, dans l'esprit de tous¹⁸¹. Mais la réalité technique est plus complexe que qu'une simple métaphore. Les technologies d'IA présentent trois niveaux d'opacité¹⁸². Le premier niveau est celui du secret. L'opacité par le secret représente le cas où l'existence même de l'information est inconnue pour toute personne située hors du secret. Dans le cas de l'IA, c'est l'exemple d'une personne interagissant ou étant sujette à une décision qui ignore totalement la présence et/ou l'utilisation d'un système d'IA dans la situation où elle se trouve. Ce niveau d'opacité n'est pas exclusif aux technologies d'IA, il est bien connu du droit notamment avec le principe du secret des affaires¹⁸³ ou encore le secret professionnel et ses différentes déclinaisons¹⁸⁴. Pour limiter ce niveau d'opacité, le droit a mis en place différentes

¹⁸⁰ PASQUALE Frank, *The Black Box Society – The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015.

¹⁸¹ Sur l'impact des métaphores dans la perception juridique des nouvelles technologies : Ryan Calo, A. Froomkin & Ian Kerr, *Robot Law*, 2016.

¹⁸² BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016.

¹⁸³ « Directive (UE) 2016/943 du Parlement européen et du Conseil du 8 juin 2016 sur la protection des savoir-faire et des informations commerciales non divulgués (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. En droit français, ce principe est énoncé au titre V du livre 1er du code de commerce (article L. 151 et suivants) depuis la transposition de la directive européenne par la « LOI n° 2018-670 du 30 juillet 2018 relative à la protection du secret des affaires (1) », 2018-670, 2018. Dans le droit de l'Union, le secret des affaires

¹⁸⁴ Par exemple, le secret professionnel médical, le secret professionnel des avocats, etc.

solutions telles que l'exception législative¹⁸⁵ ou la limitation de l'accès à l'information par le seul juge¹⁸⁶. Le deuxième niveau d'opacité est celui de la complexité technique. L'information est alors accessible, ce n'est pas un secret, mais celle-ci n'est pas compréhensible par tout le monde. Une barrière de compétence est érigée entre l'information et la connaissance qu'elle contient. Encore une fois, ce type d'opacité par la complexité est déjà connu et se trouve fréquemment dans les domaines scientifiques, l'informatique, la médecine, etc. Le troisième et dernier niveau d'opacité est l'opacité par la conception. L'information est alors intrinsèquement opaque, inintelligible dès son apparition. Cette couche d'opacité est propre aux technologies d'IA et représente les situations où même les concepteurs du système ne peuvent pas savoir exactement ce qu'il se passe dans le modèle. Ces trois couches d'opacité sont parfaitement cumulables et il est indispensable de les prendre toutes en compte sans quoi l'opacité est maintenue. Face à cette opacité sans précédent, la transparence a donc été érigée en principe éthique central pour lutter contre cette opacité sans précédent¹⁸⁷.

Ensuite, la transparence est une notion qui revêt une double nature. Il peut s'agir d'un objectif, à savoir qu'il y a une exigence propre de transparence pour les systèmes d'IA. Cette exigence est vue majoritairement comme binaire, c'est-à-dire un seuil unique au-delà duquel le système est qualifié de transparent, mais certains acteurs perçoivent plutôt la transparence comme une échelle composée de différents niveaux. Mais la transparence est aussi vue comme un moyen permettant de satisfaire, ou du moins contribuer à satisfaire, le respect d'autres principes éthiques de l'IA. Par exemple, la transparence est un outil intéressant pour permettre le contrôle humain sur le système d'IA, la prévention contre les biais algorithmiques, la redevabilité des acteurs¹⁸⁸ ou encore la légitimité des décisions dans le secteur public¹⁸⁹. La question de la transparence possède donc son existence propre mais apparaît aussi être une condition pour

¹⁸⁵ Par exemple, le code du commerce français prévoit trois exceptions permettant la divulgation d'informations soumis au secret des affaires, notamment lorsque l'atteinte au secret a été réalisée « *pour exercer le droit à la liberté d'expression et de communication, y compris le respect de la liberté de la presse, et à la liberté d'information telle que proclamée dans la Charte des droits fondamentaux de l'Union européenne ;* » - Article L. 151-8. Pour une autre exception, voir plus généralement le régime des lanceurs d'alerte, dont la directive européenne 2019/1937 : « Directive (UE) 2019/1937 du Parlement européen et du Conseil du 23 octobre 2019 sur la protection des personnes qui signalent des violations du droit de l'Union », OJ L, 2019.

¹⁸⁶ « *Le juge peut, d'office ou à la demande d'une partie ou d'un tiers, si la protection de ce secret ne peut être assurée autrement et sans préjudice de l'exercice des droits de la défense : 1° Prendre connaissance seul de cette pièce et, s'il l'estime nécessaire, ordonner une expertise et solliciter l'avis, pour chacune des parties, d'une personne habilitée à l'assister ou la représenter, afin de décider s'il y a lieu d'appliquer des mesures de protection prévues au présent article ; [...]* » - Article L. 153-1.1° du code de commerce.

¹⁸⁷ BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016. P. 6

¹⁸⁸ DE LAAT Paul B., « Algorithmic Decision-Making Based on Machine Learning from Big Data », *Philos. Technol.*, 31, 2018.

¹⁸⁹ CHESTERMAN Simon, « Through a Glass, Darkly », 2020.

satisfaire le respect d'autres principes. Toutefois, le principe de transparence entre régulièrement en conflit avec d'autres principes éthiques ou des dispositions présentes en droit positif¹⁹⁰. La transparence pour la transparence est devenue une vision minoritaire du fait des risques qu'elle crée. Pour certains, la transparence du raisonnement d'un système peut faciliter les tentatives de contournement par les personnes affectées¹⁹¹. D'autres auteurs en revanche pensent que la transparence des « règles du jeu » peut s'avérer bénéfique dans certaines situations, en poussant les personnes affectées à améliorer leur comportement dans une certaine direction¹⁹². Pas toujours perçu comme nécessaire¹⁹³, le respect du principe doit par conséquent être mesuré, proportionné.

Explication du principe de transparence. Appliquée à la phase de conception, le principe éthique de la transparence renvoie à l'idée que les systèmes d'IA doivent être conçus de manière à ne pas être opaques. La conception du logiciel doit rendre possible la surveillance des actions par un être humain. Il s'agit notamment de garder disponible et/ou communiquer des documents relatifs au système en question. Le principe de transparence s'étend également durant tout le cycle de vie du logiciel. De ce fait, la transparence s'adresse à de nombreux acteurs : le concepteur, l'utilisateur, l'autorité de contrôle le cas échéant et/ou les personnes affectées. Le principe de transparence est souvent adjoint au principe d'explicabilité, ce-dernier possédant ainsi une fréquence d'apparition très proche de celle de la transparence au sein du corpus de textes éthiques¹⁹⁴. Bien que la définition de l'explicabilité varie selon les textes, celle-ci peut se définir comme la « *traduction de concepts techniques et des résultats produits en format intelligible et compréhensible pour être appréciés* »¹⁹⁵. Le groupe d'experts de haut niveau inscrit également l'exigence de traçabilité comme une composante du principe de transparence¹⁹⁶. L'idée est de produire des systèmes au sein desquels il est possible de retracer

¹⁹⁰ À titre d'exemple, le cas de la propriété intellectuelle : BENSAMOUN Alexandra et LOISEAU Grégoire, *Droit de l'intelligence artificielle*, 2019. Chapitre 9 : Intelligence artificielle et propriété intellectuelle. CASTETS-RENARD Céline, « The Intersection Between AI and IP », *IIC*, 51, 2020.

¹⁹¹ KROLL Joshua A., HUEY Joanna, BAROCAS Solon *et al.*, « Accountable Algorithms », *University of Pennsylvania Law Review*, 165, 2017. P.1. À titre d'exemple : KHAZAN Olga, How to Fake Your Workout, [<https://www.theatlantic.com/health/archive/2015/09/unfit-bits/407644/>].

¹⁹² « *Knowing which features count can only induce those involved to do better about them. Hotel owners start cleaning their rooms more thoroughly, restaurant ownerstake care to serve better food [...]* ». DE LAAT Paul B., « Algorithmic Decision-Making Based on Machine Learning from Big Data », *Philos. Technol.*, 31, 2018.

¹⁹³ KROLL Joshua A., HUEY Joanna, BAROCAS Solon *et al.*, *op. cit.*

¹⁹⁴ JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019.

¹⁹⁵ FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, « Principled Artificial Intelligence », 2020.

¹⁹⁶ GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019. PP. 22 et 37.

le déroulé des actions ayant conduit à la production d'un résultat spécifique. L'objectif sous-jacent est notamment de pouvoir identifier la cause d'une décision algorithmique qui aurait généré un dommage. La transparence est alors un moyen de faciliter le respect du principe de responsabilité.

B- Le principe de responsabilité

La problématique de la responsabilité dans le cadre de l'IA se divise généralement en deux aspects. Le premier aspect est celui représenté par le terme de « redevabilité » (*accountability* en anglais) qui désigne la capacité du débiteur d'une ou plusieurs exigences à rendre des comptes pour leurs actions. Le second aspect est la responsabilité *ex post*, plus exclusif au droit. Il s'agit alors de l'attribution de la responsabilité et des sanctions associées en cas de préjudice.

La redevabilité. La redevabilité est une problématique de long terme du numérique. Déjà en 1996, Helen Nissenbaum¹⁹⁷ mettait en garde de la dégradation de la redevabilité dans une société du numérique¹⁹⁸. Dans ce contexte, Nissenbaum liste quatre obstacles, toujours présents vingt-cinq ans plus tard. Premièrement la multiplicité des acteurs impliqués dans la conception d'un système informatique dilue la responsabilité entre eux¹⁹⁹. Cette caractéristique est exacerbée par les systèmes d'IA qui impliquent plus d'éléments (les différentes données utilisées par exemple), chacun exploité par différents acteurs, situés dans différentes zones du monde²⁰⁰. Le deuxième obstacle est la conséquence de la présence constante de défauts (*bugs*) informatique. Parce que ces dysfonctionnements ne sont pas et ne seront jamais entièrement évitables, les professionnels du domaine sont moins enclins et incités à les identifier ainsi que chercher et appliquer des corrections à ces défauts²⁰¹. La présence de ces dysfonctionnements informatiques offre également une excuse idéale pour blâmer l'ordinateur plutôt que l'individu en cas de dommage, un réflexe bien intégré qui entraîne les concepteurs à généralement refuser de rendre tout compte pour les systèmes qu'ils développent²⁰². Enfin, l'opacité intrinsèque des

¹⁹⁷ NISSENBAUM Helen, « Accountability in a computerized society », *Sci Eng Ethics*, 2, 1996.

¹⁹⁸ Traduction française de l'auteur de l'expression « *computerized society* ».

¹⁹⁹ NISSENBAUM Helen, *op. cit.*

²⁰⁰ SCHERER Matthew U., « Regulating Artificial Intelligence Systems », *Harvard Journal of Law & Technology*, 29, 2015.

²⁰¹ NISSENBAUM Helen, *op. cit.*

²⁰² *Ibid.*

technologies d'IA accroît encore plus cet affaiblissement de la redevabilité des concepteurs²⁰³. La redevabilité est généralement associée à l'idée de responsabilité *ex ante*, à savoir exiger la mise en place de mécanismes afin de garantir une bonne conception du système d'IA. L'acteur doit s'organiser pour évaluer les risques, décider ensuite si la situation est conforme ou non aux exigences et enfin documenter ce processus en cas de futur contrôle. Une illustration traditionnelle de ce principe est l'analyse d'impacts²⁰⁴. Elle peut être réalisée aussi bien pour des risques techniques, des risques pour les droits de l'Homme que pour des risques plus globaux et systémiques. Outre l'analyse d'impact, le recours aux audits est aussi mentionné dans la moitié des textes éthiques²⁰⁵. Le principe n'a d'intérêt qu'au regard de l'objectif de conformité poursuivi. Si l'obligation à respecter est insuffisante, le principe de responsabilité *ex ante* ne sera pas en capacité de pallier cette limite.

La responsabilité *ex post*. Le second aspect du principe éthique de responsabilité est la responsabilité *ex post* qui renvoie généralement à l'idée d'indemnisation et de réparation du préjudice causé. Les nombreux risques liés aux technologies d'IA, leur opacité et la multitude d'acteurs intervenant dans le cycle de vie d'un système d'IA sont autant de facteurs qui ont entraîné une réflexion sur la problématique de la responsabilité des acteurs. L'incertitude actuelle sur la responsabilité *ex post* des différents acteurs conduit à deux conséquences. Premièrement, elle établit un climat de méfiance néfaste pour l'investissement dans le secteur d'activité²⁰⁶. Deuxièmement, cette incertitude questionne la possibilité pour un client ayant acheté un système d'IA ou une personne affectée par une décision issue d'un système d'AI d'obtenir réparation pour le préjudice subi. L'interrogation est directement liée à la question de l'acceptabilité sociale des technologies d'IA. La capacité à être indemnisé joue sur la confiance dans l'utilisation de ces technologies et donc sur la compétitivité de l'Union européenne. Au-delà de la recherche de l'auteur d'un dommage²⁰⁷, il est aussi question de l'adéquation des cadres juridiques actuels de la responsabilité face aux nouveautés de l'IA. Le droit positif regroupe plusieurs régimes de responsabilité déjà bien en place. Le principe de responsabilité *ex post* renvoie au questionnement sur le régime de responsabilité adapté pour l'IA, à la répartition de la responsabilité entre les acteurs et aussi de la possibilité d'élaboration

²⁰³ DE LAAT Paul B., « Algorithmic Decision-Making Based on Machine Learning from Big Data », *Philos. Technol.*, 31, 2018.

²⁰⁴ *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, par Dillon Reisman et al, 2018.

²⁰⁵ FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, « Principled Artificial Intelligence », 2020.

²⁰⁶ COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021.

²⁰⁷ COMMISSION DES AFFAIRES JURIDIQUES, « Rapport contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle », 2020.

d'un régime de responsabilité sur mesure. À titre d'exemple, le parlement européen a produit une résolution proposant une responsabilité sans faute pour les systèmes d'IA à haut risque et une responsabilité pour faute pour les autres systèmes d'IA²⁰⁸. Aujourd'hui la doctrine européenne s'accorde pour un régime inspiré de la logique de la responsabilité du fait des produits défectueux²⁰⁹. Toutefois l'aspect *ex post* du principe de responsabilité professionnelle ne se limite pas à la responsabilité en cas de dommage par les systèmes d'IA. Le principe éthique pose aussi la question de la responsabilité d'autres acteurs comme celle des lanceurs d'alerte²¹⁰ ou des équipes pédagogiques intervenant dans les formations en intelligence artificielle²¹¹.

C- Le principe de contrôle humain

Les besoins derrière le principe de contrôle humain. Une des spécificités des technologies d'IA est leur capacité à produire des résultats de façon autonome. Couplée à leur opacité, l'autonomie décisionnelle des systèmes fait émerger un risque spécifique aux technologies d'IA, leur imprévisibilité. En l'absence de compréhension du raisonnement de l'algorithme, il devient très difficile de surveiller le traitement qui aboutira aux résultats qui seront produits. De plus, si une même donnée est entrée à dix reprises, les logiciels recourant aux technologies d'IA sont capables de produire dix résultats différents. La reproductibilité du raisonnement n'est absolument pas garantie. Cette imprévisibilité contribue à exclure l'humain des traitements automatisés de données²¹² et pose donc immédiatement la question du contrôle, théorique et réel, de l'humain sur la machine. La forte présence du principe de contrôle humain dans les textes éthiques peut aussi s'expliquer par la peur de la perte d'autonomie humaine ou d'auto-détermination de l'individu face à la machine²¹³. Cette crainte est confirmée par l'étude des interactions entre l'humain et la machine qui révèle l'existence de comportements allant dans le sens de la perte d'autonomie. L'un des exemples les plus connus est le biais d'automatisation.

²⁰⁸ PARLEMENT EUROPÉEN, « Un régime de responsabilité civile pour l'intelligence artificielle », 2020.

²⁰⁹ MAZEAU Laurène, « Responsabilité. Quel(s) droit(s) pour les systèmes d'intelligence artificielle ? », *Cahiers Droit, Sciences & Technologies*, PUP, 2020.

²¹⁰ Rapport AI Now, 2018

²¹¹ VILLANI Cédric, « Donner un sens à l'intelligence artificielle : Pour une stratégie nationale et européenne », 2018.

²¹² Daniel Solove, « À Taxonomy of Privacy » (2006) 154:3 *University of Pennsylvania Law Review* 477.

²¹³ Résultat de l'enquête réalisée dans le cadre du débat public par la CFE-CGC, COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Comment permettre à l'Homme de garder la main ? Les enjeux éthiques de l'intelligence artificielle. », CNIL, 2017. P. 24

Celui-ci désigne la tendance de l'utilisateur humain à ignorer ou ne pas rechercher des informations contradictoires à la recommandation produite par le logiciel²¹⁴. Dès lors que le décideur humain se contente de suivre le résultat algorithmique, la limite entre l'aide à la décision et la prise de décision par un algorithme est franchie²¹⁵. Le contrôle humain vient garantir que l'utilisation de systèmes d'IA ne conduit en réalité ni à une perte d'autonomie, ni à une déresponsabilisation de l'humain²¹⁶.

L'explication du principe. Ce principe peut se décliner sous plusieurs aspects²¹⁷ correspondant aux différents sujets de ce contrôle : l'opérateur, à savoir l'individu qui utilise le système d'IA, la personne affectée, c'est-à-dire la personne faisant l'objet de la décision algorithmique. Premièrement, le contrôle humain par l'utilisateur. L'idée est de concevoir des systèmes qui permettent à un être humain d'intervenir dans le processus décisionnel. L'utilisateur humain doit comprendre le raisonnement du système dans une certaine mesure, afin d'avoir la capacité de produire une solution à la situation traitée au départ par le logiciel. Il doit avoir les compétences nécessaires pour maintenir une certaine indépendance vis-à-vis du système d'IA, afin de garantir que son intervention soit significative. Ensuite, la personne affectée. Il s'agit de lui garantir un contrôle sur la décision qu'il reçoit équivalent au cas où la décision aurait été prise par un être humain. Pour développer ce contrôle, plusieurs textes éthiques mentionnent la possibilité pour l'individu de refuser de faire l'objet d'une décision automatisée, garantissant une alternative humaine s'il le souhaite. Cette option prend notamment la forme d'*opt-out*. La personne affectée doit pouvoir également contester sa décision, voire en obtenir une nouvelle, prise par un être humain.

Conclusion de la sous-section. Sous la forme de principes, les auteurs des textes éthiques mettent en avant l'ensemble des besoins, et des problèmes, principaux concernant les technologies d'IA. Cette liste de principes consensuels offre au législateur une proposition instructive des éléments de fond à traiter en priorité au sein de la réglementation sur l'IA. Cependant, il est important de considérer ce consensus pour ce qu'il est, à savoir une

²¹⁴ CUMMINGS M. L., *Automation Bias in Intelligent Time Critical Decision Support Systems*, AIAA 3rd *Intelligent Systems Conference*, AIAA, 2004.

²¹⁵ À titre d'exemple, le groupe de travail de l'article 29, aujourd'hui le comité européen de la protection des données considère dans ses lignes directrices qu'une décision est prise de façon automatisée si le contrôle humain réalisé n'est pas significatif mais seulement symbolique. Contrôleur européen à la protection des données, « Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679 », 2018.

²¹⁶ COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, *op. cit.*

²¹⁷ Le principe de contrôle humain peut également être divisé selon la forme prise lors de sa mise en application : TSAMADOS Andreas, FLORIDI Luciano et TADDEO Mariarosaria, « Human control of AI systems », *AI Ethics*, 2024.

convergence de considérations générales et théoriques²¹⁸ sur l'éthique de l'IA. Une fois passées les grandes lignes de ces principes, le consensus entre les différents acteurs s'affaiblit et de nombreuses difficultés font leur apparition.

Sous-section 2. Les limites à l'apport des textes éthiques sur l'IA

Annonce de plan. L'émergence d'un ensemble relativement consensuel de principes éthiques pour l'ensemble des technologies et systèmes d'IA est le fruit d'un travail considérable de la part des acteurs pour répondre à des besoins liés à l'IA. Cet ensemble de notions et de principes présents dans les textes forge même une base de travail intéressante pour le droit. Aux premiers abords, l'application de ces principes par les entreprises semble facilitée par ce processus d'élaboration collectif. Pour certains acteurs, cette application ne nécessiterait même pas une intervention juridique contraignante. Toutefois, l'étude globale de ces textes éthiques fait émerger des limites non négligeables. La réutilisation de ces principes éthiques par le droit révèle donc des besoins spécifiques qu'il convient d'identifier pour compléter notre grille d'analyse du droit des systèmes d'IA. Ces limites, autant liées aux principes (§1) qu'à la nature des textes qui les contiennent (§2), réduisent l'apport immédiat des principes éthiques comme alternatives au droit et laissent ainsi de nombreux travaux à accomplir pour les législateurs.

§1. Les limites sur le fond des principes éthiques d'IA

Annonce de plan. Si ces six principes éthiques sont énoncés quasi systématiquement dans les textes, l'interprétation de chacune des notions se révèle beaucoup moins uniforme. Même si aucune d'elles n'a été inventée spécifiquement pour les technologies d'IA, il s'agit toutefois d'un secteur particulier qui élabore un vocabulaire spécifique. Des variations entre les auteurs sur la signification des principes apparaissent rapidement lorsque l'on entre dans les détails. Le consensus sur les principes devient alors vite un consensus de principe fondé sur un flou définitionnel problématique (A). De plus, les textes éthiques formulent leurs principes sous forme d'énumération. Les principes sont détaillés, le cas échéant, de façon isolée en envisageant

²¹⁸ « PRINCIPE : Définition de PRINCIPE », sur *Cnrtl.fr*.

trop peu leurs interactions. La réutilisation de ces principes éthiques, notamment au travers du droit, nécessite de comprendre les conflits pouvant émerger de ces articulations **(B)**.

A– Le manque de clarté des principes

Les facteurs de divergences sémantiques. Les divergences de signification pour un même principe peuvent s'expliquer par différents facteurs. Selon *Jobin et al.*²¹⁹, ils sont au nombre de quatre. Premièrement, l'interprétation des principes changent entre les différents textes, c'est-à-dire que les détails théoriques de la notion sont discutés. Deuxièmement, les raisons derrière l'établissement de ces notions en principes éthiques sont différentes. C'est parce que les différents auteurs ne brandissent pas le principe pour la même raison que celui-ci voit son contenu précis varier. Troisièmement, les principes sont perçus et interprétés différemment en fonction des destinataires visés et les secteurs concernés par les textes éthiques. En effet, le secteur public présente des intérêts et des besoins différents du secteur privé. Le secteur d'activité entre aussi en compte. Par exemple, le secteur hospitalier n'aura pas les mêmes besoins que le secteur de l'enseignement ou le secteur militaire. Chacun possède ses spécificités qui impactent nécessairement la conception des principes éthiques de l'IA, bien que les risques auxquels ils renvoient puissent être similaires entre certains secteurs. Enfin, la manière de penser la mise en place entraîne une divergence entre les textes éthiques. Ces divergences se cristallisent sur deux points : les circonstances d'application et les modalités d'applications des principes.

Les circonstances d'application du principe. Un élément de divergence d'interprétation entre les textes éthiques est le champ d'application des principes. Bien qu'un consensus éthique ressorte, les circonstances dans lesquelles un principe doit être respecté peuvent varier d'un auteur à l'autre. Dans le cas du principe de transparence, la divulgation d'informations, mécanisme central au respect du principe, n'est pas toujours exigée dans les mêmes situations. Pour certains textes, les informations doivent être communiquées lorsqu'un système d'IA produit une décision²²⁰. Si le système traite des données relatives à un individu sans pour autant produire de décisions ayant un impact sur sa vie, ce mécanisme de transparence n'a pas à être

²¹⁹ JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019.

²²⁰ Selon le règlement européen sur l'IA, un logiciel d'IA peut produire quatre types de résultats : la décision, la prédiction, la recommandation et le contenu.

appliqué. En revanche, plusieurs acteurs comme Amnesty International²²¹ proposent de fournir l'information, dès lors qu'il y a une interaction entre un individu et le système d'IA. Cette approche de la transparence est beaucoup plus large car elle énonce le besoin de transparence envers l'individu pour les systèmes actifs, comme des *chatbots*²²², mais aussi pour des logiciels « passifs », des caméras à reconnaissance faciale dans un lieu public par exemple.

La *fairness* est aussi sujette à ce type de divergences. Rassembler des représentants des différentes catégories de population au moment de la conception est un processus long et coûteux pour les concepteurs²²³. Est-il alors judicieux d'imposer ce genre d'obligations pour tous les systèmes ? À nouveau, la réponse varie en fonction des textes éthiques. La plupart des auteurs des auteurs de ces textes sont favorables à l'application du principe de *fairness* le plus souvent possible. L'objectif est une répartition la plus équitable possible des bénéfices de l'IA et pas seulement des conséquences négatives²²⁴. Les principes d'Asilomar parle de « *bénéfices partagés*²²⁵ ». Pris sous le prisme juridique de la non-discrimination, la *fairness* est depuis longtemps l'objet de confrontations entre différentes interprétations. Aux États-Unis par exemple, deux écoles s'affrontent. D'un côté se trouve ceux qui considèrent que la discrimination renvoie à traiter les individus de façon égale, sans distinction sur des éléments discriminants. De l'autre côté, d'autres soutiennent un traitement équitable qui passe notamment par la correction des inégalités existantes entre les catégories de population. Cette distinction entre égalité et équité est classique²²⁶. Si aux États-Unis la tendance tend vers la première école²²⁷, les tensions persistent et chaque pays peut privilégier une notion plutôt que l'autre selon sa conception de la discrimination.

Les mécanismes de mise en application du principe éthique. De manière générale, les discussions des textes sur les principes éthiques de l'IA restent à un niveau théorique et n'abordent pas en profondeur les modalités de leur application concrète. Pour autant, les auteurs énoncent certains mécanismes qui permettraient, selon eux, de contribuer à satisfaire leurs principes éthiques. Cependant, tous les auteurs n'ont pas la même vision de la manière de respecter un principe et, par conséquent, les mécanismes proposés diffèrent entre les textes. Il

²²¹ Access Now (n 55) p. 9.

²²² « Déclaration de Montréal pour un développement responsable de l'IA », 2018.

²²³ COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021.

²²⁴ Partnership on AI (n 93) Principe 1 ;

²²⁵ Future of Life Institute (n 89) (Principe 14.)

²²⁶ Voir, comme exemples parmi beaucoup d'autres : ARISTOTE, *Éthique de Nicomaque*, FLAMMARION, 2004. Livre V. ; RAWLS John, *Théorie de la justice*, POINTS, 2009.

²²⁷ SIEGEL Reva B., « Equality Divided », *Harvard Law Review*, 127, 2013.

arrive que les divergences de conception d'un même principe apparaissent explicitement. En revanche, lorsque le principe n'est pas suffisamment détaillé, les divergences conceptuelles finissent toujours par apparaître indirectement à la lecture des mesures à mettre en place. En d'autres termes, la conception d'un principe par l'auteur du texte est autant révélée par ses choix de mise en application que par la description du principe lui-même. Là où des formulations floues pourraient laisser penser à des conceptions similaires d'un même principe, les mécanismes associés clarifient souvent l'intention des différents auteurs.

Dans le cadre de la transparence, le recours à la divulgation d'informations est globalement partagé par toute la communauté comme mécanisme de transparence. Mais les divergences apparaissent sur le contenu de l'information divulguée. Les éléments possibles à intégrer dans une divulgation d'information sont nombreux et cumulables. La déclaration de Toronto propose, comme beaucoup, d'indiquer l'utilisation d'un système d'IA²²⁸. Les lignes directrices éthiques de l'entreprise SONY intègrent les conséquences possibles pour les personnes affectées²²⁹. D'autres acteurs encore demandent la divulgation le code source du logiciel²³⁰ et/ou les limites de fonctionnement du système²³¹. De son côté, le groupe d'experts de haut niveau sur l'IA considère que toutes ces informations sur le système d'IA sont importantes à communiquer²³².

Le même constat est observé pour les autres principes éthiques, comme l'illustre aussi le principe de responsabilité. L'étude d'impact est un mécanisme majoritairement adopté par les auteurs. Toutefois son contenu et l'auteur des potentiels contrôles varient selon les textes. La création d'un organe de contrôle dédié à l'IA est parfois soulevée, notamment en raison des compétences complexes et spécifiques que demande cette fonction et le manque d'expertise des autorités existantes²³³. Celui-ci se verrait confier des compétences de contrôle voire de sanctions du non-respect de la réglementation et/ou des engagements pris par l'entreprise elle-même. Entre la responsabilité *ex ante* et la responsabilité *ex post*, l'importance du juge diffère. Lorsque la responsabilité *ex ante* est priorisée, l'auteur du texte éthique favorise l'intervention

²²⁸ « The Toronto Declaration : Protecting the right to equality and non-discrimination in machine learning systems », 2018.

²²⁹ SONY, « AI Ethics Guidelines », 2018.

²³⁰ COMMISSION EUROPÉENNE POUR L'EFFICACITÉ DE LA JUSTICE (CEPEJ), « Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement », Conseil de l'Europe, 2018.

²³¹ GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019. P.18

²³² *Ibid.*

²³³ COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021.

d'autorités de contrôle. À l'inverse, l'aspect *ex post* de ce principe prévoit logiquement une place plus grande pour le juge en ce qui concerne l'évaluation des préjudices causés²³⁴.

B- La présence de conflits entre les principes

L'existence de tensions dans l'application des principes. Pour dépasser l'apport théorique et symbolique des principes éthiques, il est nécessaire de les faire respecter en pratique. La mise en application des principes éthiques apporte avec elle de nouveaux défis, en particulier celui de l'articulation des principes entre eux. L'articulation parfaite des principes lors de leur opérationnalisation n'est presque jamais possible et conduit inévitablement à des tensions. La reconnaissance de ces conflits est essentielle car elle permet, d'une part, de constater le travail à réaliser pour opérationnaliser ces principes, c'est-à-dire combler l'écart entre la théorie et la pratique²³⁵. D'autre part, cela permet aux acteurs concernés, concepteurs, utilisateurs des logiciels d'IA ainsi que les producteurs de normes, d'identifier les points de tensions afin de concentrer leurs efforts dessus²³⁶.

Présentation des tensions entre les principes éthiques. Les tensions dans l'articulation des principes éthiques peuvent prendre plusieurs formes. Premièrement, la tension peut apparaître entre les principes du fait de l'application des principes aux spécificités de l'IA²³⁷. Dans ce cas, c'est l'essence même de ce qui fait l'IA qui est à l'origine du conflit en question. Par exemple, les technologies d'apprentissage automatique nécessitent aujourd'hui une immense quantité de données pendant tout le processus de conception (données d'entraînement, de test et de validation). Lorsque ces données sont des données à caractère personnel, leur collecte massive doit s'articuler avec le droit au respect de la vie privée. Une grande quantité de données à tendance à améliorer la fiabilité du logiciel mais ce choix peut alors se heurter à la réglementation sur les données personnelles. De la même manière, l'interprétabilité du système et sa performance s'oppose depuis longtemps. Dans certaines situations comme l'analyse d'image, les technologies les plus performantes sont aussi les plus opaques. Enfin, une dernière

²³⁴ SCHERER Matthew U., « Regulating Artificial Intelligence Systems », *Harvard Journal of Law & Technology*, 29, 2015.

²³⁵ WHITTLESTONE Jess, NYRUP Rune, ALEXANDROVA Anna *et al.*, The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions, *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, Association for Computing Machinery, 2019.

²³⁶ *Ibid.*

²³⁷ *Ibid.*

tension importante se situe parfois entre le principe de fiabilité d'un algorithme et le principe de *fairness*. Par exemple, l'utilisation de données moins diverses et plus spécifiques aux situations majoritaires assure une meilleure exactitude mais augmente le risque de biais pour les situations minoritaires²³⁸.

Deuxièmement, les technologies d'IA peuvent venir aggraver une tension déjà existante entre deux principes²³⁹. Par exemple, le choix entre l'exactitude et son équité est un point de tension déjà existant dans tout processus décisionnel concernant des individus. Dans ce contexte, la capacité sans égal des modèles d'IA à différencier les individus et les situations exacerbe encore plus le conflit entre le principe d'exactitude, contenu dans le principe de fiabilité technique, et le principe de non-discrimination.

Enfin, la dernière forme de tension éthique autour des principes de l'IA est celle de la distribution inégale des conséquences des technologies utilisées²⁴⁰. En premier lieu, cette inégalité peut être perçue sous un aspect temporel. En d'autres termes, il s'agit des tensions entre des conséquences positives à court-terme et négatives à long-terme, ou inversement. Un recours aux technologies d'IA qui serait bénéfique immédiatement ne le sera pas nécessairement à long terme pour la société²⁴¹. En second lieu, cette inégalité de distribution peut affecter la population. Il est essentiel de considérer les tensions entre principes sous un prisme global en fonction des personnes affectées. Même si un équilibre peut être trouvé entre deux principes, il faut de se demander qui porte la charge des atteintes au premier principe et qui bénéficie du respect du second principe. Il existe un risque important que les mêmes catégories de personnes soient systématiquement du mauvais côté de la mise en balance²⁴². Aujourd'hui, ce phénomène est déjà observable dans de nombreuses situations²⁴³. Il s'agit alors d'une tension globale sur la manière d'équilibrer les conflits entre principes.

La transparence est probablement le principe éthique qui connaît le plus de conflits. Tout d'abord, le droit d'auteur n'étant pas envisageable pour les algorithmes, les concepteurs ne disposent comme moyen de protection que le recours au secret des affaires. En fonction du

²³⁸ Illustration de cette tension dans le cadre de systèmes d'aide au recrutement : RAGHAVAN Manish, BAROCAS Solon, KLEINBERG Jon *et al.*, *Mitigating bias in algorithmic hiring: evaluating claims and practices, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2020.

²³⁹ WHITTLESTONE Jess, NYRUP Rune, ALEXANDROVA Anna *et al.*, *op. cit.*

²⁴⁰ *Ibid.*

²⁴¹ GREEN Ben, "Good" isn't good enough, 2019.

²⁴² CALO Ryan, « Artificial Intelligence Policy », *U.C. Davis Law Review*, 51, 2017.

²⁴³ EUBANKS Virginia, *Automating Inequality*, St Martin's Press, 2018.

niveau exigé, le principe de transparence peut devenir un risque pour le patrimoine économique et intellectuel de son possesseur²⁴⁴. Débattue dans la doctrine²⁴⁵, ce conflit a été apporté devant les juges américains dans l'affaire *Loomis v. Wisconsin*²⁴⁶. De même, le risque de « triche » par les personnes sujettes aux décisions du système d'IA est un argument soulevé à l'encontre d'une transparence déraisonnable²⁴⁷. L'idée derrière cet argument est que si les individus étaient parfaitement conscients de la logique de réflexion d'un algorithme, ils parviendraient à trouver une faille pour obtenir le résultat qu'ils souhaitent sans respecter la logique recherchée par les concepteurs. Un tel risque pourrait diminuer considérablement l'utilité pratique du recours aux systèmes algorithmiques. Enfin, la vie privée est menacée devant une transparence déraisonnable. Ce-dernier conflit n'est pas nouveau et il a déjà été abordé lors de l'ouverture de certaines données. Par exemple, la réglementation française relative au principe d'*open data* connaît certaines limitations relatives à la vie privée²⁴⁸.

La gestion des tensions entre principes éthiques. Lorsque survient une tension lors de l'application de ces principes éthiques, il est inévitable de devoir les régler. La situation la plus facile et évidente est le règlement par un compromis, de façon à ce que les atteintes et bénéfices pour chacun des principes soient équilibrées de façon satisfaisante. Il est aussi possible que le conflit soit impossible à équilibrer et qu'il faille nécessairement trancher en faveur d'un des principes en jeu. Ce genre de situation nécessite une décision éminemment politique. Enfin, il existe une dernière possibilité, malheureusement trop peu souvent envisagée en pratique, en cas de conflit fort de principes, le non-déploiement du système d'IA. Les conséquences en termes d'innovation sont importantes mais elles permettent également de concentrer immédiatement l'attention du secteur sur ce qui est devenu un problème bloquant. S'il est impossible d'obtenir un compromis satisfaisant entre les principes, agir sur la possibilité de déploiement, en la réduisant ou en l'interdisant, doit toujours faire partie des options envisageables.

²⁴⁴ SERVICE DE RECHERCHE DU PARLEMENT EUROPÉEN, « A governance framework for algorithmic accountability and transparency - Think Tank », Parlement Européen, 2019.

²⁴⁵ Voir notamment WEXLER Rebecca, « Life, Liberty, and Trade Secrets », 70, 2018.

²⁴⁶ Cour suprême du Wisconsin, 13 juillet 2016, n° 881 N.W.2d 749.

²⁴⁷ KROLL Joshua A., HUEY Joanna, BAROCAS Solon *et al.*, « Accountable Algorithms », *University of Pennsylvania Law Review*, 165, 2017. P.1. À titre d'exemple : KHAZAN Olga, How to Fake Your Workout, [<https://www.theatlantic.com/health/archive/2015/09/unfit-bits/407644/>].

²⁴⁸ « Ne sont communicables qu'à l'intéressé les documents administratifs : 1° Dont la communication porterait atteinte à la protection de la vie privée [...] » - Article L. 311-6 du code des relations entre le public et l'administration.

§2. Les limites liées à la nature des principes éthiques

Annonce du plan. Au-delà des difficultés liées aux principes, la nature éthique des textes qui les contiennent nécessite un travail important pour les législateurs qui souhaiteraient s'en inspirer. D'une part, l'absence de définition claire des principes éthiques discutés permet aux acteurs d'interpréter à leur avantage ces principes **(A)**. D'autre part, l'éthique possède une force contraignante extrêmement faible qui se limite à la simple incitation. Ce niveau de contrainte ne peut garantir à lui seul une solution efficace contre les dangers des technologies d'IA et le recours au droit s'avère donc nécessaire **(B)**.

A– L'appropriation des principes éthiques par les acteurs

L'implication forte des acteurs dans l'éthique de l'IA. Le nombre colossal de textes éthiques produits sur la thématique de l'IA démontre les efforts significatifs fournis par les différents acteurs pour considérer les problématiques de ces technologies, dans l'attente ou non d'une intervention juridique par la suite. Ces efforts peuvent s'expliquer par deux raisons : d'une part, les problématiques de l'IA ont créé un besoin fort et partagé d'un cadre éthique de l'IA et, d'autre part, la forme que prendra ce cadre éthique est un enjeu majeur pour les acteurs, chacun essayant de le modeler conformément à sa vision de l'éthique de l'IA²⁴⁹. Ce combat pour la personnalisation du cadre fait ressortir l'importance des auteurs de normes éthique dans la base normative élaborée par l'éthique. Par conséquent, il convient de s'intéresser à ces acteurs, leur nature, leurs intérêts et les conséquences sur les principes éthiques.

Présentation des acteurs de l'éthique de l'IA. Afin de comprendre la diversité d'acteurs impliqués dans l'élaboration de la normativité éthique de l'IA, il est proposé ici une liste non exhaustive de catégories afin de permettre de mieux identifier les différents types d'acteurs. La première catégorie est celle regroupant les entreprises privées. Ce sont les acteurs les plus prolifiques en termes de production de textes éthiques²⁵⁰. Ce sont généralement des entreprises technologiques qui cherchent à élaborer des lignes directrices pour la conception de leurs

²⁴⁹ JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019. HICKOK Merve, « Lessons learned from AI ethics principles for future actions », *AI Ethics*, 1, 2021. « TOOLBOX », sur *AI ETHICS LAB*, publié le 25 février 2020. ALGORITHM WATCH, « AI Ethics Guidelines Global Inventory », sur *AI Ethics Guidelines Global Inventory*.

²⁵⁰ JOBIN Anna, IENCA Marcello et VAYENA Effy, *op. cit.*

systèmes d'IA ou pour leur utilisation. Elles font alors partie des destinataires des exigences contenues dans leur norme éthique. Deuxièmement, on retrouve les corps professionnels. Ces organismes visent à guider les professionnels ou le secteur qu'ils représentent au travers notamment de codes éthiques. Leurs normes s'adressent généralement aux acteurs du même secteur²⁵¹. Troisièmement, il y a les organismes de production de normes techniques. Du fait de l'approche par les principes, ces acteurs ont peu été mis en avant jusqu'à présent. Ils s'occupent de la production de normes techniques destinés à garantir le développement ou l'utilisation des systèmes d'IA de façon éthique. Les normes techniques sont indispensables pour l'applicabilité des principes éthiques, ceux élaborés par les organismes de normalisation ou d'autres contenus dans des chartes, codes, etc. La quatrième catégorie d'acteurs regroupe les institutions gouvernementales ou organismes liés aux législateurs et/ou gouvernement. Il s'agit de la deuxième catégorie d'acteurs en ce qui concerne le nombre de production de textes éthiques²⁵². Leur objectif est d'identifier les enjeux et préparer une base de réflexion pour une intervention juridique, que ce soit la production d'une nouvelle réglementation et/ou la modification de dispositions existantes. Un exemple de ce type d'acteur est le groupe d'experts de haut niveau sur l'IA. Constitué en 2018 par la Commission européenne, il réunit 52 experts issus de différents secteurs afin de produire des documents (recommandations, lignes directrices, etc.) dans le but d'aider à la mise en place de la stratégie européenne sur l'IA²⁵³. Cinquièmement et dernièrement se trouve les organisations intergouvernementales et les organisations transnationales²⁵⁴. Moins concernées directement par les lignes directrices éthiques que les précédents acteurs, elles produisent des textes à destination de l'ensemble des individus concernés par l'IA et visent rarement un secteur particulier. La variété des auteurs présentés s'accompagne d'une diversité d'intérêts dans la production de normes éthiques. Cette diversité explique les divergences d'interprétation et de hiérarchisation des principes car tous les auteurs n'ont pas la même proximité avec les enjeux et difficultés de l'IA. Néanmoins, cela rend d'autant plus intéressant l'existence d'un consensus autour de principes éthiques.

Le manque de diversité géographique. Il est également très important de noter que cette diversité de nature entre les auteurs ne se retrouvent absolument pas lorsque les textes éthiques sont analysés au travers du prisme de l'origine géographique. Les textes éthiques sont produits

²⁵¹ *Ibid.*

²⁵² HICKOK Merve, *op. cit.* « TOOLBOX », sur *AI ETHICS LAB*, publié le 25 février 2020. ALGORITHM WATCH, *op. cit.*

²⁵³ « Commission appoints expert group on AI and launches the European AI Alliance | Shaping Europe's digital future », publié le 14 juin 2018.

²⁵⁴ *Ibid.*

en grande majorité par des acteurs situés dans des pays du Nord. En effet, très peu de textes éthiques ont été produits et/ou pris en compte dans les analyses sont issus du continent africain, alors que la population fait et/ou fera l'objet de décisions par des systèmes d'IA comme partout dans le monde. De la même manière, l'Amérique du Sud, l'Amérique centrale et l'Asie centrale sont également des régions largement sous-représentées en tant qu'auteurs de normes dans les débats internationaux²⁵⁵. Il ne faut donc pas oublier que si l'élaboration de principes éthiques a atteint aujourd'hui un certain consensus sur les principes importants en IA, ce consensus ne représente pas un accord sur des valeurs universelles²⁵⁶. Il n'est pas fondé sur une participation égalitaire des personnes concernées par le sujet de l'IA mais représente une convergence d'intérêts dans un contexte d'asymétrie de pouvoir entre les différentes régions du monde.

Les divergences entre acteurs au sein des principes éthiques. Les divergences sémantiques observées au sein des principes éthiques sont largement corrélées avec la nature des acteurs à l'origine des textes. Il s'agit désormais de mettre en lumière certaines des différences qui opposent les acteurs sur le contenu des principes éthiques, en particulier sur l'opposition entre auteurs du secteur privé et auteurs du secteur public. Par exemple, l'exigence d'audit liée au principe de transparence est portée principalement par les institutions publiques, en particulier les autorités de contrôle. De même, si les solutions techniques sont proposées principalement par le secteur privé, celui-ci est moins actif lorsqu'il est question de stratégies de contrôle sociales comme la participation des personnes affectées par le logiciel d'IA dans le processus de développement²⁵⁷, la surveillance par une autorité ou encore l'amélioration de la protection des lanceurs d'alerte²⁵⁸.

Ensuite, le secteur privé associe bien moins le principe de *fairness* au droit à la non-discrimination²⁵⁹. Il préfère ainsi se concentrer sur les aspects techniques de gestion des biais algorithmiques et de *fairness* algorithmique au moment de la conception. Le prisme de réflexion sur la question de la *fairness* est celui du logiciel d'IA pris de façon isolée. La prise en compte d'un contexte plus large et des interactions que le système entretient avec son environnement

²⁵⁵ HICKOK Merve, *op. cit.*

²⁵⁶ « *More economically developed countries are shaping this debate more than others, which raises concerns about neglecting local knowledge, cultural pluralism and the demands of global fairness.* » - JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nat Mach Intell*, 1, Nature Publishing Group, 2019. P. 396.

²⁵⁷ COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Comment permettre à l'Homme de garder la main ? Les enjeux éthiques de l'intelligence artificielle. », CNIL, 2017.

²⁵⁸ Artificial intelligence and machine learning: policy paper. *Internet Society*. The Japanese Society for Artificial Intelligence Ethical Guidelines (Japanese Society for Artificial Intelligence, 2017).

²⁵⁹ JOBIN Anna, IENCA Marcello et VAYENA Effy, *op. cit.*

est limitée. Cette approche majoritairement calculatoire et technique des principes éthiques peut s'expliquer notamment²⁶⁰ par la logique d'optimisation sous-jacente aux technologies d'IA²⁶¹. Les acteurs privés concevraient alors le problème de la *fairness* comme une variable mathématique à optimiser, parmi de nombreuses autres. À contrario, le secteur public discute considérablement plus du principe de *fairness* lors de l'application, en particulier sur les conséquences pour le marché du travail et les questions démocratiques. Le secteur public traite donc davantage le principe de *fairness* au travers de ses composantes juridiques, sociales et politiques que le secteur privé.

Pour ce qui est du principe de responsabilité, l'organisme auprès duquel le concepteur du système doit rendre des comptes varie selon l'auteur du texte. Les entreprises privilégient notamment l'élaboration d'un comité de surveillance interne²⁶², à l'image des comités d'éthique. Les tentatives de mise en place de ce type de bureaux internes sont très populaires dans les grandes entreprises du secteur de l'IA²⁶³. Ces essais se sont toutefois soldés fréquemment par des échecs, que ce soit en raison du choix des membres constituant le comité²⁶⁴ et/ou du départ desdits membres après un conflit avec la direction de l'entreprise²⁶⁵. Le secteur public, en particulier les organismes gouvernementaux, sont bien plus favorables à l'établissement d'une autorité de contrôle, indépendante²⁶⁶, qui élaborerait et contrôlerait les normes techniques sur les systèmes d'IA²⁶⁷. Chacun des secteurs propose ici un aspect de la redevabilité qui lui permet de conserver le plus de pouvoir et de liberté sur la nature et la forme des exigences éthiques à respecter. De manière générale, le secteur privé pousse fortement la question de la confiance dans les technologies d'IA, notamment au travers de travaux de sensibilisation et d'explication de ces technologies au public. La confiance est en effet indispensable pour le succès commercial des systèmes d'IA et pour les investissements dans le secteur. Toutefois, la mise en avant de la confiance sociale dans les technologies d'IA est

²⁶⁰ Une explication complémentaire proposée par certains auteurs est la proportion de femmes parmi les autrices des textes éthiques. Ces dernières auraient une logique moins « calculatoire et rationnelle » mais davantage « empathique ». HAGENDORFF Thilo, « The Ethics of AI Ethics -- An Evaluation of Guidelines », *Minds & Machines*, 30, 2020.

²⁶¹ GREEN Ben, "Good" isn't good enough, 2019.

²⁶² Microsoft AI principles

²⁶³ Sam Shead, « Facebook Reportedly Has A Dedicated AI Ethics Team », en ligne: *Forbes* <<https://www.forbes.com/sites/samshead/2018/05/03/facebook-reportedly-has-a-dedicated-ai-ethics-team/>>.

²⁶⁴ PIPER Kelsey, « Exclusive », sur *Vox*, publié le 4 avril 2019.

²⁶⁵ METZ Cade et WAKABAYASHI Daisuke, « Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I. », *The New York Times*, 3 décembre 2020.

²⁶⁶ Amnesty International, Access Now (n 55) p. 10.

²⁶⁷ German Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy, and the Federal Ministry of Labour and Social Affairs (n 9) p. 26

contestée par certains auteurs qui y voient un risque de diminution du niveau de surveillance et d'obligations sociales exigés aux entreprises conceptrices de ces systèmes²⁶⁸. La divergence d'intérêts entre auteurs, notamment entre les acteurs du secteur public et ceux du secteur privé, illustre les enjeux dans la détermination des valeurs et des principes éthiques pour l'IA. Les choix des législateurs lors de l'élaboration d'un cadre juridique spécifique à l'IA prioriseront nécessairement les intérêts de certains acteurs par rapport à d'autres. De plus, cette analyse met en évidence l'importance des acteurs impliqués dans l'élaboration de la norme sur le contenu de cette-dernière.

B- Le faible niveau de contrainte des textes éthiques

La force contraignante de l'éthique. Les textes éthiques, que ce soient des chartes éthiques, des codes de conduite d'entreprise ou encore des rapports d'institutions gouvernementales, ne sont pas contraignants. Contrairement au droit, ce corpus éthique n'est qu'incitatif. L'objectif est d'aiguiller, proposer, discuter de problématiques et de solutions autour des technologies d'IA. C'est en partie grâce à ce manque de caractère contraignant qu'autant de textes ont été produit et qu'un consensus a pu être identifié. Le caractère contraignant aurait pu empêcher ou dissuader bon nombre d'acteurs d'apporter leur pierre à l'édifice. Du point de vue de la légitimité, les institutions gouvernementales et les autorités de contrôle n'ont pas toutes la compétence et/ou capacité de produire des normes générales contraignantes juridiquement. Également, les entreprises auraient été bien moins actives et donc représentées si des sanctions étaient automatiquement associées aux textes éthiques qu'elles produisent. Cette absence de force contraignante présente bien évidemment des inconvénients qui imposent de mettre en perspective la proactivité de certains acteurs du secteur privé. En effet, la forte implication du secteur privé dans l'élaboration de textes éthiques peut s'analyser comme un moyen d'éviter la mise en place d'une réglementation sur le sujet de l'IA²⁶⁹. L'objectif est de démontrer qu'elles sont conscientes des enjeux éthiques derrière les technologies d'IA et qu'elles sont capables de lutter seules. De ce fait il ne serait pas nécessaire d'imposer une réglementation, ou alors

²⁶⁸ Bryson, J. AI & global governance: no one should trust AI. *United Nations University* <https://cpr.unu.edu/ai-global-governance-no-one-should-trust-ai.html> (2018).

²⁶⁹ « *In this context [corporate and technology policy], ethical frameworks that provide a way to go beyond existing legal frameworks can also provide an opportunity to ignore them.* » - WAGNER Ben, « Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping? », in *Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?*, Amsterdam University Press, 2018.

sous l'aspect d'une auto-régulation²⁷⁰. Cette démarche est connue sous le nom d'*ethical lobbying*²⁷¹. La vision de l'éthique comme un norme concurrente et pouvant remplacer le droit comme est en réalité une conception erronée fréquemment présente dans les textes éthiques rédigés par le secteur privé²⁷². En outre, l'absence de contrôle du respect de ces normes éthiques et de sanctions en cas de manquement avéré expose au risque d'*ethical washing*²⁷³. Inspiré du concept de *greenwashing* relatif à l'écologique, l'*ethical washing* désigne la pratique consistant à communiquer des informations erronées au public ou de mettre en place des mesures superficielles, de façon consciente et délibérée, dans le but de paraître plus éthique que la réalité²⁷⁴.

Ce phénomène s'appelle l'*ethical washing*, appellation directement inspirée de son équivalent pour l'écologie, le *green washing*. La seconde utilité trouvée par les entreprises dans la proposition de normes non contraignantes serait la volonté de déguiser sous les traits de problèmes techniques des problématiques en réalité sociales et politiques²⁷⁵. Cet aspect est particulièrement présent au sein des normes techniques. Cette technicisation des enjeux permet, d'une part, de dévier l'attention d'acteurs sur d'autres problématique. D'autre part, cela permet d'affirmer la légitimité du secteur privé à s'occuper desdits problèmes, celui-ci se présentant comme le plus compétent pour régler la difficulté²⁷⁶.

La contrainte au sein des textes éthiques. Malgré l'absence de force contraignante de l'éthique, la question de la contrainte est tout de même discutée dans certains textes éthiques mais de manière indirecte. Lorsque le manquement à un principe est énoncé comme rédhibitoire pour le déploiement du système d'IA en question, il s'agit d'un facteur de contrainte du principe. C'est le cas par exemple du rapport de la Chambre des Lords sur l'IA qui énonce qu'un certain niveau d'explicabilité doit être atteint pour le déploiement d'un système d'IA présentant un risque élevé²⁷⁷. De façon similaire, énoncer que le manquement à un principe

²⁷⁰ BENKLER Yochai, « Don't let industry write the rules for AI », *Nature*, 569, 2019.

²⁷¹ FLORIDI Luciano, « Translating Principles into Practices of Digital Ethics », *Philos. Technol.*, 32, 2019.

²⁷² En réalité, les deux sources normatives peuvent parfaitement coexister et agir de façon complémentaire vers un objectif commun. En outre, l'absence de contrôle. WAGNER Ben, *op. cit.*

²⁷³ CASTETS-RENARD Céline, « Comment construire une intelligence artificielle responsable et inclusive ? », *Recueil Dalloz Sirey*, 4, 2020.

²⁷⁴ « *The malpractice of making unsubstantiated Or, misleading claims about, Or, implementing superficial measures in favour of, the ethical values and benefits of digital processes, products, services, Or, other solutions in order to appear more digitally ethical than one is.* » - FLORIDI Luciano, *op. cit.*

²⁷⁵ GREENE Daniel, HOFFMANN Anna Lauren et STARK Luke, *Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning*, 2019.

²⁷⁶ BARRAUD Boris, « La corégulation d'Internet (ou comment répondre à la plurinormativité par l'internormativité) — Une contribution française », *cd1*, 59, Faculté de droit de l'Université Laval, 2018.

²⁷⁷ UK House of Lords, Select Committee on Artificial Intelligence (n 7) p. 40.

entraînera le retrait du système d'IA du marché ou de sa mise en service est un moyen, ici *ex post*, de discuter de la force contraignante que devrait revêtir un principe. La présence d'éléments relatifs à la contrainte des principes peut sembler étonnant au sein de textes éthiques à la vocation plus incitative que contraignante. En réalité, il s'agit d'une mise en lumière du rôle fort et assumé d'inspiration législative de l'éthique. Il s'agit d'exemples concrets de l'utilisation de l'éthique par certains acteurs pour préfigurer et influencer des règles juridiques futures. Ces règles plus contraignantes, auront alors le choix de s'inspirer du corpus éthique (ou non) ou d'accorder la force contraignante du droit à ces textes par divers moyens.

CONCLUSION DE LA SECTION I :

Aux premiers abords, les principes énumérés dans les textes éthiques sur l'IA mettent en lumière des objectifs à poursuivre pour lutter contre les risques de ces technologies. Équité, vie privée, fiabilité technique, transparence, responsabilité et contrôle humain sont les six principes partagés par l'ensemble des auteurs des textes éthiques sur l'IA. Le consensus autour de ces principes éthiques de l'IA permet d'affirmer qu'il s'agit de points de fond importants attendus dans la réglementation à venir. Au-delà de ces objectifs substantiels, l'analyse approfondie des limites des textes éthiques a permis d'identifier également des besoins importants relatifs à la forme de l'encadrement. Le manque de précision des principes éthiques camoufle en réalité une divergence importante d'interprétation entre les auteurs des textes, chacun mettant en avant une conception favorable aux intérêts qu'il défend. Négligée par les textes éthiques sur l'IA, le besoin d'articulation des principes entre eux apparaît aussi comme un point de forme essentiel dans l'encadrement de l'IA. L'éthique de l'IA révèle donc aussi bien des besoins de fond que des besoins de forme dans l'encadrement de l'IA. Parce que l'éthique et le droit peuvent être complémentaires, il convient désormais de compléter notre étude des besoins normatifs de l'IA avec l'analyse des règles de droit positif applicables à ces technologies.

Section II : Les besoins du droit positif relatif à l'utilisation d'outils algorithmiques

Annonce de plan. L'absence d'un droit *sui generis* aux technologies d'IA ne signifie pas pour autant que leur utilisation soit dépourvue de contraintes juridiques. Certains risques générés par l'IA ne sont pas totalement nouveaux et existaient déjà avec l'arrivée de l'automatisation et

l'utilisation des algorithmes déterministes. Des dispositions juridiques ont alors été élaborées, en particulier autour des résultats produits à l'aide et par des algorithmes. Elles seront regroupées ici sous l'expression de « droit des décisions impliquant un algorithme ». Leur existence et leur application à l'IA pose deux enjeux pour l'encadrement de l'IA. Tout d'abord, et bien que la très grande majorité de ces dispositions n'ait pas été conçue spécifiquement pour lutter contre les risques de l'IA, certaines d'entre elles demeurent des mesures de protection importantes pour les acteurs concernés. Il existe donc un besoin d'articulation entre ces règles pertinentes de droit positif et le droit de l'IA afin de conserver au mieux leur efficacité (**Sous-section 1**). Ensuite, les particularités des technologies d'IA mettent à mal l'application des obligations et droits contenus dans ce cadre, produisant une insécurité juridique pour les individus concernés. Le cadre juridique de l'IA doit tenir compte de ces lacunes afin de pas les reproduire et de les corriger. En d'autres termes, les futurs textes ont la charge de répondre à un besoin d'amélioration du droit positif appliqué l'IA (**Sous-section 2**).

Sous-section 1 : Un besoin d'articulation avec le droit positif

Annonce du plan. Les dispositions de droit positif applicables, directement ou indirectement, aux systèmes d'IA sont bien trop importantes pour espérer en produire ici une analyse exhaustive. La présente démarche vise à mettre en avant l'existence de dispositions juridiques pertinentes en droit de l'IA afin de souligner l'importance d'une bonne articulation avec le droit spécifique à l'IA. Les mécanismes présentés ci-dessous ont été sélectionnés car ils représentent, d'une part, une volonté législative de considérer les risques liés aux décisions automatisées et, d'autre part, ils concernent des préoccupations centrales en ce qui concerne l'IA. Les deux exemples que sont la transparence (§1) et le contrôle humain (§2) font ainsi partie du consensus éthique énoncé précédemment par les membres de la communauté des technologies d'IA.

§1. Les mécanismes juridiques de la transparence

Annonce du plan. Depuis l'arrivée des technologies informatiques dans les processus de prise de décision, la complexité technique et le secret autour de leur usage ont entraîné la mise en place de garanties de transparence par les législateurs. Ces garanties sont apportées par des

mécanismes dont l'efficacité dépend de la combinaison entre leurs modalités de forme, à savoir la manière dont sont communiquées les informations **(A)**, et les modalités de fond, soit le contenu de l'information communiquée **(B)**.

A – Les formes de la transparence

Annonce du plan. Les obligations imposées en cas de décision automatisée peuvent varier selon trois critères, à savoir le nombre de destinataires de l'information, le moment de sa délivrance et l'acteur devant initier le mécanisme de transparence. Au travers des différentes dispositions sur la transparence algorithmique, trois catégories de mécanismes peuvent être identifiées : l'obligation d'information individuelle **(1)**, la communication d'informations sur demande **(2)** et la publication des informations **(3)**.

1) L'obligation d'information individuelle.

Les rôles de l'obligation d'information. Le premier mécanisme de transparence aujourd'hui en place est l'obligation de d'information de la personne qui sera affectée par la future décision. Parfois garante du caractère éclairé du consentement donné par la personne affectée, ce rôle est aujourd'hui critiqué. Entre la surcharge d'informations²⁷⁸ et l'intelligibilité réelle²⁷⁹ de ces dernières, l'obligation d'information semble une mesure de protection de l'individu de plus en plus superficielle. Toutefois, l'obligation d'information dans le cadre de décision impliquant un algorithme n'est pas limitée à cette seule fonction. L'obligation d'information est imposée dans toutes les situations, qu'un consentement soit requis ou non. La fourniture de renseignements déterminés sur la situation et son contexte permet de réduire, certes faiblement, l'asymétrie de pouvoir et d'informations entre l'utilisateur du système algorithmique et la personne sujette à la décision.

²⁷⁸ « À regression analysis identifies information overload as a significant negative predictor of reading TOS upon signup, when TOS [Terms of Service] changes, and when PP [Privacy Policy] changes. » - Jonathan A Obar & Anne Oeldorf-Hirsch, « The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services » (2020) 23:1 Information, Communication & Society 128-147. P. 2. OEHLER Andreas et WENDT Stefan, « Good Consumer Information », *J. Consumer Pol'y*, 40, 2017.

²⁷⁹ BEN-SHAHAR Omri, « The Myth of the 'Opportunity to Read' in Contract Law », 5, *De Gruyter*, 2009. BAILLOUX Cassandre, « The Average Consumer in European Consumer Law », *Exeter L. Rev.*, 44, 2017.

L'obligation d'information en droit positif. Le règlement européen sur la protection des données personnelles²⁸⁰ impose ainsi à ses articles 13 et 14 une obligation d'information à la charge du responsable de traitement. Dans le cadre d'une décision individuelle automatisée, au sens de l'article 22 du RGPD, l'information doit contenir des informations supplémentaires. Ainsi les articles 13.2.f) et 14.2.g) prévoient tous deux une obligation d'informer la personne concernée préalablement au début du traitement des données personnelles. Le début du processus décisionnel est un moment propice à l'information de l'individu, il permet à l'individu d'utiliser cette connaissance nouvelle tôt dans le traitement et donc potentiellement limiter les préjudices non souhaités le plus vite possible. De ce fait, ce choix du moment d'information est très fréquent dans les textes juridiques.

La loi pour une République numérique de 2016 (ci-après LRN) impose une mention d'information explicite à l'intéressé pour toute décision administrative individuelle fondée sur un traitement algorithmique²⁸¹. Un choix similaire est réalisé au sein de la loi française relative à la bioéthique qui dispose d'une obligation d'information au patient et aux professionnels de santé avant tout « *acte de prévention, de diagnostic ou de soin* » qui impliquerait le recours à un traitement de données par un logiciel exploitant des technologies d'apprentissage automatique²⁸². Il arrive aussi que l'information individuelle soit fournie après la prise de décision²⁸³. Toutefois, ce choix d'information individuelle *ex post* est rare. D'une part, il peut représenter une charge lourde pour le débiteur de l'obligation (entreprise, administration, etc.). D'autre part, le destinataire n'est pas forcément intéressé par ces renseignements, rendant le cas échéant inutile le travail supplémentaire d'information effectué. En ce qui concerne la transparence après la décision, le mécanisme de la communication sur demande est majoritairement privilégié.

²⁸⁰ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

²⁸¹ « LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique », 2016-1321, 2016. Article 4, codifié à l'article L. 311-3-1 du CRPA.

²⁸² « LOI n° 2021-1017 du 2 août 2021 relative à la bioéthique (1) », 2021-1017, 2021. Article 17.

²⁸³ « Directive sur la prise de décision automatisée (Canada) », 2019.

2) La communication d'informations sur demande

Contrairement aux dispositions précédentes, la communication d'informations sur demande n'est pas une obligation mais un droit²⁸⁴. Par conséquent, cette prérogative n'est pas automatique et doit être actionnée par son bénéficiaire. Ce besoin de proactivité de la part de ce-dernier expose à un risque de faible recours à ce droit²⁸⁵ et donc d'une effectivité décevante *in fine*. Cependant il présente aussi de nombreux avantages en termes de transparence. Il apparait donc fréquemment dans le droit positif des décisions impliquant un algorithme.

La communication sur demande en droit des données personnelles. Ainsi, le RGPD offre à la personne concernée un droit d'accès à son article 15. Dans le texte, les informations spécifiques au système automatisé sont identiques avec celles contenues dans l'obligation d'information²⁸⁶. L'un des avantages de ce droit est qu'il peut être exercé à plusieurs reprises par la personne concernée au cours du traitement de ses données personnelles. La personne concernée peut alors suivre une potentielle évolution d'un traitement de données personnelles de longue durée et avoir un accès à l'ensemble des données concernant le traitement depuis son démarrage. Toutefois l'exercice de ce droit fait peser une charge de travail potentiellement importante sur le responsable de traitement²⁸⁷. Il possède donc un délai d'un mois pour répondre à la demande, délai extensible à trois mois sous certaines conditions²⁸⁸. Dans la même logique, l'usage répété et injustifié du droit d'accès par un individu peut justifier un refus ou exiger un paiement pour la demande d'accès par le responsable de traitement²⁸⁹.

La communication sur demande en droit administratif. La directive canadienne sur la prise de décision automatisée met en avant une autre possibilité de mise en œuvre d'une communication d'informations sur demande. Avant de poursuivre cet exemple, il convient d'expliquer la nature d'une directive en droit canadien. Contrairement en droit européen, les directives canadiennes ne sont pas des règles de droit mais des actes « *de nature et de portée essentiellement administratives* »²⁹⁰. Les exigences et sanctions qu'elles contiennent ne sont

²⁸⁴ AUBERT Jean-Luc et SAVAUX Eric, *Introduction au droit et thèmes fondamentaux du droit civil*, Sirey, 2023.
²⁸⁵ ROCHFELD Judith, *Les grandes notions du droit privé*, Presses Universitaires de France - PUF, 2022

²⁸⁶ EDWARDS Lilian et VEALE Michael, « Enslaving the Algorithm », *IEEE Security & Privacy*, 16, 2018. P. 6.

²⁸⁷ Pour plus de détails sur le contenu, voir supra. Il est important de noter que les informations générales concernant le traitement de données ne sont pas identiques en pratique.

²⁸⁸ En réalité, la demande d'accès peut aussi être réalisée par le sous-traitant en fonction de la répartition des obligations inscrite dans le contrat de sous-traitance imposé à l'article 28.3 du règlement européen.

²⁸⁹ Article 12.3 du RGPD

²⁹⁰ Article 12.5 du RGPD. Il s'agit des situations où les demandes d'accès sont « *manifestement infondées ou excessives, notamment en raison de leur caractère répétitif [...]* ».

²⁹¹ Federal Court, 7 mai 2002, n° T-1119-01. Par. 37.

destinés qu'aux fonctionnaires de l'administration, les directives canadiennes n'ouvrent aucun droit actionnable pour les administrés²⁹¹. En dépit de cette limite significative, les exigences contenues dans le texte se révèlent pertinentes pour notre étude. Contrairement au RGPD, la directive canadienne ne permet pas les demandes multiples pendant le processus de traitement car elle ne peut être demandée qu'après le rendu de la décision.²⁹² Le texte dispose d'une possibilité de demander une explication pour les décisions de refus dont le niveau d'incidence est modéré²⁹³. Toujours dans le contexte du droit administratif, la LRN propose aussi un droit de demander la communication d'informations sur les décisions algorithmiques. Le destinataire de cette communication est un enjeu crucial dans la portée de l'objectif de transparence. La LRN permet à tout individu « *intéressé* » d'obtenir des informations précises sur le traitement algorithmique²⁹⁴. Contrairement aux précédentes dispositions, ce terme renvoie à plus de destinataires que la seule personne affectée par la décision. Toutefois cet élargissement des destinataires n'est pas garanti. De nombreuses exceptions sont présentes comme dans le cas de la procédure nationale de préinscription à l'université, appelée Parcoursup, qui rend l'acceptation des bacheliers aux études supérieures partiellement automatisée²⁹⁵. Les dispositions concernant Parcoursup ont également réduit les destinataires de ce mécanisme de transparence aux seuls candidats²⁹⁶. Un syndicat étudiant, qui aurait pu revendiquer la qualité de partie intéressée s'est vu refuser le droit de demander la communication des informations spécifiques aux traitements algorithmiques en œuvre²⁹⁷.

3) La publication d'informations

Une transparence à grande échelle. La publication est le mécanisme de transparence avec la portée la plus large car, par définition, son contenu est destiné à tous. Elle présente plusieurs avantages par rapport à l'information individuelle. Premièrement, elle offre la possibilité pour des personnes tierces expérimentées en informatique d'expliquer et/ou réexpliquer le contenu de la publication aux personnes moins à l'aise avec la technologie. Secondement, les

²⁹¹ SCASSA Teresa, « Administrative Law and the Governance of Automated Decision-Making », *University of British Columbia Review*, 54, 2021.

²⁹² Article 6.2.3 « *Parvenir des explications après les décisions* »

²⁹³ Pour plus détails, voir l'annexe C de la directive.

²⁹⁴ Article L. 311-3-1 du CRPA

²⁹⁵ Article L. 612-3 du code de l'éducation

²⁹⁶ Article L. 612-3 alinéa 6 du code de l'éducation.

²⁹⁷ Conseil d'État, 12 juin 2019, n° 427916.

renseignements publiés peuvent servir de référence pour identifier des tendances globales (tels que les risques systémiques²⁹⁸) plus difficilement détectables par le biais d'informations individuelles. La publication semble actuellement gagner en popularité lorsqu'une décision est prise par ou à l'aide d'un système algorithmique. Par exemple, la directive canadienne sur la prise de décision automatisée prévoit la publication de l'utilisation d'un système décisionnel automatisé au travers d'un avis sur Internet²⁹⁹. Sauf exceptions, l'administration canadienne doit également publier le code source de ces systèmes automatisés³⁰⁰. Le code source est fait également partie des documents à publier en droit français, comme tout document administratif³⁰¹. De plus, le CRPA impose la publication des « *règles définissant les principaux traitements algorithmiques utilisés dans l'accomplissement de leurs missions [...]* »³⁰². Le règlement européen sur les services numériques contient également une obligation de publication d'informations. En effet, le texte européen impose aux opérateurs de plateforme en ligne la publication de « *rapports de transparence* » contenant des informations sur :

« (c) tout recours à des moyens automatisés à des fins de modération de contenus, y compris une spécification des objectifs précis, des indicateurs de la précision des moyens automatisés pour atteindre ces objectifs et des éventuelles mesures de sauvegarde appliquées. »³⁰³

La publication est la mise en œuvre du principe de transparence la plus aboutie en termes de portée de l'information. Cependant, cette ampleur de diffusion pose plusieurs inconvénients quant à la nature de l'information pouvant être publiée. Une mise en balance entre la portée et le contenu de la publication est alors réalisée par les législateurs.

L'impact de la forme sur le fond des mécanismes de publication. L'inconvénient de la publication est le risque de violation de confidentialité des informations. Cela impose de vérifier que le contenu de la publication desdites informations n'enfreint aucune réglementation relative aux données. À l'exception du DSA, les autres textes prévoient les situations de conflits entre le principe de publication qu'ils imposent et les différents secrets qu'ils pourraient enfreindre.

²⁹⁸ « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. Article 34.

²⁹⁹ Articles 6.2.1 et 6.2.2 de la directive

³⁰⁰ Article 6.2.6 de la directive.

³⁰¹ Article L. 300-2

³⁰² Article L. 312-1-3 du CRPA. Cette publication s'impose pour les administrations employant plus de cinquante personnes équivalents temps pleins

³⁰³ « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. Article 23.1.c)

Ces solutions s'illustrent, soit comme des compromis, soit comme des exceptions strictes à l'obligation. Les compromis peuvent prendre la forme de « *restrictions d'accès appropriées* »³⁰⁴, en réduisant le nombre de destinataires³⁰⁵ ou rendant illisibles les informations qualifiées de sensibles. Les exceptions strictes se trouvent principalement dans les secteurs comme la sécurité nationale³⁰⁶. Ces différents aménagements démontrent que la forme que prennent les mécanismes de transparence doit toujours prendre en considération leur contenu, l'efficacité et les risques de celle-ci y sont intimement liés.

B – Le contenu de la transparence

Annnonce de plan. Outre la forme que prend le mécanisme de transparence, son contenu est un élément significatif pour déterminer l'efficacité et l'objectif de la garantie mise en place par le droit. Comme illustré par l'analyse des textes éthiques, le contenu de l'information peut varier de façon importante selon la conception du principe de transparence choisie par les acteurs. Les décisions impliquant un algorithme présentent deux niveaux d'opacité : le secret et la complexité technique. Chacun de ces niveaux nécessite des informations particulières pour être surmonté. Pour le premier, il est essentiel d'informer la personne affectée du contexte dans lequel elle se trouve, à savoir une décision prise totalement ou partiellement par un système automatisé **(1)**. Pour le second, le destinataire de la transparence ne peut vaincre l'opacité qu'à l'aide d'informations sur le processus décisionnel ayant abouti à la décision, en particulier des éléments sur le fonctionnement de l'algorithme utilisé **(2)**.

1) L'information d'utilisation d'un algorithme

La notification de la présence d'un algorithme dans la prise de décision. Premièrement, l'information de la présence d'une décision automatisée permet d'éviter toute dissimulation du recours à un algorithme par son utilisateur. Il s'agit du strict minimum de la transparence qui se retrouve dans de nombreuses réglementations. Par exemple, dans le domaine de la santé, la

³⁰⁴ Article 6.2.7 de la directive canadienne.

³⁰⁵ Article L. 311-6 du CRPA énumère les situations où la publication est restreinte à la simple communication à l'intéressé.

³⁰⁶ Article 6.2.6.1 de la directive canadienne, Article L. 311-5.2 du CRPA

loi française relative à la bioéthique³⁰⁷ prévoit à son article 17 alinéa 1³⁰⁸ la même notification à la personne concernée :

« I.- Le professionnel de santé qui décide d'utiliser, pour un acte de prévention, de diagnostic ou de soin, un dispositif médical comportant un traitement de données algorithmique dont l'apprentissage a été réalisé à partir de données massives s'assure que la personne concernée en a été informée [...] »

L'article 17 mentionne explicitement un autre destinataire de cette obligation d'information, les professionnels de santé : *« II.-Les professionnels de santé concernés sont informés du recours à ce traitement de données. [...] »*. Avec ce second destinataire, la disposition distingue intelligemment entre l'utilisateur du système algorithmique et les autres individus participant au processus décisionnel, tous n'ayant pas nécessairement le même niveau d'informations et de compétences.

L'information concernant le fonctionnement est également communiquée dans le cadre d'une décision prise par l'administration publique. En France, l'article 4 de la loi pour une république numérique³⁰⁹, codifié à l'article L. 311-3-1 du Code des relations entre le public et l'administration prévoit que : *« une décision individuelle prise sur le fondement d'un traitement algorithmique comporte une mention explicite en informant l'intéressé »*. En droit canadien, la directive relative à la prise de décision automatisée impose à l'administration de communiquer l'utilisation d'un système algorithmique par le biais d'un avis publié sur les sites web du service administratif concerné³¹⁰. Cet avis n'est cependant pas obligatoire pour les décisions n'ayant pas ou peu d'effets sur l'administré.

Le règlement européen sur les services numériques³¹¹ (DSA ci-après) prévoit pour les fournisseurs de services intermédiaires une obligation de mentionner dans les conditions générales les modalités de modération réalisées, notamment les : *« [...] mesures et outils utilisés à des fins de modération des contenus, y compris la prise de décision fondée sur des algorithmes [...] »*.

La mention seule de l'existence d'un algorithme dans le processus décisionnel est loin d'être suffisante pour octroyer un niveau de transparence satisfaisant à l'individu. Par conséquent dans

³⁰⁷ « LOI n° 2021-1017 du 2 août 2021 relative à la bioéthique (1) », 2021-1017, 2021.

³⁰⁸ Codifié à l'article L. 4001-3.-I du Code de la santé publique français.

³⁰⁹ « LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique », 2016-1321, 2016.

³¹⁰ Article 6.2 de la directive sur la prise de décision automatisée

³¹¹ DSA > Ajouter la ref lorsque le texte sera publié officiellement au JO UE

le RGPD comme pour la directive canadienne et bien d'autres textes, d'autres informations sont fournies avec cette notification. L'objectif ici n'est pas de lister l'ensemble des éléments présents dans les obligations d'information imposées dans l'ensemble des textes existants. Notifier la présence d'un algorithme permet d'informer la personne concernée, et parfois d'autres intervenants dans la décision, du contexte dans lequel elle(s) se situe(nt). La connaissance de ce contexte permet ensuite à l'individu de recevoir les autres informations relatives à la situation, de la finalité du traitement aux droits spécifiques qu'il possède. Cette mention d'information permet de lutter contre le premier niveau d'opacité, l'opacité par le secret³¹². Sans cela, lutter contre les autres niveaux d'opacité resterait inutile pour la personne affectée par la décision.

2) L'information sur le fonctionnement de l'algorithme

La lutte contre la complexité algorithmique. Le second élément traité concerne la deuxième couche d'opacité dans les systèmes algorithmiques : la complexité technique. Ce niveau d'opacité est commun aux systèmes algorithmiques intégrant de l'IA ou non. Par conséquent, le droit positif contient déjà des obligations destinées à rendre compréhensible le fonctionnement complexe de ces systèmes.

Il est possible de retrouver une obligation d'information relative au fonctionnement d'un système algorithmique dès 1995. Dans son article 12, la directive européenne sur les données personnelles prévoit un droit d'accès à la personne concernée. Lorsqu'il est exercé, ce droit lui permet d'obtenir « *la connaissance de la logique qui sous-tend tout traitement automatisé des données la concernant [...]* »³¹³. Abrogée en mai 2018 par le RGPD, cette disposition se retrouve dans les articles 13.f), 14.g) et 15.h) du règlement européen. L'expression « *logique sous-jacente* » n'est pas davantage détaillée dans le reste du règlement, donnant lieu à des débats doctrinaux passionnés³¹⁴, les garanties liées à cette disposition restent donc encore incertaines.

³¹² BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016.

³¹³ « Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données », OJ L, 1995. Article 12.a) alinéa 3

³¹⁴ Voir supra

Les dispositions existantes en droit administratif sont plus récentes et plus détaillées sur les éléments à produire pour aider à l'administré à lutter contre cette technicité. Cela s'explique par l'importance particulière que revêtent les décisions algorithmiques administratives par rapport à celles du secteur privé. Contrairement à ces dernières, les algorithmes utilisés par l'administration le sont au profit de l'intérêt général et non privés. Ils sont quasi systématiquement imposés à l'administré, sans alternative possible pour lui. Enfin, ces systèmes algorithmiques représentent l'interface entre l'État et le public, ce qui en fait donc des outils importants de la confiance de la population en l'État³¹⁵.

Le droit administratif français offre ainsi un droit d'accès, sur demande de l'intéressé, portant sur « *Les règles définissant ce traitement ainsi que les principales caractéristiques de sa mise en œuvre sont communiquées à l'intéressé s'il en fait la demande.* »³¹⁶. Les termes flous sont cette fois précisés au travers d'un décret d'application³¹⁷ :

« L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes :

1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision ;

2° Les données traitées et leurs sources ;

3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé ;

4° Les opérations effectuées par le traitement. »³¹⁸

Le « *degré et le mode de contribution du traitement algorithmique* » sont deux éléments d'information permettant à l'administré de cerner le rôle et l'influence du système algorithmique dans le processus décisionnel. Les données traitées désignent les données entrées dans le système. L'objectif est de permettre à l'individu d'identifier des potentielles données erronées concernant sa situation. Le troisième alinéa renvoie aux différents critères utilisés par le système et aux poids de chacun d'entre eux. La précision « *appliqués à la situation de*

³¹⁵ ENA, « Ethique et responsabilité des algorithmes publics », Etalab, 2019.

³¹⁶ Article L. 311-3-1 du CRPA

³¹⁷ « Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique », 2017-330, 2017.

³¹⁸ Article R. 311-3-1-2 du CRPA

l'intéressé » montre clairement que ces paramètres et leur pondération doivent être communiqués afin d'informer du fonctionnement personnalisé, voire individualisé, du traitement dont est sujet l'administré. Enfin, l'expression « *opérations effectuées* » parle d'elle-même.

Contrairement au droit français et européen, la directive canadienne sur la prise de décision automatisée par les administrations dispose d'obligations qu'elle qualifie explicitement d'explications : « *Fournir une explication significative aux personnes concernées sur la façon dont la décision a été prise et la raison pour laquelle elle a été prise [...]* »³¹⁹. Les raisons de la prise de décision peuvent renvoyer à une obligation de justification de la décision tandis que les informations sur la manière dont la décision a été prise désigne plutôt une obligation d'explication du fonctionnement³²⁰.

Enfin, l'on retrouve ce type de dispositions dans le droit des fournisseurs de services intermédiaires, en particulier concernant les opérateurs de plateformes en ligne³²¹. Dès 2016 en France, la LRN impose à ces opérateurs de plateforme en ligne une obligation d'information :

« II.- Tout opérateur de plateforme en ligne est tenu de délivrer au consommateur une information loyale, claire et transparente sur :

1° Les conditions générales d'utilisation du service d'intermédiation qu'il propose et sur les modalités de référencement, de classement et de déréférencement des contenus, des biens ou des services auxquels ce service permet d'accéder ;

2° L'existence d'une relation contractuelle, d'un lien capitalistique ou d'une rémunération à son profit, dès lors qu'ils influencent le classement ou le référencement des contenus, des biens ou des services proposés ou mis en ligne ; [...] »³²²

Cette rubrique relative aux critères de référencement est précisée par un décret paru en 2017³²³ qui indique son contenu :

³¹⁹ Article 6.2.3 de la directive canadienne

³²⁰ SCASSA Teresa, « Administrative Law and the Governance of Automated Decision-Making », *University of British Columbia Review*, 54, 2021.

³²¹ ROCHFELD Judith, « Le projet de loi pour une République numérique : entre espoirs et regrets », Dalloz IP/IT, 2016. ROCHFELD Judith et ZOLYNSKI Célia, « La « loyauté » des « plateformes ». Quelles plateformes ? Quelle loyauté ? », *Dalloz IP/IT*, 2016.

³²² Article L. 111-7.II du code de la consommation français

³²³ « Décret n° 2017-1434 du 29 septembre 2017 relatif aux obligations d'information des opérateurs de plateformes numériques », 2017-1434, 2017.

« 1° Les conditions de référencement et de déréférencement des contenus et des offres de biens et services, notamment les règles applicables pour être référencé et les obligations dont le non-respect conduit à être déréférencé ;

2° Les critères de classement par défaut des contenus et des offres de biens et services, ainsi que leurs principaux paramètres ;

3° Le cas échéant, l'existence d'un lien capitalistique ou d'une rémunération entre l'opérateur de plateforme et les offreurs référencés dès lors que ce lien ou que cette rémunération exerce une influence sur le référencement ou le classement des contenus, des biens ou des services proposés ou mis en ligne. »

Cette information doit être contenue dans une rubrique spécifique de la plateforme mais également présent sur chaque page de résultats et même pour chaque résultat lorsqu'il s'agit de relations capitalistiques ou de rémunération.

Depuis l'entrée en vigueur du DSA le 17 février 2024³²⁴, les plateformes en ligne sont également contraintes au niveau européen, en ce qui concerne la publicité en ligne, d'informer les internautes des « *principaux paramètres utilisés pour déterminer le bénéficiaire auquel la publicité est présentée* »³²⁵. La logique de personnalisation des publicités est ainsi plus transparente vis-à-vis de son destinataire. Les opérateurs en charge de très grandes plateformes au sens du texte ont également une obligation d'information aux consommateurs relative aux systèmes de recommandations. Inscrite dans les conditions générales, ces informations concernent à nouveau les « *principaux paramètres utilisés dans les systèmes de recommandations* ».

Conclusion. Que ce soit son fonctionnement global ou local, les mécanismes de transparence concentrés sur le fonctionnement du système décisionnel sont d'une importance cruciale. Malgré leurs limites pratiques, ils permettent d'assurer une certaine compréhension minimale par la personne affectée sur le processus décisionnel dont elle fait l'objet. Cette conscience de la logique du système peut l'aider à soupçonner un dysfonctionnement ou une utilisation néfaste du logiciel, d'agir en justice ou encore à demander à avoir un plus grand contrôle humain sur le modèle.

³²⁴ « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. Article 93.2.

³²⁵ *Ibid.* Article 24.

§2. Les mécanismes juridiques de contrôle humain sur la décision produite

Annonce du plan. Le contrôle humain est une garantie juridique importante dans le cadre des prises de décisions impliquant un algorithme et ce sous plusieurs aspects. Tout d'abord, le contrôle humain est un mécanisme permettant à l'opérateur humain, intervenant aux côtés du logiciel, de veiller au bon déroulement du processus décisionnel **(A)**. Ensuite, l'existence d'un contrôle humain permet aux personnes affectées par la décision de conserver une certaine maîtrise sur la prise de décision partiellement ou totalement automatisé. Le contrôle peut aussi être réalisé *a posteriori*, soit par la personne faisant l'objet de la décision, soit par le biais d'un intervenant humain tiers **(B)**.

A – Le contrôle humain pendant la prise de décision

L'intervenant humain comme condition d'interdiction. La présence d'un être humain au cours d'un processus décisionnel algorithmique a très vite été identifiée comme un élément central pour la préservation d'une autonomie humaine. Très tôt les législateurs ont admis qu'une prise de décision dépourvue d'intervenant humain entraînait des risques trop élevés pour les individus. En France, la loi informatique et libertés (LIL ci-après), dès sa première version de 1978, posait comme principe l'interdiction de recourir uniquement à un traitement automatisé pour certaines décisions critiques fondés sur le comportement d'un individu :

« Aucune décision de justice impliquant une appréciation sur un comportement humain ne peut avoir pour fondement un traitement automatisé d'informations donnant une définition du profil ou de la personnalité de l'intéressé.

Aucune décision administrative ou privée impliquant une appréciation sur un comportement humain ne peut avoir pour seul fondement un traitement automatisé d'informations donnant une définition du profil ou de la personnalité de l'intéressé.³²⁶ »

Dans le contexte des décisions de justice, le recours à ce type de traitement automatisé est totalement prohibé, l'expression « *seul fondement* » à l'alinéa 2 indique clairement qu'un

³²⁶ « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Article 2 (version abrogée)

intervenant humain est obligatoire pour la production de décisions administratives et privés. Cette disposition existe encore aujourd'hui à l'article 47 de la version actuelle de la LIL. Depuis la version de 1978, deux exceptions ont été ajoutées à l'alinéa 2 lors de l'adaptation du RGPD en droit français. La première transpose les exceptions européennes prévues dans le RGPD que sont le consentement et la conclusion ou l'exécution d'un contrat entre la personne concernée et le responsable de traitement³²⁷. La seconde résulte de l'utilisation par la France de sa marge de manœuvre accordée par le RGPD³²⁸. Il s'agit des décisions administratives individuelles qui peuvent être prises sans l'intervention d'un être humain³²⁹.

Contrairement à la France, le Canada impose la présence d'un intervenant humain dans le processus décisionnel d'un service administratif externe³³⁰. Cette obligation de participation humaine vaut pour tous les systèmes dès lors que les décisions produites peuvent avoir une incidence élevée sur l'administré³³¹.

L'intervenant humain comme condition d'accès à un droit par la personne concernée. Le RGPD reconnaît aussi l'importance de la présence humaine aux côtés de l'algorithme mais décide d'articuler cette préoccupation de façon un peu différente des deux précédents textes :

« La personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire. »³³²

Le texte subordonne le droit de la personne à l'absence d'acteur humain dans la prise de décision. Ce n'est pas un critère d'interdiction comme les législations canadiennes et françaises mais une condition négative. Malgré cette différence, le législateur reconnaît les risques d'une

³²⁷ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 22.2 a) et c).

³²⁸ « Le paragraphe 1 ne s'applique pas lorsque la décision : [...] b) est autorisée par le droit de l'Union ou le droit de l'État membre auquel le responsable du traitement est soumis et qui prévoit également des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée; ou » - Article 22.2.b) du RGPD.

³²⁹ « Aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative ne peut être prise sur le seul fondement d'un traitement automatisé de données à caractère personnel, y compris le profilage, à l'exception : [...] 2° Des décisions administratives individuelles prises dans le respect de l'article L. 311-3-1 et du chapitre Ier du titre Ier du livre IV du code des relations entre le public et l'administration » - Article 47 al. 2 de la LIL.

³³⁰ Article 6.3.9 directive canadienne

³³¹ Annexe C de la directive canadienne. Les systèmes décisionnels automatisés concernés sont ceux relevant du niveau 3 et 4 d'incidence.

³³² Article 22 alinéa 1 du RGPD. Emphase ajoutée par l'auteur.

décision entièrement automatisée et ce droit de refus expose le responsable de traitement à un refus explicite de la personne concernée³³³. De plus, l'absence d'une présence humaine entraîne le déclenchement des mécanismes de transparence spécifiques à ce type de traitement automatisé énoncées précédemment. Cette disposition peut également être vue comme une prérogative accordée à la personne concernée pour influencer indirectement la mise en place d'un participant humain par le responsable de traitement.

B – Le contrôle humain après la prise de décision

Le contrôle humain après la prise de décision en droit positif. Malgré ces dispositions juridiques imposant une intervention humaine dans le processus décisionnel, de nombreuses décisions sont prises quotidiennement sur le seul fondement d'un système algorithmique. Dans ces situations, les législations prévoient des mesures de protection sous la forme d'un contrôle humain après le rendu de la décision auprès de l'individu affecté. Ainsi lorsqu'un responsable de traitement se situe dans les exceptions de l'article 22 du RGPD, il doit mettre en place :

*« des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision. »*³³⁴

Sans pour autant entrer dans les débats doctrinaux sur ladite disposition, il en ressort que cette disposition renvoie au minimum à une révision humaine de la décision produite exclusivement sur la base d'un algorithme si la personne concernée en fait la demande. De la même manière, le droit offre à la personne concernée d'exprimer son point de vue et de contester la décision afin de garantir un certain rôle à la personne affectée par la décision, comme ce serait le cas si la décision avait été prise par un être humain.

Le DSA reprend la même logique concernant l'utilisation des algorithmes pour effectuer la modération des contenus illicites sur les plateformes en ligne. Si la qualification du contenu comme illicite et la sanction délivrée conséquemment peuvent être prises par un système algorithmique seul, cette possibilité s'arrête lors de la contestation par la personne sanctionnée :

³³³ Sauf les exceptions énoncées à l'article 22.2. et exposées plus en détails dans le paragraphe précédent.

³³⁴ Article 22.3 du RGPD. Emphase ajoutée par l'auteur.

« Les plateformes en ligne veillent à ce que les décisions visées au paragraphe 4 [décisions concernant les réclamations à une sanction] ne soient pas uniquement prises par des moyens automatisés.³³⁵ »

À nouveau le contrôle humain, bien qu'absent dans le processus décisionnel, est imposé en faveur de la personne affectée dans un second temps. Le contrôle humain est un outil essentiel pour préserver la maîtrise de l'humain sur des systèmes de plus en plus complexe. Déjà discutée lors de l'élaboration de ces articles, la question de l'efficacité de cet intervenant humain se pose à nouveau et avec encore plus de vigueur avec l'arrivée des technologies d'IA.

Sous-section 2 : Un besoin de dépasser les limites du droit positif

Annonce du plan. Malgré les solutions apportées par le droit positif, les dispositions juridiques souffrent de limites importantes, créées ou exacerbées par le recours à des technologies d'IA. Dans un premier temps, des limites textuelles viennent réduire la portée de ces dispositions, résultant en un cadre juridique éparse et lacunaire (§1). Dans un second temps, ces dispositions sont confrontées à des contraintes techniques non prévues qui mettent à mal leur application par les acteurs concernés. Cette inadéquation du droit avec la technique empire avec le déploiement des technologies d'IA qui apportent avec elles de nouveaux obstacles techniques (§2).

§1. Les limites relatives à la portée des dispositions juridiques existantes

Annonce du plan. Le corpus des dispositions sur le droit des décisions impliquant un algorithme regroupe des règles juridiques éparées sous une thématique commune. Ces obligations font parties de textes contenus dans des branches du droit différentes. Ces articles révèlent des limites face aux enjeux transversaux de l'IA. La portée lacunaire de ces outils trouve son origine dans la définition changeante de l'objet réglementé (A) et dans les différences de champ d'application des textes dans lesquels se trouve lesdites dispositions (B).

³³⁵ Article 17 du DSA : système interne de traitement des réclamations

A –La définition de la décision impliquant un algorithme

L’existence d’un impact sur les individus. La totalité des dispositions énoncées précédemment se concentrent sur les décisions qui ont des conséquences sur les droits et libertés des personnes affectées par la décision. Un résultat algorithmique qui n’intervient pas dans un processus décisionnel visant une personne ne sera pas concernée par ces obligations. En ce sens, le « *droit de ne pas faire l’objet d’une décision* » de l’article 22 du RGPD illustre clairement cette condition. Les lignes directrices produites par le groupe de travail sur la protection des personnes à l’égard du traitement de données à caractère personnel³³⁶ énoncent ainsi la notion de « prise de décision automatisée », ce qui renforce la présence de cette condition. Depuis l’arrêt SCHUFA rendue par la Cour de justice de l’UE le 7 décembre 2023, la notion de « décision » est interprétée extensivement et peut désormais désigner un score³³⁷, en l’espèce un score de solvabilité produit par un système algorithmique bancaire³³⁸. Pour reprendre l’interprétation du professeur Emmanuel Netter : « *la décision se caractériserait ainsi par ses effets, extrinsèques, plutôt que par des caractéristiques intrinsèques.* »³³⁹.

Dans le même sens, la directive canadienne sur la prise de décision automatisée s’applique exclusivement aux systèmes décisionnels automatisés produisant une recommandation ou une décision au sujet d’un client³⁴⁰. De plus, les résultats algorithmiques produits dans un environnement de test sont explicitement exclus³⁴¹. Enfin, les articles L. 311-3-1 et L. 312-1-3 du CRPA évoquent une « *décision individuelle prise sur le fondement d’un traitement algorithmique* ». Toutes ces dispositions conditionnent leur application à la présence d’une décision ayant un impact sur un individu.

La criticité de la décision algorithmique. Le risque encouru par les sujets d’une décision algorithmique est une condition d’application partagée par la plupart des textes. Cependant, le niveau de risque à partir duquel les obligations doivent être respectées est un point de

³³⁶ Aussi appelé G29. Il est désormais devenu le comité européen de la protection des données (ci-après CEPD).

³³⁷ NETTER Emmanuel, « Quand la force de conviction du scoring bancaire provoque sa chute. L’interprétation extensive, par la CJUE, de la prohibition des décisions entièrement automatisées », *RTD Com.*, 2024.

³³⁸ « *cette notion renvoie non seulement à des actes qui produisent des effets juridiques concernant la personne en cause mais également des actes qui affectent celle-ci de manière significative de façon similaire.* » - Cour de justice de l’Union européenne, 7 décembre 2023, n° C-634/21. Par. 44-46.

³³⁹ NETTER Emmanuel, *op. cit.* Voir aussi : DOUVILLE Thibault, « De l’approche extensive de la prise de décision exclusivement automatisée », *Recueil Dalloz*, 2024. NETTER Emmanuel, « La part de l’Homme et celle de la machine dans les décisions “ automatisées ”. Propositions pour une réécriture de l’article 22 du RGPD », dans BARBÉ Vanessa et MAUCLAIR Stéphane, *Vers un droit de l’algorithme ?*, Mare&Martin, 2022.

³⁴⁰ « Directive sur la prise de décision automatisée (Canada) », 2019. Article 5.2.

³⁴¹ *Ibid.* Article 5.3

divergence important. Dans le RGPD, les mécanismes de contrôle humain et de transparence énoncés précédemment ne sont disponibles que si la décision automatisée dépasse un certain seuil de criticité. Ce seuil est atteint lorsque la décision rendue produit à l'encontre de la personne concernée « *des effets juridiques [...] ou l'affectant de manière significative de façon similaire* ». Dans ses lignes directrices, le comité européen de la protection des données, organe de l'Union en charge de veiller à l'application cohérente du RGPD³⁴², indique que l'expression « *effets juridiques* » renvoie aux décisions automatisées qui affectent les droits juridiques ou le statut juridique d'une personne³⁴³. L'annulation d'un contrat ou encore le refus de la citoyenneté à un demandeur sont des exemples d'effets juridiques au sens de l'article 22. En ce qui concerne la seconde partie de la condition, cela renvoie à l'idée que la conséquence doit avoir un niveau d'importance pour la personne concernée équivalent à un effet juridique. Par exemple, « *le rejet automatique d'une demande de crédit en ligne ou des pratiques de recrutement en ligne [...]* » sont des situations qui affectent la personne concernée de manière significative similairement à un effet juridique. Il est important de rappeler que les considérants comme les lignes directrices du CEPD ne sont pas des textes contraignants juridiquement et ne servent qu'à éclairer les dispositions.

Une vision différente de la criticité de l'impact d'une décision algorithmique est proposée par la directive canadienne sur la prise de décision automatisée sous la forme d'une gradation. Dans son annexe B, le texte décrit quatre niveaux de risques possibles, calculés en fonction de quatre critères : la nature de l'incidence, sa gravité, sa durée et sa réversibilité. Les effets pris en compte sont ceux concernant les droits, la santé ou le bien-être, les intérêts économiques et la durabilité continue de l'environnement. Ainsi une décision avec aucun ou peu d'effets sur ces sujets et dont les effets seront brefs et réversibles sera située au premier niveau tandis qu'une décision aux impacts importants et irréversibles appartiendra au niveau quatre. Pour certaines exigences, leur intensité varie en fonction du niveau d'incidence des décisions automatisées. Par exemple, l'obligation de publier un avis informant du recours à un système automatisé n'est pas applicable si l'incidence de la décision est de niveau un. Le contenu de cet avis est plus dense pour les niveaux trois et quatre que pour le niveau deux. De la même manière, il est

³⁴² « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Articles 68 et 70.

³⁴³ Contrôleur européen à la protection des données, « Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679 », 2018.

obligatoire d'impliquer un être humain dans le processus décisionnel pour les niveaux trois et quatre alors que c'est parfaitement possible pour les niveaux un et deux.

Enfin, la dernière possibilité relative au niveau de criticité de la décision est de ne pas la prendre en compte. Les exigences permettant de protéger les personnes affectées sont obligatoires peu importe les conséquences de la décision. C'est le cas des obligations de transparence imposées par le code des relations entre le public et l'administration (CRPA) et le code de la santé publique (CSP) en France et plus généralement par le droit des plateformes numériques en France³⁴⁴ et dans l'Union européenne³⁴⁵. Dans ces situations, le niveau de risque n'est pas une condition d'application de la disposition spécifiques aux décisions automatisées. Ce sont des textes aux champs d'applications plus restreints qui ont pu ainsi identifier le niveau de risque moyen de leur secteur et par conséquent adapter les obligations du texte par ailleurs à cette criticité. L'exigence de criticité est alors indirecte car elle est dissimulée au travers d'un choix dans le champ d'application du texte.

Conclusion. Ces différents facteurs limitant l'application des exigences représentent autant de situations où les mécanismes de protection des droits et libertés des individus ne sont pas applicables. Au-delà de leurs divergences, l'établissement de ces limites et leur justification est remise en question avec l'utilisation toujours plus croissante des technologies d'IA. Les risques nouveaux ou intensifiés apportés par l'IA ne s'arrêtent pas aux limites posées alors par ces dispositions, lesquelles présentent des lacunes à combler.

B – Le champ d'application restreint des textes existants

La deuxième catégorie de facteurs limitatifs du droit des décisions algorithmiques est celle du périmètre du champ d'application des textes. Le champ d'application du texte ajoute ses propres limites à celles contenues dans les dispositions présentées précédemment.

La restriction aux algorithmes externes de l'administration. La Loi pour une république numérique de 2016 établit des règles spécifiques aux décisions administratives algorithmiques. Le but poursuivi est de protéger l'administré d'un préjudice lié au recours au système dans le

³⁴⁴ Voir notamment l'article L. 111-7 du code de la consommation.

³⁴⁵ « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022.

processus décisionnel. Cet objectif laisse par conséquent de côté les utilisations des algorithmes utilisés en interne dans l'administration. L'utilisation d'un système algorithmique pour analyser des données, élaborer une politique ou même prendre des décisions sur les employés, fonctionnaires et contractuels, de l'administration ne sera pas soumise aux règles prévues dans le CRPA. Cette limitation du périmètre est également présente dans la directive canadienne sur la prise de décision automatisée. Son application exclusive aux décisions automatisées fournies dans le contexte de « services externes » de l'administration exclut les utilisations des systèmes algorithmiques sur les employés, sur les candidats pour entrer dans l'administration et sur l'usage de ces technologies pour réfléchir à des orientations politiques. Lors de son examen biennuel de la directive, le comité du trésor canadien propose par ailleurs d'élargir le texte aux services internes de l'administration, en particulier sur les processus de recrutement³⁴⁶. Les risques générés par les technologies d'IA sont aussi présents lors de l'utilisation d'algorithmes en interne, la recherche d'exemplarité de l'administration ne doit donc pas s'arrêter aux relations avec le public. Les membres du comité canadien craignent ainsi que l'exclusion des algorithmes utilisés par les services internes nuise à l'objectif du gouvernement de renforcer la confiance des citoyens dans l'administration³⁴⁷. De la même manière, l'article 17 de la loi française relative à la bioéthique de 2021³⁴⁸ se concentre sur les systèmes algorithmiques utilisés « *pour un acte de prévention, de diagnostic ou de soin* ». À nouveau, les mécanismes en place ne peuvent pas être sollicités lors de l'utilisation interne des systèmes algorithmiques. Ainsi l'utilisation de systèmes algorithmiques pour d'autres tâches comme la gestion de l'hôpital ou le recrutement de personnel de santé tombe en-dehors du champ d'application. Concernant la directive canadienne en particulier, celle-ci connaît des limites supplémentaires du fait de sa nature juridique. Une directive canadienne n'est pas texte de « droit dur » mais plutôt de « droit souple ». L'objectif de ce type de textes est d'influencer la manière dont les décisions administratives sont prises. Ce genre de documents est contraignant en interne mais les sanctions qui en découlent restent elles aussi en interne³⁴⁹. Aucun droit n'est actionnable pour l'administré.

L'emphase mise sur la donnée. Bien que le droit des données ne se limite pas aux données personnelles, c'est longtemps uniquement sous ce prisme que se trouvent les mesures de

³⁴⁶ BITAR Omar, DESHAIES Benoit et HALL Dawn, « 3rd Review of the Treasury Board Directive on Automated Decision-Making », 2022.

³⁴⁷ *Ibid.*

³⁴⁸ « LOI n° 2021-1017 du 2 août 2021 relative à la bioéthique (1) », 2021-1017, 2021.

³⁴⁹ SCASSA Teresa, « Administrative Law and the Governance of Automated Decision-Making », *University of British Columbia Review*, 54, 2021.

protection de ces données imposées aux responsables de traitement. Par conséquent, les systèmes algorithmiques utilisant des données non personnelles, bien moins protégées³⁵⁰, étaient épargnés de la plupart des exigences de ce cadre alors qu'ils peuvent eux aussi entraîner des conséquences pour la société. Cette lacune en droit positif a été comblée récemment au niveau européen par le règlement sur la gouvernance des données³⁵¹ (ci-après DGA) et, plus secondairement, par le règlement sur les données (ci-après DA). Au travers du DGA, le législateur européen vise notamment l'utilisation nouvellement massive des algorithmes d'apprentissage automatique³⁵² sans pour autant imposer des obligations spécifiques à l'utilisation des données dans le contexte des technologies d'IA ou aux systèmes qui traiteront ces données.

Conclusion. Chacun de ces textes est élaboré avec des objectifs spécifiques pour un ou plusieurs problèmes identifiés dans un secteur d'activité délimité. Les solutions contenues dans ces textes sont donc nécessairement limitées par ce cadre. La protection des personnes affectées par les décisions algorithmiques se retrouve donc fragmentée, variant dans son intensité et dans son fonctionnement d'un texte à l'autre. De nombreux secteurs ne sont pas considérés. Pourtant les risques juridiques et éthiques liés aux technologies d'IA suivent son utilisation multisectorielle et ils sont donc présent à tout niveau. Si l'adaptation au contexte d'utilisation semble nécessaire, ces solutions présentent de nombreuses lacunes face aux risques transversaux de l'IA.

³⁵⁰ « Règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2018.

³⁵¹ « Règlement (UE) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724 (règlement sur la gouvernance des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022.

³⁵² « *Le présent règlement vise à contribuer à l'émergence de réserves de données mises à disposition selon le principe de l'altruisme en matière de données, qui soient d'une taille suffisante pour permettre l'analyse des données et l'apprentissage automatique, y compris au-delà des frontières de l'Union.* » - « Règlement (UE) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724 (règlement sur la gouvernance des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022, considérant 35.

§2. Les difficultés d'application : l'inadéquation entre le droit et la technique

Annnonce de plan. Comme son nom l'indique, le cadre juridique sur les décisions impliquant un algorithme porte sur les décisions, c'est-à-dire le résultat d'un processus décisionnel dans lequel est intervenu un système automatisé. Actuellement, le droit positif se concentre sur la l'utilisation du logiciel car c'est à ce moment-là qu'un dommage peut effectivement être causé. Ce choix d'encadrement se fait au détriment du logiciel lui-même dont le fonctionnement n'est jamais abordé précisément dans les obligations, en dehors d'exigences de transparence. Pourtant, le risque de préjudice par une décision automatisé (partiellement ou entièrement) est en grande partie dû au fonctionnement du système algorithmique utilisé. Il existe un lien fort entre le fonctionnement du système et les conséquences de la décision prise. Délaisse par les législateurs, le lien entre le texte juridique et l'outil produisant le contenu réglementé est donc toujours confié aux personnes en charge de l'application concrète de ces mécanismes. Or, l'absence quasi systématique d'éléments techniques dans les dispositions rend très difficile la qualification juridique des situations par les acteurs **(A)**. De la même manière, cette déconnexion avec la technologie abandonne les acteurs face à des obstacles techniques difficilement surmontables, voire à des dispositions juridiques inapplicables. Si cela évite une obsolescence technique apparente des textes avec l'utilisation des technologies d'IA, cela camoufle en réalité l'obsolescence dans les faits des solutions existantes **(B)**.

A – L'imprécision technique des termes juridiques

Annnonce de plan. L'informatique et le droit sont deux domaines distincts et tous les deux un vocabulaire propre. Si rien n'interdit au législateur de recourir à un vocabulaire technique, il est très rare qu'il exploite cette possibilité dans le cadre des systèmes de prise de décision. En effet, la cohabitation du langage informatique et juridique n'est pas sans risque. Il est difficile de trouver un équivalent parfait pour chaque terme et le risque d'obsolescence juridique n'est jamais loin. Les débiteurs des obligations juridiques sont donc contraints de régler ces enjeux de traduction au moment de l'application. Les difficultés démarrent dès l'identification de l'objet visé par les textes **(1)** et continuent pour comprendre la signification des obligations qui en découlent **(2)**.

1) La qualification juridique complexe des décisions concernées

La diversité des notions juridiques. Les dénominations des décisions impliquant un algorithme et les systèmes à l'origine de ces décisions sont aussi nombreux que diverses en droit positif. Le RGPD évoque de « *décision individuelle automatisée* » et de « *traitement automatisé* ». La LRN vise de « *décisions administratives algorithmiques* » alors que la loi bioéthique renvoie à l'expression de « *traitement algorithmique dont l'apprentissage a été réalisé à partir de données massives* ». Le code de la consommation français contient la formule de « *algorithmes informatiques* » dans le cadre du droit des plateformes en ligne. Il est très surprenant de constater que pour tous ces textes, aucune définition claire n'est fournie par le législateur. Le constat est encore plus marquant en ce qui concerne la définition des systèmes capables de produire lesdites décisions. Ce manque de définition est d'autant plus problématique qu'il concerne un objet central des dispositions en question. En effet chacune de ces définitions contribue significativement à établir le champ d'application des mécanismes et donc la portée de la protection qu'ils fournissent.

La décision algorithmique en droit positif. La LRN ne donne pas la définition de la décision algorithmique. Des précisions sont cependant apportées en-dehors du texte par Etalab, un organisme ministériel français³⁵³, sur la signification de la notion de décision algorithmique. L'algorithme est alors défini comme « *une suite finie et non ambiguë d'instructions permettant d'aboutir à un résultat à partir de données fournies en entrée.* »³⁵⁴. Ainsi, les décisions algorithmiques peuvent aussi bien résulter d'un traitement informatique que d'un traitement papier. Cette caractéristique élargit alors considérablement le contexte d'application de la notion par rapport aux autres textes qui se limitent à l'utilisation d'algorithmes numériques. De la même manière, le terme de « décision » est entendu au sens large, à savoir qu'il regroupe la prise de décision mais également l'aide à la décision. À nouveau, cette interprétation s'éloigne de celles retenues par d'autres textes qui porte sur la prise de décision impliquant un algorithme. Ces éclaircissements pertinents sont la conséquence des retours faits par les acteurs du secteur public. Ces derniers avaient fait notamment remonter un besoin essentiel de clarté dans les termes employés par la réglementation³⁵⁵.

³⁵³ Etalab est un département de la direction interministérielle du numérique en charge de l'ouverture, du partage et de la valorisation des données publiques : <https://www.etalab.gouv.fr/>

³⁵⁴ COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Comment permettre à l'Homme de garder la main ? Les enjeux éthiques de l'intelligence artificielle. », CNIL, 2017.

³⁵⁵ ENA, « Ethique et responsabilité des algorithmes publics », Etalab, 2019.

La décision automatisée en droit positif. La définition de la décision automatisée au sens du RGPD est composée de différents critères, chacun flou et sujet à interprétation. Bien que le comité européen de la protection des données (CEPD) ait tenté de donner une explication précise et des exemples concrets, son avis reste non contraignant et des interprétations divergentes persistent³⁵⁶.

Les conséquences applicatives du flou notionnel. Si le problème d'articulation a été relativement réglé entre les textes, la mise en application de ces derniers restent difficiles pour les débiteurs des obligations en question. Ainsi un rapport de juin 2019 dans le cadre de la mission Etalab révèlent une difficulté significative à identifier les « *traitements algorithmiques* » de l'administration³⁵⁷. Les interlocuteurs interrogés au cours de l'enquête fournissent des définitions différentes de la notion, notamment entre les services juridiques et les services informatiques. L'absence de définition légale rend par ailleurs très difficile la résolution de cette divergence sémantique.

Le cas de la LRN n'est pas un cas isolé. Bien que récemment précisée par la Cour de justice³⁵⁸, les débats sur ce qui constitue une décision individuelle automatisée au sens de l'article 22 du RGPD³⁵⁹ et ses multiples conditions imprécises ont un impact sur l'application des mécanismes par les responsables de traitement. Une autre illustration de ce problème d'identification de l'objet de la réglementation est survenue au sein d'un groupe de travail sur les algorithmes de la ville de New York. Ce groupe avait pour mission d'auditer les algorithmes publics dans le but de lutter contre le risque de biais algorithmiques et les discriminations³⁶⁰. Deux ans après sa constitution, un fort débat a divisé l'équipe sur la définition du terme « *algorithme* » et par conséquent sur la portée des compétences qui leurs étaient confiées par la loi³⁶¹. Comme pour la LRN, le rapport du groupe de travail insiste sur les difficultés apportées par ces lacunes

³⁵⁶ Pour les détails des conditions d'application de l'article 22.1 du RGPD, voir §1. Les limites relatives à la portée des dispositions juridiques existantes.

³⁵⁷ ENA, *op. cit.* P. 9

³⁵⁸ Cour de justice de l'Union européenne, 7 décembre 2023, n° C-634/21.

³⁵⁹ NETTER Emmanuel, « La part de l'Homme et celle de la machine dans les décisions “ automatisées ”. Propositions pour une réécriture de l'article 22 du RGPD », dans BARBÉ Vanessa et MAUCLAIR Stéphane, *Vers un droit de l'algorithme ?*, Mare&Martin, 2022. DOUVILLE Thibault, « De l'approche extensive de la prise de décision exclusivement automatisée », *Recueil Dalloz*, 2024. NETTER Emmanuel, « Quand la force de conviction du scoring bancaire provoque sa chute. L'interprétation extensive, par la CJUE, de la prohibition des décisions entièrement automatisées », *RTD Com.*, 2024.

³⁶⁰ <https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>

³⁶¹ LECHER Colin, « New York City's algorithm task force is fracturing », sur *The Verge*, publié le 15 avril 2019.

définitionnelles et recommande fortement d'élaborer des critères de qualifications pour les systèmes algorithmiques ainsi que des éléments de hiérarchisation par importance³⁶².

Ces exemples mettent en avant l'importance centrale de produire une définition claire de l'objet du droit pour l'efficacité des mécanismes à l'œuvre. Élaborer une définition unique et précise représente un défi qu'aucun texte contraignant n'a permis de surmonter. Par conséquent, un texte juridique sur les systèmes d'IA pourrait accomplir cette tâche et faciliter considérablement l'application de ces dispositions, soit en propageant sa définition dans les autres textes, soit en servant de référentiel pour que chaque texte précise sa propre définition.

2) La portée incertaine des exigences juridiques.

En plus des difficultés de qualification, le manque de clarté rend très difficile la compréhension des obligations auxquelles sont soumis les utilisateurs de ces systèmes. Celles-ci sont exprimées en termes vagues tout en désignant des objets techniques. La traduction de formulations juridiques déjà floues dans le vocabulaire informatique moins enclins à ce genre de formulation est un travail extrêmement complexe pour les acteurs chargés de leur mise en œuvre.

L'absence de clarté dans les obligations de transparence. Dans ses mécanismes de transparence, situés aux articles 13. f), 14.g) et 15.h), le RGPD impose notamment de communiquer « *des informations utiles concernant la logique sous-jacente* ». Aucun détail supplémentaire du sens de la notion de « *logique sous-jacente* » n'est présent dans le texte. Le CEPD explique que la logique sous-jacente ne signifie pas « *nécessairement une explication complexe des algorithmes complets ou la divulgation de l'algorithme complet*³⁶³ ». Même si ce ne sont que des recommandations, cette description par la négative met en avant le fait que la portée du mécanisme de transparence est limitée, tout en ne proposant pas d'endroit de moyens d'établir le niveau de transparence attendu. Cette incertitude ne vient pas seule, une autre question se pose alors : de quelle logique est-il question ? Le texte pourrait tout autant cibler la logique sous-jacente du modèle, renvoyant alors à une transparence globale, ou la logique sous-

³⁶² AUTOMATED DECISION SYSTEMS TASK FORCE, « New York City Automated Decision Systems Task Force Report », 2019, P. 26.

³⁶³ Contrôleur européen à la protection des données, « Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679 », 2018, P. 28.

jacente de la décision individuelle subie par la personne, plutôt synonyme d'une transparence locale³⁶⁴. Chacune de ces options renvoie à une solution pratique différente.

La LRN contient une autre approche dans le détail de l'information attendue. Au travers d'un décret de 2017, elle précise à son article R. 311-3-1-2 du CRPA les éléments techniques précis qui composent toute réponse à l'intéressé en vertu de l'article L.311-3-1 du CRPA. Cette prise de position par le législateur apporte une certitude importante pour les personnes en charge de son exécution. De ce fait, la disposition du CRPA fait office d'inspiration pour d'autres législateurs qui y voient une solution à leurs problématiques³⁶⁵.

Toutefois, cibler des éléments techniques précis n'est pas une solution parfaite. Il existe une différence majeure entre cibler techniquement les outils et produire des critères précis de réalisation et de contrôle d'une obligation. En ce sens, le droit des plateformes en est une bonne illustration. Lors d'une étude en 2018, il a été révélé que les mécanismes de transparence relatifs aux publicités publiées sur la plateforme Facebook produisaient des informations souvent incomplètes et parfois même trompeuses³⁶⁶. Ainsi ce sont les critères les plus fréquents qui sont communiqués aux internautes et non les plus impactant dans la personnalisation de la publicité en question. L'absence de critères pour sélectionner les caractéristiques pertinentes rend impossible la garantie d'un seuil de qualité, et donc d'utilité de l'information communiquée. Dans cette illustration, c'est l'interprétation de la notion floue de pertinence qui a conduit à ce résultat. Cette lacune a laissé une liberté de choix dans l'exécution qui permet à la personne de réaliser l'obligation sans satisfaire l'objectif sous-jacent. Il a notamment été montré qu'il était parfaitement possible d'adapter les algorithmes chargés de l'explication afin qu'ils personnalisent l'explication pour les individus et évitent, au cas par cas, la communication d'informations pertinentes mais potentiellement préjudiciables pour l'utilisateur de l'algorithme³⁶⁷. Lister les éléments attendus dans une obligation n'est pas suffisant seul pour combler une exigence définie de façon floue. Dans ce cas, l'exigence sera définie lors de son application concrète par son débiteur. Il sera alors très difficile de démontrer que l'application de l'exigence à dénaturer l'exigence de départ, celle-ci n'ayant pas été définie.

³⁶⁴ EDWARDS Lilian et VEALE Michael, « Enslaving the Algorithm », *IEEE Security & Privacy*, 16, 2018.

³⁶⁵ BITAR Omar, DESHAIES Benoit et HALL Dawn, « 3rd Review of the Treasury Board Directive on Automated Decision-Making », 2022.

³⁶⁶ Athanasios Andreou et al, *Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations*, 2018.

³⁶⁷ MERRER Erwan Le et TRÉDAN Gilles, « Le problème du videur », *Interstices*, 2021.

Conclusion. L'étude de ces différents obstacles pratiques met en lumière l'importance de la précision technique dans la mise en application des mécanismes relatifs aux décisions impliquant un algorithme. À l'exception de la LRN, les textes présentés n'ont pas incorporé de lexique technique, une position préjudiciable qui oblige un travail de longue durée de la part d'organismes afin de préciser à la place du législateur les situations concernées par la loi et l'ampleur des mécanismes à l'œuvre. La précision des termes juridiques est indispensable pour faciliter leur traduction informatique lors de l'application. Ce besoin de clarté ne doit pas se limiter aux seules définitions mais doit également être pris en compte dans la formulation des obligations. Ce n'est qu'après ce long processus que les acteurs peuvent commencer la mise en œuvre des obligations, une phase qui contient elle aussi son lot d'obstacles.

B – L'aggravation des difficultés d'exécutions des obligations dans le contexte des technologies d'IA

Annonce de plan. Au-delà de ces problèmes de définitions, les spécificités de l'IA ajoutent une couche de difficultés concernant la faisabilité de ces dispositions par leurs débiteurs. Les solutions techniques qui ont pu être imaginées par les acteurs (entreprises, les autorités de régulation et autres personnes concernées) sont mises à mal par les technologies d'IA. En reprenant les mécanismes de droit positif présentés précédemment, il s'agira de montrer les complications rencontrées par les mécanismes de transparence dans le cadre d'un système d'IA **(1)**. Puis il sera observé que ces difficultés impactent aussi le contrôle humain sur la décision, que celui-ci soit exercé par l'utilisateur du système ou bien par une autorité de régulation **(2)**.

1) L'opacité problématique des technologies d'IA

L'augmentation de la complexité technique. La transparence et l'explicabilité représentent un principe et un mécanisme essentiels pour l'accomplissement de nombreux autres principes. Toutefois la réalisation de cette exigence connaît de nombreux obstacles lors de son application aux systèmes d'IA. Premièrement, la complexité des tâches réalisées par le système algorithmique qui incorpore de l'IA. L'avantage des systèmes d'IA par rapport aux systèmes plus traditionnels repose sur leur capacité à établir des corrélations entre des éléments qui seraient invisibles pour des êtres humains, à proposer des liens dans des situations qui semblent

chaotiques. Par conséquent, les tâches pour lesquelles les systèmes d'IA sont et seront le plus utilisés ne sont pas des tâches simples à expliquer. Dès le départ, l'obligation de produire une explication ou une information claire relative aux décisions prises peut être complexifiée par la difficulté d'expliquer la tâche en question, qu'elle soit ou non réalisée informatiquement³⁶⁸. Avec les technologies d'IA, on assiste à une augmentation de la quantité de caractéristiques prises en compte pour produire le résultat. Certains systèmes peuvent prendre en compte des dizaines voire des centaines de milliers de caractéristiques. Par exemple, l'entreprise LinkedIn expliquait en 2014 que son modèle d'apprentissage prenait plus de 100.000 variables en compte³⁶⁹. Cette augmentation du nombre de critères met en conflit l'exhaustivité de l'explication ou de l'information et sa compréhensibilité³⁷⁰. Le débiteur de l'obligation de transparence doit réaliser des choix encore plus difficiles pour sélectionner les variables les plus pertinentes pour obtenir une explication ou une information satisfaisante au regard de la loi. D'une part, ce travail de sélection des critères pertinents expose aux risques de tromperie discutées précédemment. D'autre part, cela conditionne la réalisation par l'utilisateur du système à la méthode et aux choix réalisés en amont par le concepteur. L'utilisateur ne pourra satisfaire son obligation de transparence uniquement si le système d'IA qu'il a acheté ou qu'il exploite est suffisamment interprétable.

2) Le contrôle difficile des technologies d'IA

La capacité de contrôle de l'intervenant humain. L'interaction entre l'humain et la machine a toujours été un enjeu important dans l'utilisation de l'informatique. L'autonomie importante des systèmes d'IA entraîne un risque de perte de contrôle effectif de la personne sur la décision. Une cause bien connue de cette diminution de la maîtrise de l'humain est le biais d'automatisation qui entraîne le participant humain au processus décisionnel à donner progressivement sa confiance à l'outil décisionnel qu'il utilise, au point d'y accorder une confiance aveugle et absolue. Les simples recommandations de la machine deviennent alors de réelles prises de décision. Les raisons d'utilisation des algorithmes, souvent pour augmenter la productivité en réduisant le temps nécessaire pour prendre une décision, décuple la présence de

³⁶⁸ EDWARDS Lilian et VEALE Michael, « Slave to the Algorithm? », *Duke Law Technology Review*, 16, 2017.

³⁶⁹ KUN LIU, DEVELOPING WEB-SCALE ML AT LINKEDIN—FROM SOUP TO NUTS, PRESENTED AT THE NIPS SOFTWARE ENGINEERING FOR MACHINE LEARNING (Dec. 13, 2014).

³⁷⁰ « Understanding algorithmic decision-making », Parlement Européen, 2019.

ces biais et aggravent encore plus la recherche d'une décision maîtrisée³⁷¹. Les mécanismes de contrôle humain prévus dans le droit positif font alors face à des difficultés d'application d'un niveau non anticipé avant l'utilisation des technologies d'IA.

L'article 22 du RGPD précise ainsi que les mesures de protection s'appliquent seulement si la décision est « *fondée exclusivement sur un traitement automatisé, y compris le profilage, [...]* ». Par conséquent, si un être humain intervient dans le processus décisionnel, la décision tombe hors de la portée de la disposition. La pertinence de l'intervention humaine nécessaire pour en arriver à ce résultat est sujet à discussion. Le CEPD déclare dans ses lignes directrices que l'intervenant humain doit assurer un « *contrôle significatif* ». Mais ces lignes directrices ne sont pas contraignantes et d'autres interprétations ont déjà été effectuées. Ainsi la SCHUFA, l'agence générale de protection du crédit allemande, a déclaré que l'implication nominale d'un être humain à tout moment du processus automatisé constituait une intervention humaine au sens de l'article 22 du RGPD³⁷².

La caractéristique « significatif » de l'action de l'être humain dans le processus décisionnel représente la délimitation entre le remplacement de l'humain par la machine ou le système comme support décisionnel de l'humain. Si l'individu ne se contente que de suivre aveuglément les recommandations ou les résultats produits par le logiciel, il n'est alors qu'une fiction pouvant servir à faciliter les questions de responsabilité ou à rassurer les personnes affectées, mais ne représente en aucun cas une quelconque garantie juridique. Il est essentiel de réaliser que la perte d'une intervention significative fait basculer la situation dans une délégation de prise de décision au système algorithmique. Le preneur de décision n'est plus l'être humain. Or, son statut de décideur peut être le fruit d'une procédure juridique, vote ou nomination par exemple. Placer un système d'IA en remplacement ne peut pas simplement se justifier par des raisons économiques ou de praticité. Dans ces situations, l'automatisation significative de la prise de décision doit possiblement offrir des garanties particulières sous peine de tomber dans l'illégalité. Le secteur public offre de nombreuses situations pouvant rencontrer ce problème. Plusieurs auteurs s'accordent sur l'idée qu'une délégation à la machine d'une fonction occupée par une personne nommée n'est pas légale. Le système algorithmique ne peut être utilisé qu'en

³⁷¹ CUMMINGS M. L., Automation Bias in Intelligent Time Critical Decision Support Systems, *AIAA 3rd Intelligent Systems Conference*, AIAA, 2004.

³⁷² WACHTER Sandra, MITTELSTADT Brent et FLORIDI Luciano, « Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation », *International Data Privacy Law*, 1, 2017.

tant qu'aide à la décision pour le décideur nommé³⁷³. La directive canadienne impose un intervenant humain en fonction du niveau d'incidence pour la personne affectée. Le statut du décideur remplacé n'est pas explicitement envisagé comme élément contenu dans l'analyse de l'incidence³⁷⁴. La problématique n'est donc pas abordée explicitement par la directive à l'heure actuelle³⁷⁵. Le droit français semble avoir mieux anticiper l'enjeu de la substitution du décideur nommé par le système algorithmique en interdisant par exemple la prise de décision exclusivement par algorithme dans le cas des décisions de justice sans aucune exception envisageable³⁷⁶. Au travers de la loi française de programmation 2018-2022 et de réforme de la justice³⁷⁷, le même régime d'interdiction s'applique pour les modes alternatifs de résolution des conflits en France que sont les services de conciliations, de médiation et d'arbitrage³⁷⁸. Toutefois, ces dispositions pourraient bénéficier d'améliorations en ce qu'elles manquent elles aussi d'un critère d'intervention significative afin d'éviter une présence humaine illusoire.

La capacité de contrôle des autorités de régulation. La problématique de contrôle se pose également pour l'intervention de personnes extérieures à l'utilisateur du système, en particulier les autorités de contrôle. La vérification d'un système d'IA peut s'avérer bien plus difficile que celle d'un algorithme déterministe et nécessite des compétences spécifiques pour les agents de contrôle qu'ils ne possèdent pas ou du moins n'en avaient pas la nécessité auparavant. Les autorités de contrôle et autres institutions peuvent toujours acquérir ces nouvelles compétences requises par le biais de recrutements d'experts. Cependant les technologies d'IA sont un domaine émergent et les experts du domaine travaillent déjà bien souvent dans le secteur privé. Les autorités de contrôle doivent donc entrer dans une concurrence féroce avec le secteur privé et/ou la recherche pour acquérir cette expertise³⁷⁹, un combat bien souvent perdu en raison des différences d'attractivité, notamment sur le plan financier³⁸⁰. De plus les technologies d'IA n'ont pas de secteur de prédilection. Par conséquent, les institutions concernées par le contrôle

³⁷³ "Administrative law and the machines of government: judicial review of automated public-sector decision-making", (2019) 39 Legal Studies 636, P. 644-645. Makoto Hong Cheng & Hui Choon Kuen, "Towards a Digital Government: Reflections on Automated Decision-Making and the Principles of Administrative Justice", (2019)

³⁷⁴ « Directive sur la prise de décision automatisée (Canada) », 2019, annexes B et C.

³⁷⁵ BLAKE Sara, *Administrative Law in Canada*, LexisNexis Canada, 2017, P. 153.

³⁷⁶ « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978, article 47 alinéa 1.

³⁷⁷ « LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice », 2019-222, 2019.

³⁷⁸ *Ibid.* Article 4.3

³⁷⁹ LYNDON Mary L., « Tort Law and Technology Response », *Yale J. on Reg.*, 12, 1995. P. 157-158 : « Agencies do not have a monopoly on expert knowledge. Neither courts nor agencies possess expertise when technologies are new Or, changing. At early stages of innovation, only researchers involved in R&D have this knowledge. »

³⁸⁰ SCHERER Matthew U., « Regulating Artificial Intelligence Systems », *Harvard Journal of Law & Technology*, 29, 2015.

des prises de décisions impliquant un algorithme sont elles aussi multiples. D'une part, cela accentue la difficulté à trouver des experts qui doivent non seulement avoir une compétence en IA mais également dans le secteur d'activité visé par la compétence de l'autorité. D'autre part, cela signifie que l'autorité n'est pas dédiée à la question des prises de décisions par des systèmes d'IA. Sa mission ne se limite pas aux risques des technologies d'IA mais englobe l'entièreté d'un secteur. Le temps et le budget octroyables à ces nouveaux risques se retrouvent donc nécessairement limités par les activités déjà réalisées par l'institution. Par conséquent, les experts recrutés, le cas échéant, se retrouvent en surcharge de travail par rapport aux systèmes décisionnels à contrôler³⁸¹. Les réglementations éparses sur les décisions algorithmiques entraînent logiquement un éparpillement des ressources et des politiques publiques sur la question de l'IA. Sans cadre réglementaire spécifique, il est peu probable que des budgets et des missions spécifiques soient octroyés aux autorités de contrôle.

CONCLUSION DE LA SECTION II :

En l'attente de règles spécifiques à l'IA, le droit positif contient tout de même un certain nombre de dispositions applicables à ces nouvelles technologies. L'étude de cet environnement juridique a permis d'identifier plusieurs besoins importants à prendre en compte lors de l'analyse de la réglementation relative à l'IA. Dans un premier temps, le droit positif révèle un besoin d'articulation important entre les exigences concernant l'utilisation des systèmes algorithmiques et le cadre juridique de l'IA. Cette articulation est indispensable pour permettre une adaptation et une application facilitée du droit positif envers les spécificités des technologies d'IA. Une seconde leçon issue du droit positif est l'importance d'utiliser des termes juridiques précis dans l'encadrement de technologies informatiques. Que ce soit en droit français, canadien ou européen, le recours à des notions juridiques imprécises entraîne systématiquement des difficultés importantes d'application pour les acteurs concernés. Les conséquences de ce flou notionnel affectent aussi bien les débiteurs des obligations que ceux en charge de leur contrôle. L'identification de ces deux besoins de forme, que sont l'articulation et la précision des exigences, vient ainsi renforcer les conclusions établies à la suite à l'analyse des textes éthiques.

³⁸¹ À titre d'exemple, le défenseur des droits en France, institution en charge de nombreuses missions et notamment de la lutte contre les discriminations ne possède qu'un seul agent dédié à la question des discriminations algorithmiques pour tout le pays.

CONCLUSION DU CHAPITRE 1 :

Entre les textes éthiques et le droit positif, l'environnement normatif relatif aux technologies d'IA se caractérise par une diversité et une richesse importante de règles. L'étude de ces normes a mis en avant des inspirations à poursuivre mais aussi des limites à éviter pour l'encadrement de l'IA. Malgré leur différence de nature, des besoins et des enjeux similaires émergent de ces textes. À l'aide de ces éléments, il est possible de constituer une grille d'analyse du droit de l'IA. Sur le fond, les besoins de transparence, d'équité, de protection de la vie privée, de règles de responsabilité, de contrôle humain et de fiabilité technique des logiciels d'IA sont partagés autant par le droit que par l'éthique. Le second apport de cette analyse se trouve dans l'identification de besoins sur la forme de l'encadrement. Dans le contexte des technologies informatiques, les analyse de l'éthique et du droit positif ont mis clairement en lumière l'importance de la précision des termes juridiques employés. Ainsi, l'éthique de l'IA permet de constater que des concepts imprécis sont interprétés, appropriés et mis en pratique de façon non uniforme par les acteurs. Le droit des décisions algorithmiques conforte ce constat en montrant que l'imprécision des termes conduit toujours à une application, soit partielle, soit absente, des exigences. Un autre élément central de notre prisme d'analyse est l'articulation desdites exigences. Cette articulation s'envisage alors sous deux aspects avec, d'un côté, l'articulation au sein-même des exigences spécifiques à l'IA et, de l'autre côté, l'articulation entre ces nouvelles règles juridiques et les anciennes. En plus des enjeux de fond, la précision des notions juridiques employées, leur articulation entre elles et avec d'autres notions sont des éléments que l'éthique et le droit positif ont fait ressortir comme essentiels dans l'encadrement de l'IA. Par conséquent, ces éléments sont constitutifs de notre grille de lecture. C'est donc sous ce prisme que seront identifiés et étudiés le ou les éléments juridiques centraux dans la future réglementation de l'IA.

CHAPITRE 2 : L'ADOPTION DE NORMES JURIDIQUES SPECIFIQUES AUX SYSTEMES D'IA

Introduction. Face aux lacunes du droit positif, c'est sans surprise que la Commission européenne décide d'intégrer dans sa politique de renforcement de la souveraineté numérique de l'Europe lors de la « décennie numérique » les impacts de l'IA³⁸². Débuté en 2018 par une communication sur de l'IA dans la politique européenne³⁸³, l'axe stratégique de l'IA se poursuit par de nombreux rapports commandés par la Commission elle-même³⁸⁴, des résolutions du Parlement européen³⁸⁵ ainsi que des avis du groupe d'experts de haut niveau sur l'IA (HLEG sur l'IA ci-après)³⁸⁶. Ces différentes contributions ont abouti à l'élaboration d'un livre blanc sur l'IA en 2020³⁸⁷. Ce livre blanc constitue la dernière étape de la Commission européenne avant les propositions de législations³⁸⁸.

Annnonce de plan. L'Union européenne a décidé de proposer un régime juridique spécifique pour encadrer les technologies d'IA. L'apparition de ces dispositions spécifiques conduit à étudier et analyser la manière dont la Commission européenne propose de réglementer ces technologies. Dans un premier temps, il appert qu'elle fait le choix de poser un régime juridique transversal applicable à tous les systèmes d'IA. Pour autant, la lecture du règlement révèle rapidement que l'attention de la Commission est concentrée sur les systèmes d'IA à haut risque (**Section I**). Le règlement européen sur l'IA dévoile la mise en place d'exigences de conception imposée aux systèmes d'IA européens dans le but de garantir des objectifs d'excellence et digne de confiance. À l'aide de la grille d'analyse élaborée précédemment, les exigences relatives au fonctionnement des système d'IA seront étudiées afin d'examiner dans quelle mesure le texte

³⁸² « Les priorités de la Commission », sur *Commission.europa.eu*, publié le 9 février 2023.

³⁸³ *Communication de la Commission européenne au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions - L'intelligence artificielle pour l'Europe*, 2018.

³⁸⁴ Voir par exemple : ZUIDERVEEN BORGESIU Frederik, « Discrimination, Artificial Intelligence and Algorithmic Decision-Making », Conseil de l'Europe, 2018.

³⁸⁵ Voir notamment les résolutions sur le responsabilité de l'IA et l'éthique de l'IA : PARLEMENT EUROPÉEN, « Un régime de responsabilité civile pour l'intelligence artificielle », 2020 PARLEMENT EUROPÉEN, « Cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes », 2020.

³⁸⁶ Le HLEG sur l'IA a élaboré jusqu'à présent quatre livrables : « High-level expert group on artificial intelligence | Shaping Europe's digital future », publié le 6 février 2023.

³⁸⁷ COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020.

³⁸⁸ Sont visées ici la proposition de la Commission européenne sur les systèmes d'IA ainsi que les deux propositions sur la responsabilité du fait des systèmes d'IA.

répond directement aux besoins de l'IA et dans quelle mesure leur satisfaction est déléguée aux normes techniques (**Section II**).

Section I : Des règles juridiques centrées sur les systèmes d'IA à haut risque

Annonce de plan. Les réponses juridiques disponibles pour faire face aux défis de l'IA sont multiples et impliquent donc de réaliser des choix forts quant à l'encadrement de ces technologies. Par exemple, faire évoluer séparément les différentes branches du droit pour les adapter aux évolutions apportées par l'IA était une possibilité. Avec le règlement européen sur l'IA, l'Union européenne réalise ainsi deux décisions significatives quant aux caractéristiques de cette réglementation. Premièrement, elle décide d'encadrer spécifiquement les technologies d'IA en choisissant comme objet juridique central le système d'IA (**Sous-section 1**). Deuxièmement, l'Union européenne reconnaît les difficultés pratiques et les conséquences néfastes pour l'innovation qu'entraîne un encadrement de tous les systèmes d'IA. Par conséquent, elle adopte une approche par les risques et concentre son attention sur les systèmes d'IA dont l'usage expose les individus à des risques considérés, au minimum, comme élevés (**Sous-section 2**).

Sous-section 1 : Le système d'IA comme objet de réglementation

Annonce de plan. Avec le règlement sur l'IA³⁸⁹ et les propositions de directives relatives à la responsabilité de l'IA³⁹⁰, l'Union européenne choisit de se concentrer sur les technologies d'IA. Instaurer le système d'IA comme objet de droit est une décision audacieuse. L'encadrement d'une technologie particulière est une stratégie justifiable dès lors que cette technologie présente des spécificités importantes pour le droit. Ainsi, une partie de la doctrine et la Commission européenne justifient la pertinence du règlement sur l'IA en énumérant certaines de ses caractéristiques (**§1**). L'étape suivante est alors de définir la notion de système d'IA.

³⁸⁹ « Règlement sur l'intelligence artificielle », 2024.

³⁹⁰ Sur l'évolution du régime de responsabilité des produits défectueux pour intégrer les problématiques de l'IA, voir Commission européenne, *Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux*, 28 septembre 2022. Sur la responsabilité extracontractuelle pour un dommage causé par un système d'IA, voir *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022.

Cette-dernière représente le critère central du champ d'application du texte. Issu de la discipline informatique, ce terme porte avec lui un flou sémantique que le droit devra résoudre pour répondre aux besoins de précision et d'articulation du règlement (§2).

§1. Le choix de réglementer spécifiquement les technologies d'IA

Annonce du plan. La décision d'encadrer spécifiquement les technologies d'IA au travers de nouveaux textes est donc la stratégie choisie par la Commission européenne parmi les autres possibilités à disposition. Dès lors, il est intéressant de comprendre dans quelle mesure un tel choix peut être juridiquement justifiable. Les enjeux juridiques autour de cette idée font l'objet de débats depuis longtemps. Au travers d'une étude de la doctrine, plus particulièrement des écrits relatifs au droit de la robotique, nous mettrons en évidence l'importance d'identifier le caractère juridiquement disruptif d'une nouvelle technologie pour justifier son encadrement spécifique (A). Une fois le concept identifié et expliqué, nous nous attarderons sur les arguments utilisés par la doctrine et par la Commission européenne pour justifier l'intérêt de cet encadrement spécifique des technologies d'IA (B).

A- La justification de l'encadrement spécifique d'une technologie

Le choix significatif d'encadrer transversalement une ou plusieurs technologies. Préalablement à toute initiative, il est pertinent pour le législateur de s'interroger sur l'existence d'un ou plusieurs besoins intrinsèques à la technologie justifiant le recours à des dispositions nouvelles et spécifiques. En l'absence d'arguments en ce sens, le choix d'un texte spécifique peut se révéler dispensable et d'autres solutions moins coûteuses et moins complexes peuvent alors être envisagées. Par exemple, il existe une stratégie juridique alternative qui consiste à repérer dans chaque secteur du droit les insuffisances et réfléchir uniquement à des améliorations *ad hoc*.

Cela n'a pas été le choix de la Commission européenne. En proposant sa stratégie européenne sur l'IA, elle a décidé d'ajouter aux évolutions du droit positif la création d'un ensemble de règles spécifique aux technologies d'IA. Ce choix juridique et politique s'illustre particulièrement au travers de la proposition d'un règlement européen 2021/0106 établissant

des règles harmonisées concernant l'intelligence artificielle³⁹¹. En matière d'IA, l'Union européenne décide donc à la fois de faire évoluer son droit positif³⁹² et de créer un cadre juridique nouveau spécifique aux technologies, les deux réponses juridiques étant complémentaires l'une de l'autre³⁹³. Approcher de façon globale la question des technologies d'IA permet de maintenir la réflexion sur les conséquences de l'IA sur un temps plus long. En effet, résoudre des défis entre branches du droit plutôt que des problèmes plus précis relève d'une plus grande complexité. Cette durée étendue des travaux législatifs permet de dépasser la simple anticipation et d'être encore réactif au moment du déploiement de la technologie et de ses conséquences sur les relations sociales. Cette chronologie a été avancée par la professeure Lyria Bennet Moses qui explique que les problèmes juridiques sont réglés préalablement à la démocratisation de l'utilisation desdites technologies³⁹⁴. Or, les étapes d'application exponentielle d'une technologie et les réactions que cela produit dans les relations sociales sont deux événements générateurs de défis juridiques. L'approche au cas par cas présente le risque de penser que les problématiques juridiques posées par une technologie sont définitivement résolues parce que le dernier symptôme en date vient d'être réglé³⁹⁵. Au contraire, l'approche globale allonge la durée de vigilance et permet de maintenir la réflexion pour les moments clés survenant après le déploiement de la technologie.

L'importance d'identifier le caractère disruptif juridiquement d'une technologie. Dès lors qu'une nouvelle technologie vient poser des difficultés d'application au droit positif, le besoin de nouvelles règles est soulevé. Les juristes se sont alors interrogés sur les caractéristiques d'une technologie permettant de justifier l'élaboration de dispositions juridiques spécifiques³⁹⁶. L'objectif sous-jacent est d'identifier dans quelles mesures une nouvelle réglementation

³⁹¹ *Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (Législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union*, 2021.

³⁹² Voir notamment les propositions de directives relatives à la responsabilité dans le contexte de l'IA ainsi que les documents ayant contribué à ces propositions : *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022. COMMISSION EUROPÉENNE, *op. cit.* PARLEMENT EUROPÉEN, « Un régime de responsabilité civile pour l'intelligence artificielle », 2020.

³⁹³ Ainsi les deux propositions européennes de directives concernant la responsabilité des systèmes d'IA réutilisent la terminologie et la notion de conformité élaborée dans le règlement européen sur l'IA. Elles apportent également des dispositions particulières sur la responsabilité qui viennent en complément d'un règlement sur l'IA grandement centré sur la conception des systèmes d'IA.

³⁹⁴ MOSES Lyria, « Why Have a Theory of Law and Technological Change? », *Minnesota Journal of Law, Science & Technology*, vol. 8, n° 2, 1 janvier 2007, p. 589.

³⁹⁵ LIU Hin-Yan, MAAS Matthijs M., DANAHER John, SCARCELLA Luisa, LEXER Michaela et VAN ROMPAEY Léonard, « Artificial Intelligence and Legal Disruption », *op. cit.*

³⁹⁶ MERABET Samir, *Vers un droit de l'intelligence artificielle*, 2020.

spécifique à une technologie est pertinente. En effet, il existe une tendance à penser les problématiques issues d'une nouvelle technologie uniquement au travers de ses mauvaises utilisations³⁹⁷. Si les difficultés sont causées par l'usage et non par les attributs de la technologie, l'élaboration d'une réglementation spécifique peut se révéler dommageable. Selon Susan W Brenner, cela risque de produire des lois redondantes et compartimentées³⁹⁸, ce qui accroît fortement la possibilité d'obsolescence³⁹⁹. Afin d'éviter ces écueils, il est central d'identifier des problèmes juridiques intimement liés à la technologie en elle-même, au-delà du bon ou mauvais usage qui en ait fait. Par conséquent, la qualification d'une technologie comme profondément disruptive juridiquement devient un argument central pour les défenseurs d'un régime juridique propre à certaines technologies⁴⁰⁰.

La démarche d'identification des spécificités disruptives juridiquement d'une technologie a été réalisée à plusieurs reprises par la doctrine. L'ensemble de ces travaux met alors en avant l'existence de différentes approches visant à identifier ce caractère disruptif⁴⁰¹. Dans le contexte de l'IA, une illustration particulièrement intéressante de ces démarches est celle du droit de la robotique. D'une part, les spécificités disruptives juridiques attribuées aux robots se retrouvent, du moins en partie, dans les technologies d'IA. D'autre part, les deux approches sélectionnées ont été appliquées sur une conception large de la robotique qui s'étend partiellement à l'IA⁴⁰². En l'espèce, l'objectif est de mettre en lumière, au travers de deux exemples, les outils théoriques permettant de révéler les arguments en faveur d'un encadrement d'une technologie particulière. Pour illustrer ces tentatives et leurs divergences théoriques, nous utiliserons deux propositions relatives l'encadrement de la robotique : la théorie des « qualités essentielles »

³⁹⁷ BRENNER Susan, *Law in an Era of « Smart » Technology*, Oxford University Press, 2007.

³⁹⁸ *Ibid.*

³⁹⁹ BENNETT MOSES Lyria, « How to Think About Law, Regulation and Technology », *Law, Innovation and Technology*, 5, 2013. P. 15. L'exemple utilisé par l'autrice est la réglementation sur les nanotechnologies.

⁴⁰⁰ À titre d'exemple, ce concept de « disruption juridique », même s'il n'était pas nommé, était un argument au cours du débat sur la pertinence d'un cyberdroit entre d'un côté Franck Easterbrook et, de l'autre côté, Lawrence Lessig. Voir respectivement : EASTERBROOK Frank, « Cyberspace and the Law of the Horse », *University of Chicago Legal Forum*, 1996, 2015. LESSIG Lawrence, « The Law of the Horse », *Harvard Law Review*, 113, The Harvard Law Review Association, 1999.

⁴⁰¹ Ces approches sont néanmoins limitées par leur absence totale de méthodologie d'identification des éléments disruptifs. Cette limite n'est pas exclusive à ces deux théories. Par la suite, d'autres modèles ont été proposés sans arriver à offrir une méthodologie précise et parfaitement reproductible. Pour un modèle s'approchant d'une méthodologie complète et aboutie, voir : LIU Hin-Yan, MAAS Matthijs M., DANAHER John *et al.*, « Artificial Intelligence and Legal Disruption », *Law, Innovation and Technology*, 12, 2020.

⁴⁰² « I do not distinguish sharply between robots and artificial intelligence (AI) agents. As innovation proceeds, the distinction between these two kinds of technologies may be far less important to the law than it seems at present [...] » dans BALKIN Jack M., « The Path of Robotics Law », *California Law Review*, 6, 2015. P. 45. « Part IIA defines robotics and makes the case for what distinguishes robots and artificial intelligence from previous and constituent technologies. » dans CALO Ryan, « Robotics and the Lessons of Cyberlaw », *California Law Review*, 103, 2015. P. 528.

proposée par le professeur Ryan Calo et celle des « saillances » élaborée par le professeur Jack M Balkin.

La théorie des « qualités essentielles ». Lors de la multiplication des robots en-dehors des usines, la doctrine a tenté d'identifier si la robotique présentait des caractéristiques suffisamment exceptionnelles pour pouvoir être considérée comme disruptive du point de vue du droit. En 2014, Ryan Calo nomme ces éléments exceptionnels recherchés les « qualités essentielles » de la technologie qu'il définit comme l'ensemble des caractéristiques spécifiques qui distinguent une technologie de toutes les autres⁴⁰³. Par cette définition, Calo associe deux attributs aux éléments disruptifs juridiquement d'une technologie. Premièrement, ces éléments sont des caractéristiques intrinsèques à ladite technologie. Deuxièmement, il n'est pas nécessaire que chaque caractéristique identifiée soit nouvelle et/ou disruptive juridiquement, c'est leur présence combinée au sein d'une même technologie qui la rend exceptionnelle. Il propose une définition claire de la robotique, à défaut d'être consensuelle⁴⁰⁴ : « *robots are mechanical objects that take the world in, process what they sense, and in turn act upon the world* »⁴⁰⁵. Il identifie alors trois qualités essentielles de la robotique : l'incarnation, l'émergence et la valence sociale. Tout d'abord, l'incarnation représente la combinaison entre la capacité des robots à causer un dommage physique, envers des biens et/ou des personnes, et leur fonctionnement, déterminé de façon complexe par la programmation informatique. Cette première « qualité essentielle » est responsable de plusieurs problématiques juridiques en responsabilité du fait des produits défectueux⁴⁰⁶. Deuxièmement, l'émergence se définit comme la capacité des robots à accomplir des tâches de façon parfois imprévisible. L'existence d'un comportement non anticipable présente à son tour des implications très importantes envers le cadre juridique de la responsabilité civile, pénale et même en droit international⁴⁰⁷. Enfin, la valence⁴⁰⁸ sociale renvoie au constat selon lequel les humains ne perçoivent les robots ni comme

⁴⁰³ CALO Ryan, *op. cit.* P. 517.

⁴⁰⁴ Critique admise par l'auteur lui-même qui considère qu'il y a très peu de technologies complexes qui peuvent posséder une définition totalement consensuelle.

⁴⁰⁵ « *Les robots sont des objets mécaniques qui captent le monde, traitent ce qu'ils perçoivent et agissent à leur tour sur le monde.* » - CALO Ryan, « Robotics and the Lessons of Cyberlaw », , 28 février 2014.

⁴⁰⁶ Sur les conséquences de l'incarnation sur la responsabilité du fait des produits défectueux, cette fois-ci dans le contexte des imprimantes 3D, voir : ENGSTROM Nora Freeman, « 3-D Printing and Product Liability », *University of Pennsylvania Law Review*, 162, 2013. Bien que les imprimantes ne présentent pas une complexité de programmation aussi importante que les robots au sens de Calo, elles sont un autre exemple intéressant concernant l'incarnation.

⁴⁰⁷ ALLEN Craig H., « The Seabots are Coming Here », *The Journal of Navigation*, 65, Cambridge University Press, 2012.

⁴⁰⁸ En psychologie, le terme « valence » est défini comme la puissance d'attraction (valence positive) ou de répulsion (valence négative) qu'un individu éprouve à l'égard d'un objet ou d'une situation : <https://www.cnrtl.fr/definition/valence>

des objets traditionnels ni comme des humains mais comme une entité intermédiaire encore incertaine⁴⁰⁹. Ce rapport nouveau des individus envers des robots peut causer à nouveau un certain nombre de situations pour lesquelles le droit n'est pas encore préparé, comme le cas des robots ayant une finalité sociale. En d'autres termes, le lien particulier qui existe entre un robot et un être humain est une source de préjudice nouvelle que le droit doit prendre en compte⁴¹⁰. Selon cet auteur, ces trois « qualités essentielles » justifient, par conséquent, le besoin d'une réponse juridique au phénomène que sont les robots.

La théorie des « saillances ». Une approche alternative de l'identification du caractère juridiquement disruptif d'une technologie est également proposée par le professeur Jack M Balkin. En réponse à la théorie de Calo, Balkin propose sa théorie des « saillances »⁴¹¹. En effet, ce-dernier considère que la théorie des « qualités essentielles » se concentre aux mauvais endroits pour identifier les éléments disruptifs. Selon lui, ce ne sont pas les caractéristiques intrinsèques de la technologie qui définissent son exceptionnalisme juridique mais l'impact de cette technologie sur la vie sociale car une technologie est avant tout un outil comme un autre de gestion sociale⁴¹². Ainsi, les « saillances de la technologie » sont les conséquences de la technologie sur deux types d'interactions : celles entre les individus et ladite technologie et les interactions entre les individus au travers de l'utilisation de ladite technologie. À son tour, il identifie deux saillances propres à la robotique et à l'IA qu'il appelle à réfléchir lors de leur encadrement : la distribution des droits et des responsabilités dans le contexte d'actions réalisées par des entités non-humaines et l'effet de substitution⁴¹³. La première saillance renvoie à l'idée que les robots et l'IA sont et seront la source de nouveautés qui peuvent être bénéfiques ou préjudiciables⁴¹⁴. L'apparition de ces nouveaux éléments est une saillance de ces technologies et il importe par conséquence de s'interroger sur la façon de répartir les droits et les devoirs dans ce contexte de création par ces technologies⁴¹⁵. L'effet de substitution, quant à lui, peut prendre deux formes. Dans un premier temps, il peut être inconscient et cela renvoie aux effets psychologiques sur l'humain lors d'une interaction avec un robot. Dans un second

⁴⁰⁹ RUCKERT Jolina, « The new ontological category hypothesis in human-robot interaction », *Proceedings of the 6th international conference on Human-robot interaction - HRI* '11, 2011.

⁴¹⁰ DARLING Kate, « Extending Legal Protection to Social Robots », in *Robot Law*, 2012.

⁴¹¹ BALKIN Jack M., « The Path of Robotics Law », *California Law Review*, 6, 2015.

⁴¹² *Ibid.*

⁴¹³ *Ibid.*

⁴¹⁴ Voir introduction.

⁴¹⁵ BALKIN Jack M., *op. cit.*

temps, la substitution peut être consciente. Le robot est alors utilisé comme un remplacement à l'individu dans la réalisation de tâches spécifiques.

B- La justification de l'encadrement spécifique de l'IA

Les spécificités disruptives de l'IA identifiées par la doctrine. Depuis plusieurs années, les problématiques que pose l'IA au droit positif sont étudiées et leurs causes sont recherchées⁴¹⁶. Dans un premier temps, l'autonomie de fonctionnement des logiciels d'IA pose toujours des difficultés juridiques. Tout d'abord, elle interroge sur les tâches qui peuvent être confiées à l'IA sans intervention humaine et les conséquences liées. Cette autonomie différencie l'IA de l'informatique classique car il ne s'agit plus d'automatisation ou d'amélioration de la phase d'exécution de la décision mais d'une véritable délégation du raisonnement décisionnel en ce qui concerne les technologies d'apprentissage⁴¹⁷. Ce passage de l'automatisation à la délégation permet aux logiciels d'IA de dépasser la résolution de problèmes prédéfinis pour s'attaquer à des problèmes non structurés, même s'ils sont encore déterminés par un individu⁴¹⁸. Cette spécificité permet ainsi une utilisation de l'IA dans des situations toujours plus nombreuses et variées. Le fonctionnement autonome pose d'autant plus problème qu'il est imprévisible, autant par les concepteurs du système que par ses utilisateurs. Certains auteurs envisagent déjà cette particularité comme une cause exonératoire de responsabilité⁴¹⁹. D'autres considèrent que cette autonomie et imprévisibilité de fonctionnement caractérisent une agentivité des systèmes d'IA, les posant alors comme des éléments hybrides à mi-chemin entre la personne et l'objet, mettant ainsi à mal la dichotomie fondamentale en droit⁴²⁰. De façon intéressante, cette agentivité spécifique aux systèmes d'IA se rapproche du concept de « valence sociale » exprimé par Calo ainsi qu'à « l'effet de substitution » énoncé par Balkin⁴²¹. La combinaison de l'autonomie et de l'agentivité des technologies semble donc être une caractéristique disruptive juridiquement de

⁴¹⁶ La plupart des auteurs qui ont tenté d'identifier les spécificités des technologies d'IA ne justifient pas d'une préférence envers l'approche des « qualités essentielles » ou des « saillances sociales ».

⁴¹⁷ SARANGI Saswat et SHARMA Pankaj, *Artificial Intelligence: Evolution, Ethics and Public Policy*, London, Routledge India, 2018.

⁴¹⁸ *Ibid.*

⁴¹⁹ SCHERER Matthew U., « Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies », , 30 mai 2015.

⁴²⁰ LIU Hin-Yan, MAAS Matthijs M., DANAHER John, SCARCELLA Luisa, LEXER Michaela et VAN ROMPAEY Léonard, « Artificial Intelligence and Legal Disruption », *op. cit.*

⁴²¹ BALKIN Jack M., « The Path of Robotics Law », *California Law Review*, 6, 2015.

l'IA. Dans un second temps, l'opacité particulièrement importante des algorithmes d'IA⁴²² et de ses résultats⁴²³, semble être une autre des particularités des technologies d'IA. Enfin, la manière de développer des outils d'IA est encore une autre spécificité pour Scherer qui complexifie l'application des dispositions juridiques existantes⁴²⁴. Selon lui, le fait que l'IA puisse être conçue sans recourir à des infrastructures visibles, par différents acteurs, situés à différents endroits, impliqués dans différents éléments du logiciel final, rend très compliquée l'application de la loi dans l'espace et facilite par la même occasion le *forum shopping* pour les fournisseurs⁴²⁵.

Les spécificités disruptives de l'IA sélectionnées par la Commission européenne. La Commission européenne, quant à elle, présente aussi ses arguments pour justifier son initiative de législation propre à l'IA. Elle liste ainsi des « facteurs de problèmes » présents dans l'IA, sélectionnés au regard de leur capacité à générer des problèmes juridiques. Cette liste de particularités des technologies d'IA proposée par la Commission européenne se fonde sur de très nombreux documents de travail sur les technologies d'IA, aussi bien au niveau européen qu'au niveau international⁴²⁶. Le groupe d'experts indépendants de haut niveau sur l'IA a notamment produit quatre de ces documents préparatoires, respectivement des lignes directrices relative à l'éthique de l'IA, des recommandations sur les politiques et les investissements pour une IA digne de confiance, des réflexions sur la mise en place de ces recommandations dans certains secteurs d'activité et une liste d'évaluation pour une IA digne de confiance (ci-après ALTAI). Le choix de la Commission européenne d'encadrer transversalement les technologies d'IA dans l'UE ne se fonde pas uniquement sur une volonté politique mais prend également appui sur des experts indépendants issus pour la plupart de l'entreprise ou de l'université⁴²⁷.

La liste des spécificités de l'IA justifiant la proposition de réglementation sur l'IA s'élève ainsi au nombre de cinq⁴²⁸. Premièrement et sans surprise on retrouve l'opacité du fonctionnement

⁴²² BURRELL Jenna, « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, vol. 3, n° 1, 1 juin 2016, p. 2053951715622512.

⁴²³ BUITEN Miriam C., « Towards Intelligent Regulation of Artificial Intelligence », *European Journal of Risk Regulation*, vol. 10, n° 1, mars 2019, p. 41-59.

⁴²⁴ SCHERER Matthew U., « Regulating Artificial Intelligence Systems », *op. cit.*

⁴²⁵ *Ibid.*

⁴²⁶ COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021. P. 9 à 12.

⁴²⁷ Parmi les 52 experts composant le groupe, 24 d'entre eux viennent d'entreprises, 17 sont des universitaires, 5 appartiennent à la société civile et les autres proviennent de diverses organisations telles que l'agence des droits fondamentaux de l'Union européenne. « 'Trustworthy AI' is not an appropriate framework », sur *AlgorithmWatch*.

⁴²⁸ COMMISSION EUROPÉENNE, *Impact assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, Brussels, 2021.

du système qui complexifie la détection, le contrôle et la preuve de violations de droits⁴²⁹. La deuxième spécificité retenue est la complexité des technologies d'IA, à savoir la multiplicité des composants qui composent un logiciel d'IA, ce qui accroît encore davantage la difficulté de contrôler effectivement le respect des dispositions juridiques. La troisième spécificité est l'imprévisibilité et l'adaptabilité continue des systèmes. La Commission européenne anticipe les nouveaux risques que pourrait générer cette caractéristique. L'avant dernière spécificité est l'autonomie de fonctionnement, c'est-à-dire la capacité à produire des résultats avec peu ou pas d'intervention humaine. Enfin, la dernière caractéristique est la place centrale des données dans le développement et l'utilisation des systèmes d'IA. En plus de représenter un traitement massif de données, les données utilisées dans la conception ont un impact considérable sur la qualité de fonctionnement du système d'IA, notamment sur les problématiques de biais⁴³⁰.

Ces spécificités de l'IA sont toutes des caractéristiques inhérentes aux technologies d'IA. La Commission européenne semble justifier que la logique même de fonctionnement des technologies représente une potentielle source de problèmes juridiques. Les motivations énoncées se rapprochent donc davantage de la théorie des « qualités essentielles » de Calo que de la théorie des « saillances » de Balkin. Cependant il ne serait pas juste de considérer que ce rapprochement est intentionnel. D'une part, la théorie n'est pas explicitement mentionnée comme source d'inspiration par l'analyse d'impact du règlement européen sur l'IA. D'autre part, la Commission européenne ne se limite pas à énumérer des spécificités intrinsèques de l'IA. En effet, la justification de l'élaboration du texte se fonde également, pour une partie importante, sur les mauvais usages pouvant être faits des technologies d'IA. Par conséquent, la justification du droit de l'IA s'appuie aussi bien sur les spécificités problématiques liées au fonctionnement intrinsèque des technologies d'IA que sur les problématiques relatives à leur utilisation.

§2. La définition du système d'IA comme objet de droit

Annonce de plan. Le choix de réglementer spécifiquement l'IA a été motivé par les particularités, de fonctionnement et d'usage, des technologies qui composent ce terme. Dès lors,

⁴²⁹ *Ibid.*

⁴³⁰ Les données d'entraînement ne sont pas la cause exclusive des biais algorithmiques. Ces derniers peuvent trouver leur source dans les biais préexistants dans la société mais également dans les biais des humains en charge de la conception du système d'IA.

c'est assez logiquement que le législateur européen a décidé de faire du système d'IA la notion centrale de son règlement. À l'image de tous les législateurs ayant opté pour un encadrement transversal de l'IA, la Commission européenne doit alors définir l'objet qu'est le système d'IA. Ce travail définitionnel soulève des enjeux importants car le système d'IA contribuera en grande partie à délimiter le champ d'application du règlement. La définition doit donc être suffisamment claire pour faciliter l'application du texte et assez large pour intégrer les problématiques spécifiques des technologies d'IA. Au départ, l'Union européenne ainsi que d'autres législateurs ont élaboré une définition de l'IA hybride, y intégrant alors des termes scientifiques **(A)**. Cependant, les versions finales de ces initiatives réglementaires ont fini par abandonner l'aspect scientifique pour privilégier une définition certes plus pérenne de l'IA mais aussi plus critiquable au regard des besoins d'une telle réglementation **(B)**.

A- Une définition juridique bipartite des systèmes d'IA.

La définition européenne de système d'IA. Définir juridiquement un objet technologique est toujours un défi. Pour reprendre les termes de la Commission européenne dans son livre blanc sur l'IA de 2020 : « *la définition de l'IA devra être suffisamment souple pour tenir compte des progrès techniques tout en étant suffisamment précise pour garantir la sécurité juridique nécessaire.* »⁴³¹. Parmi les différentes initiatives réglementaires relatives aux technologies d'IA, certains législateurs ont adopté une définition du système d'IA en deux parties. En droit européen, la proposition de règlement européen énonce au départ à son article 3 la définition suivante du système d'IA :

*« Un logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit ; »*⁴³²

On distingue bien ici le premier élément décrivant le système d'IA au travers de ses fonctions : le système génère des résultats qui, en fonction des objectifs donnés par un humain, influencent

⁴³¹ Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance, Bruxelles, Commission européenne, 2020.

⁴³² Proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (Législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union, 2021. Article 3 (1).

son ou ses environnements. Le second élément est relégué aux annexes de la proposition et entreprend de définir les technologies entrant dans l'expression « intelligence artificielle »⁴³³. La place et le contenu de ce second élément définitionnel a évolué au fil des textes de compromis et des amendements proposés mais la dichotomie semble constante. Au regard des arguments de justification du règlement européen, cette dualité apparaît cohérente. En effet, l'aspect fonctionnel de la définition renvoie aux problématiques spécifiques liées à l'usage des technologies d'IA tandis que l'aspect descriptif rappelle les caractéristiques disruptives intrinsèques au fonctionnement des technologies.

La définition canadienne du système d'IA. Cette approche définitionnelle n'est pas propre à l'Union européenne. Cette dichotomie dans la définition des systèmes d'IA se retrouve, en moins marquée, dans la partie 3 de la proposition de loi canadienne C-27 sur l'IA et les données⁴³⁴. Néanmoins, la Chambre des communes canadienne propose à son article 2 une définition en une seule partie, similaire à celle de la Commission européenne :

*« Système technologique qui, de manière autonome ou partiellement autonome, traite des données liées à l'activité humaine par l'utilisation d'algorithmes génétiques, de réseaux neuronaux, d'apprentissage automatique ou d'autres techniques pour générer du contenu, faire des prédictions ou des recommandations ou prendre des décisions. »*⁴³⁵

Le législateur canadien, comme le législateur français, met l'accent sur la capacité des systèmes à agir sur leur environnement. Les deux dispositions qualifient les systèmes d'IA en fonction de leur capacité à générer du contenu, faire des prédictions, des recommandations au travers de technologies plus ou moins identifiées selon les textes.

B- Le changement pour une définition générale encore à préciser

La perte de l'aspect descriptif de la définition des systèmes d'IA. Dans sa version finale, le règlement européen sur l'IA opère un changement important sur la définition d'un système d'IA. Désormais, l'objet juridique au centre du texte est défini comme :

⁴³³ *Ibid.* Annexe I

⁴³⁴ CHAMBRE DES COMMUNES, *Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*, 2022, p. 27.

⁴³⁵ *Loi sur l'intelligence artificielle et les données*, 16 juin 2022.

« un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels; »⁴³⁶

De façon très intéressante, cette définition juridique du système d'IA se retrouve de façon presque à l'identique dans d'autres textes sur l'IA. Ainsi, l'*Executive order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, publié le 30 octobre 2023 par le gouvernement des États-Unis définit l'IA et les systèmes d'IA comme suivant :

« The term “AI system” means any data system, software, hardware, application, tool, Or, utility that operates in whole Or, in part using AI. »⁴³⁷

« The term “artificial intelligence” Or, “AI” has the meaning set forth in 15 U.S.C. 9401(3): a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, Or, decisions influencing real Or, virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information Or, action. »⁴³⁸

Cette définition est également partagée par l'organisation de coopération et de développement économiques (OCDE) :

« un système d'IA est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels. Les systèmes d'IA peuvent présenter différents niveaux d'autonomie et de capacité d'adaptation après leur déploiement. »⁴³⁹

Malgré les légères variations entre les textes, un consensus définitionnel semble apparaître concernant le système d'IA. Ces récentes définitions ont en commun le fait de ne pas reproduire la structure hybride identifiée dans les premières versions de la proposition de règlement

⁴³⁶ « Règlement sur l'intelligence artificielle », 2024. Article 3.1.

⁴³⁷ « Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence », 2023. Section 3(e).

⁴³⁸ *Ibid.* Section 3(b).

⁴³⁹ « Updates to the OECD's definition of an AI system explained ». Traduction réalisée par l'auteur.

européen sur l'IA⁴⁴⁰. La définition retenue par la version finale du texte européen a exclu la composante descriptive scientifique, à savoir la référence à l'annexe I qui listait les techniques et approches composant l'IA. Pour cause, l'annexe I du règlement européen a rapidement été supprimée lors des discussions⁴⁴¹. Dès lors, le système d'IA n'est défini que par son aspect fonctionnel, c'est-à-dire comment il fonctionne et ce qu'il peut accomplir. Ce changement de la part de l'Union européenne n'est pas sans conséquences.

Une définition plus cohérente fondée sur les spécificités de l'IA. Le système d'IA étant l'objet juridique central de la réglementation européenne sur l'IA, sa définition a des conséquences importantes sur le champ d'application du texte. L'annexe I renvoyait à plusieurs familles d'approches et de techniques informatiques composant l'IA. Cette liste élargissait grandement, voire de façon disproportionnée, la notion de système d'IA⁴⁴², et donc le champ d'application du règlement. Le législateur européen souhaite recentrer la définition autour des caractéristiques essentielles des systèmes d'IA⁴⁴³. Les mentions « *d'une capacité d'adaptation parès sont déploiement* », de « *différents niveaux d'autonomie* » ainsi que la possibilité de fonctionnement avec des « *objectifs explicites ou implicites* » sont le résultat de cette volonté. La Commission européenne insistait sur le fait que les causes principales des problèmes de l'IA sont « *les caractéristiques spécifiques des systèmes d'IA qui les rendent qualitativement différents des avancées technologiques précédentes* »⁴⁴⁴. Avec cette nouvelle définition, la notion de système d'IA se révèle bien plus cohérente avec les motivations derrière l'élaboration du règlement⁴⁴⁵. Enfin, l'Union européenne rend la notion juridique de système d'IA neutre technologiquement. Dès lors, la définition est bien moins exposée au risque d'obsolescence technologique. Quelles que soient les nouvelles technologies informatiques qui apparaîtront, les logiciels qui y auront recours pourront être concernés par le texte européen.

La délimitation encore imprécise de la notion. Si la nouvelle définition se révèle plus adéquate sur plusieurs aspects, il demeure toutefois un problème de précision de la définition. En d'autres termes, la frontière entre les systèmes d'IA et les autres systèmes informatiques

⁴⁴⁰ Voir précédemment.

⁴⁴¹ Les deux premières familles d'approches et de techniques listées à l'annexe I n'ont toutefois pas complètement disparues. Elles sont présentes au considérant 12 du texte européen. Seul le troisième point de la liste, le plus critiqué, n'a pas été conservé. « Règlement sur l'intelligence artificielle », 2024. SÉNÉCHAL Juliette, « Vote des parlementaires européens sur l'AI Act : vers une réglementation accrue des IA, des modèles de fondation et des IA génératives, s'inspirant du DSA, du Data Act et du RGPD ? », *Dalloz actualités*, 2023.

⁴⁴² Voir notamment : VEALE Michael et ZUIDERVEEN BORGESIU Frederik, « Demystifying the Draft EU Artificial Intelligence Act », *Computer Law Review International*, 22, 2021. P. 109.

⁴⁴³ « Règlement sur l'intelligence artificielle », 2024. Considérant 12.

⁴⁴⁴ COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021. P. 28.

⁴⁴⁵ *Ibid.*

n'est pas complètement claire. Or, le manque de précision est l'un des deux principaux problèmes de forme révélés par notre analyse des textes éthiques sur l'IA et des dispositions de droit positif relatifs aux décisions impliquant un système algorithmique⁴⁴⁶. Par exemple, l'indication selon laquelle les objectifs d'un système d'IA peuvent être « *explicites ou implicites* » permet de combler une lacune par rapport à la version proposée par la Commission européenne qui ne mentionnait que les « *objectifs définis par l'homme* »⁴⁴⁷. Cependant, il n'existe pas d'objectifs qui ne soient ni explicites, ni implicites. Par conséquent, ce critère de qualification, s'il demeure plus adéquat par rapport à la réalité d'un système d'IA, ne semble pas exclure de situations possibles. De la même manière, la possibilité pour le système d'IA d'influencer « *des environnements physiques ou virtuels* » n'exclut aucune situation réelle du champ de la définition. Si la nouvelle définition se révèle plus pertinente, un approfondissement des termes utilisés pour définir un système d'IA sera néanmoins nécessaire. Sans ce travail, les fournisseurs de systèmes d'IA, les déploieurs, les autorités de contrôle et autres acteurs concernés risquent d'avoir des difficultés lors de l'identification des logiciels concernés par le règlement européen.

Sous-section 2 : Le niveau de risque comme critère majeur d'application

Annnonce de plan. Le législateur européen a refusé un encadrement uniforme des systèmes d'IA afin de ne pas porter un trop grand préjudice au marché européen⁴⁴⁸. L'application des exigences relatives aux systèmes d'IA est donc proportionnée selon le niveau de risques auquel les individus sont exposés. Cette approche par les risques se divise en quatre catégories : les risques inacceptables, les risques élevés, les risques spécifiques, et les risques moyens et faibles. Dans la version originelle du règlement européen publiée par la Commission européenne, la quasi-totalité des obligations relatives aux systèmes d'IA contenues dans le texte était imposée aux systèmes d'IA à risque élevé. Désormais, l'Union européenne a aussi ajouté plusieurs autres obligations, notamment à destination d'une nouvelle catégorie de logiciels : les systèmes d'IA à usage général. Pour autant, malgré l'augmentation des catégories juridiques de systèmes d'IA contenues dans le texte européen, les systèmes d'IA à risque élevé conservent une place centrale

⁴⁴⁶ Voir Sous-section 2 : Un besoin de dépasser les limites du droit positif.

⁴⁴⁷ « Règlement sur l'intelligence artificielle », 2024. Version initiale de la Commission européenne.

⁴⁴⁸ COMMISSION EUROPÉENNE, *op. cit.*

dans la réglementation (§2). Dès lors, il est nécessaire de s'attarder sur les modalités de qualification d'un système d'IA à risque élevé (§1).

§1. La qualification du système d'IA à risque élevé

Annonce de plan. Au regard des conséquences importantes pour les individus, le choix a été fait de concentrer notre analyse sur les systèmes d'IA à haut risque. Dès lors, il convient de s'intéresser aux conditions de qualification de cette catégorie juridique. À la lecture du règlement européen sur l'IA, il apparaît que le législateur prend en main, dans un premier temps, la désignation des critères caractérisant le risque élevé (A). Toutefois, le fournisseur du système d'IA n'est pas exclu du travail d'évaluation du niveau de risque et se voit octroyer la possibilité de déroger, dans un second temps, à la qualification de système d'IA à risque élevé (B).

A- Les critères de qualification du système d'IA à risque élevé

Le choix européen d'une description prévisible du contenu des niveaux de risques. Avec sa proposition de règlement sur l'IA, la Commission européenne a décidé de ne pas recourir à la traditionnelle délégation de l'évaluation des risques. En effet, le niveau de risque élevé est énoncé clairement dans le règlement et leur contenu est lui aussi exhaustivement indiqué. Le travail de réflexion sur l'objectif d'évaluation et la méthodologie pour y parvenir ne sont alors plus à la charge des fournisseurs de systèmes d'IA. La Commission européenne a réalisé elle-même cette charge de travail et a intégré ses résultats dans le texte.

Les pratiques impliquant des systèmes d'IA qui exposent à des risques élevés sont ainsi listés de façon exhaustive dans le texte. Les systèmes d'IA à haut risque sont alors divisés en deux catégories : les composants de sécurité de certains produits déjà réglementé par certains actes d'harmonisation européen d'un côté (annexe I), et les systèmes d'IA inscrits dans l'annexe III du règlement sur l'IA de l'autre⁴⁴⁹. La première catégorie lie la notion de systèmes d'IA à haut risque avec certains produits déjà soumis à une harmonisation européenne faisant l'objet d'un examen par un tiers. Les textes visés par le règlement sur l'IA sont listés exhaustivement dans

⁴⁴⁹ « Règlement sur l'intelligence artificielle », 2024. Article 6.1 et 6.2.

son annexe I⁴⁵⁰. Ce choix du législateur européen est évident. Le règlement sur l'IA étant lui-même un texte d'harmonisation, il serait paradoxal qu'il crée des conflits et des interrogations avec d'autres textes d'harmonisation existant. La seconde catégorie de systèmes d'IA à risque élevé représente les nouveaux systèmes d'IA à risque élevé sous la forme de pratique énumérées au sein de huit secteurs d'activité identifiés et listés au sein de l'annexe III du règlement sur l'IA. Chaque secteur contient une liste de pratiques qui seront considérées comme des pratiques à haut risque dès lors qu'elles sont réalisées par le biais de systèmes d'IA. Cette prise en charge par la Commission européenne de la responsabilité d'évaluer la criticité des systèmes d'IA est donc en décalage avec l'approche traditionnelle de mise en conformité présentée précédemment. Toutefois, les fournisseurs de systèmes d'IA ne sont pas exemptés totalement d'analyse de risques spécifiques à l'IA. En effet, les fournisseurs de systèmes d'IA à haut risque doivent malgré tout mettre en œuvre un système de gestion des risques au travers duquel ils doivent identifier et évaluer les risques divers associés à leurs systèmes d'IA :

« a) l'identification et l'analyse des risques connus et prévisibles associés à chaque système d'IA à haut risque ;

b) l'estimation et l'évaluation des risques susceptibles d'apparaître lorsque le système d'IA à haut risque est utilisé conformément à sa destination et dans des conditions de mauvaise utilisation raisonnablement prévisible ;

c) l'évaluation d'autres risques susceptibles d'apparaître, sur la base de l'analyse des données recueillies au moyen du système de surveillance après commercialisation visé à l'article 61 ;

*d) l'adoption de mesures appropriées de gestion des risques conformément aux dispositions des paragraphes suivants. »*⁴⁵¹

L'allègement des incertitudes liées à la réalisation d'une analyse d'impact est donc limité à la situation juridique du fournisseur. Les fournisseurs s'épargnent le travail de pré-évaluation des risques afin d'identifier la criticité de leurs pratiques. Ce choix de la part de la Commission européenne est important car il modifie les acteurs en charge de la détermination du seuil de criticité en droit des systèmes d'IA. Toutefois, la proposition de règlement est un texte encore en discussion et cette absence d'obligation d'évaluation de la criticité envers les fournisseurs de systèmes d'IA peut évoluer. Ainsi, parmi les amendements proposés, certains d'entre eux

⁴⁵⁰ Sont inscrits à titre d'exemple, le règlement 2017/745 sur les dispositifs médicaux ou encore le règlement 300/2008 sur la sûreté de l'aviation civile.

⁴⁵¹ « Règlement sur l'intelligence artificielle », 2024. Article 9.2.

proposent le retour à la méthode traditionnelle de délégation de l'analyse de la criticité aux fournisseurs de systèmes d'IA⁴⁵².

Les limites de la désignation législative des pratiques à haut risques. Le choix de fournir une liste des utilisations critiques des systèmes d'IA est donc une stratégie avantageuse pour les fournisseurs de systèmes d'IA. Il permet de définir des règles qui permettent, d'une part, aux acteurs d'anticiper les exigences et, d'autre part, de s'assurer que leur produit sera commercialisable et ne sera pas subitement déclaré illégal. En d'autres termes, cette décision vise à offrir une sécurité juridique aux investissements dans le domaine des technologies d'IA. Mais si la liste des pratiques critiques permet effectivement de fixer les règles du jeu du marché de l'IA, elle peine à les délimiter clairement, réduisant alors significativement leur intérêt pour les fournisseurs mais également pour la protection des personnes affectées.

Un manque de clarté similaire se retrouve dans la liste des systèmes d'IA à haut risque, en particulier les pratiques des nouveaux secteurs à haut risque proposés par le règlement⁴⁵³. Le choix des secteurs critiques et des pratiques considérées comme à haut risque qu'ils contiennent n'est pas justifié par la Commission européenne. De ce fait, il est très difficile de déduire les critères qui ont permis de qualifier ces secteurs et ces pratiques comme présentant des risques élevés⁴⁵⁴. Or, la Commission européenne présente une méthodologie qu'elle utilisera pour mettre à jour la liste des pratiques de systèmes d'IA à haut risque. Cette méthode prend notamment en compte :

« (a) la destination du système d'IA ;

(b) la mesure dans laquelle un système d'IA a été utilisé ou est susceptible de l'être ;

c) la nature et la quantité des données traitées et utilisées par le système d'IA, en particulier si des catégories particulières de données à caractère personnel sont traitées ;

(d) la mesure dans laquelle le système d'IA agit de manière autonome et la mesure dans laquelle l'homme peut intervenir pour annuler une décision ou des recommandations susceptibles de causer un préjudice potentiel ;

⁴⁵² « Document - Les corapporteurs revoient la classification des systèmes à haut risque - Contexte ».

⁴⁵³ Il s'agit donc de l'annexe III du règlement. L'annexe II renvoie à d'autres textes européens. De ce fait, les définitions et les systèmes d'IA concernés sont plus faciles à qualifier car ce sont ceux qui s'intègrent à des produits clairement identifiés dans des textes dédiés. Ce n'est pas le cas de l'annexe III.

⁴⁵⁴ HUTTNER Liane, *La décision de l'algorithme*, Dalloz, 2024.

[...] »⁴⁵⁵

Il existe donc des conditions pour qualifier une pratique d'IA à haut risque, bien qu'elles soient destinées à être utilisées par la Commission seulement. Cette méthodologie semble révéler les critères sous-jacents derrière la notion de risque élevé, bien plus divers que le seul secteur d'utilisation du système d'IA. Au total, ce sont treize éléments qui sont listés dans le règlement européen. On observe ainsi un changement d'angle par rapport à la liste de l'annexe III qui se concentrait uniquement sur l'usage⁴⁵⁶. L'existence d'un double prisme d'évaluation des risques élevés est difficile à appréhender pour les fournisseurs. Dans le cas où ils s'appêtent à développer un système d'IA dans un secteur critique, qui semble présenter des risques élevés mais est étranger aux pratiques listées exhaustivement dans l'annexe III du règlement, comment doivent-ils qualifier leurs futurs systèmes d'IA ? La logique de liste offre, d'un côté, une sécurité juridique dans les éléments qu'elle liste mais crée, d'un autre côté, des incertitudes sur les systèmes d'IA absents de la liste mais dont l'usage est potentiellement à haut risque.

B- L'exception de l'absence de risques significatifs

La création d'une exception relative aux systèmes d'IA utilisés dans des secteurs d'activités à haut risque. Au sein du règlement européen sur l'IA, la qualification d'un système d'IA comme étant à haut risque présente une exception :

« Par dérogation au paragraphe 2, un système d'IA visé à l'annexe III n'est pas considéré comme étant à haut risque lorsqu'il ne présente pas de risque important de préjudice pour la santé, la sécurité ou les droits fondamentaux des personnes physiques, y compris en n'ayant pas d'incidence significative sur le résultat de la prise de décision. »⁴⁵⁷

Cette dérogation à la qualification de système d'IA à haut risque concerne seulement les systèmes d'IA qui sont utilisés dans des secteurs et pour des finalités considérées à haut risque par l'annexe III du texte européen. Les systèmes d'IA concernés par l'annexe II du texte relative aux produits soumis à la législation d'harmonisation de l'Union européenne ne peuvent bénéficier de cette exception.

⁴⁵⁵ « Règlement sur l'intelligence artificielle », 2024. Article 7.1.b)

⁴⁵⁶ HUTTNER Liane, *op. cit.*

⁴⁵⁷ « Règlement sur l'intelligence artificielle », 2024. Article 6.3 alinéa 1.

Afin d'identifier la présence de risques significatifs, ou leur absence le cas échéant, le législateur européen fournit une liste exhaustive, mais non définitive⁴⁵⁸, de critères :

- « a) le système d'IA est destiné à exécuter une tâche procédurale étroite ;*
- b) le système d'IA est destiné à améliorer le résultat d'une activité humaine déjà réalisée ;*
- c) le système d'IA est destiné à détecter des schémas de prise de décision ou des écarts par rapport à des schémas décisionnels antérieurs et n'est pas destiné à remplacer ou influencer l'évaluation humaine l'évaluation humaine effectuée précédemment, sans examen humain approprié ; ou*
- d) le système d'IA est destiné à effectuer une tâche préparatoire à une évaluation pertinente aux fins des cas d'utilisation énumérés à l'annexe III ; »*⁴⁵⁹

La satisfaction d'un seul de ces critères permettra ainsi au fournisseur du système d'IA de contester la qualification « à risque élevé » de son logiciel. Au premier abord, l'apparition de la notion de « risque significatif » peut paraître déroutant, notamment du fait de sa proximité sémantique avec celui de « risque élevé ». D'un côté, le terme de « risque élevé » désigne une catégorie juridique de systèmes d'IA au sein du règlement européen. De l'autre, le terme de « risque significatif » renvoie à un des multiples critères qu'un système d'IA peut remplir pour être qualifié de système d'IA à risque élevé.

Dès la version du texte proposé par le parlement européen, la conception d'un système d'IA dans le but d'être utilisé dans un secteur dit « à haut risque » n'était pas une condition suffisante pour que le logiciel soit qualifié de système d'IA à risque élevé. Un second critère avait été ajouté : la présence d'un risque significatif de dommage. La notion est donc déjà présente depuis plusieurs versions du règlement mais, là où elle était un critère de qualification, elle devient un motif de dérogation. En d'autres termes, la notion de « risque significatif » a subi un changement de fonction dans le règlement. Par conséquent, les fournisseurs de systèmes d'IA sont exemptés de la charge, mais aussi du pouvoir, de qualifier leurs logiciels comme des

⁴⁵⁸ « La Commission peut adopter des actes délégués ajoutant de nouveaux critères à ceux énoncés aux points a) à d) du premier alinéa du paragraphe 2 bis, premier alinéa, ou les modifier, uniquement lorsqu'il existe des preuves concrètes et fiables de l'existence de systèmes d'IA relevant du champ d'application de l'annexe III, mais qui ne présentent pas de un risque significatif d'atteinte à la santé, à la sécurité et aux droits fondamentaux. » - *Ibid.* Article 6.2.d) alinéa 2 du compromis final du texte.

⁴⁵⁹ *Ibid.* Version finale du compromis.

systèmes d'IA à haut risque⁴⁶⁰. Le caractère limitatif de la liste de situations dérogatoires fournie dans le règlement européen renforce cette conclusion.

Le critère de qualification sous-jacent du risque élevé. Il est possible de résumer la qualification de système d'IA à haut risque dans le cadre de l'annexe III de la façon suivante autour de la notion de risque significatif. Les secteurs d'activité listés à l'annexe III du règlement européen représentent une présomption de la présence de risques significatifs. Dès lors qu'un système d'IA est destiné à être utilisé pour l'un des usages énoncés, la charge de la preuve de l'absence de risques significatifs est transmise au fournisseur du système d'IA. Si celui-ci parvient à prouver cette absence de risques significatifs à l'aide des critères fournis par la Commission européenne, le système d'IA n'est plus considéré comme à haut risque. Par conséquent, dans le contexte de l'annexe III, il semble que le critère concret de qualification du risque élevé soit le risque significatif. Les secteurs listés à l'annexe III apparaissent alors comme une manière approximative de comprendre tous les usages possibles présentant un risque significatif pour les individus sans pour autant confier aux fournisseurs de systèmes d'IA la tâche d'évaluer le niveau de risque.

§2. La place du risque élevé en droit des décisions impliquant un algorithme

Le concept de seuil critique de risques. Le concept de « criticité » exprime ici le seuil de gravité d'une pratique à partir duquel la réglementation impose des contraintes supplémentaires à l'action dite « critique ». Au sein du règlement européen sur l'IA, le seuil de criticité est celui à partir duquel un système d'IA est considéré à haut risque. Par conséquent un système d'IA critique désignera à la fois les systèmes d'IA à haut risque et ceux présentant un risque inacceptable. Ce concept de « criticité » ne renvoie pas à une équivalence entre différents niveaux de risques présents dans différentes dispositions. Il exprime l'usage d'un seuil de gravité des risques comme condition d'application de mesures juridiques particulières. La criticité est employée pour désigner le mécanisme juridique conditionnant des obligations et non pour évoquer le contenu de ladite condition. Une pratique peut donc être critique pour un texte et non critique pour un autre. Cette variation peut également se constater entre des

⁴⁶⁰ Cela aurait été le cas si la double condition proposée par le parlement européen avait été conservée.

dispositions d'un même texte. Toutefois, la présence de ce mécanisme dans de nombreuses réglementations, européennes et internationales, applicables à l'IA en fait un outil d'articulation particulièrement important.

L'identification du seuil critique de risque laissée aux débiteurs des exigences.

Traditionnellement, l'évaluation du niveau de risques qu'une pratique fait courir aux individus est à la charge des débiteurs de l'obligation. La réglementation européenne sur les données à caractère personnel est une parfaite illustration de ce phénomène de délégation de l'analyse de la criticité aux auteurs desdits risques. Malgré les critères de criticité dévoilés par les autorités de contrôle des données personnelles⁴⁶¹, le responsable de traitement ou le sous-traitant demeure en charge de qualifier l'opération de traitement comme présentant un risque élevé et donc d'identifier la nécessité d'effectuer une analyse d'impact. De la même manière, le législateur européen donne les grandes lignes du contenu de cette analyse⁴⁶² mais sa réalisation incombe à nouveau aux débiteurs de l'obligation de réaliser l'étape d'évaluation de la criticité des impacts⁴⁶³. L'approche par les risques suit la même logique en droit canadien des décisions automatisées. La directive canadienne sur la prise de décision automatisée impose aux administrations concernées d'effectuer une évaluation d'incidence⁴⁶⁴, c'est-à-dire de classer le système de prise de décisions automatisées parmi les quatre catégories énoncées par la loi⁴⁶⁵. En outre, le législateur canadien conserve cette obligation d'analyse de risques dans son projet de loi sur l'IA et les données⁴⁶⁶.

La notion d'« incidence élevée » représente le seuil de criticité de la future réglementation canadienne⁴⁶⁷. Or, cette expression n'est pas définie par le texte et l'identification de ce niveau d'incidence est à nouveau à la charge du responsable du système d'IA : « *Le responsable d'un*

⁴⁶¹ Cf lignes directrices G29 paragraphe précédent

⁴⁶² RGPD, article 35.7 : « *L'analyse contient au moins: a) une description systématique des opérations de traitement envisagées et des finalités du traitement, y compris, le cas échéant, l'intérêt légitime poursuivi par le responsable du traitement; b) une évaluation de la nécessité et de la proportionnalité des opérations de traitement au regard des finalités; c) une évaluation des risques pour les droits et libertés des personnes concernées conformément au paragraphe 1; et d) les mesures envisagées pour faire face aux risques, y compris les garanties, mesures et mécanismes de sécurité visant à assurer la protection des données à caractère personnel et à apporter la preuve du respect du présent règlement, compte tenu des droits et des intérêts légitimes des personnes concernées et des autres personnes affectées.* »

⁴⁶³ Logique similaire pour l'évaluation du risque élevé en cas de violation de données personnelles. Une analyse cruciale qui permet de déterminer l'ampleur des obligations dont le responsable de traitement ou le sous-traitant fait l'objet. Articles 33 et 34 RGPD.

⁴⁶⁴ « Directive sur la prise de décision automatisée (Canada) », 2019, article 6.1.1.

⁴⁶⁵ À nouveau, ce classement détermine la quantité et l'ampleur des obligations à réaliser pour l'administration. La décision revêt par conséquent des enjeux financiers, humains et techniques très importants.

⁴⁶⁶ *Loi sur l'intelligence artificielle et les données*, 16 juin 2022.

⁴⁶⁷ SCASSA Teresa, « Canada's Proposed AI & Data Act - Purpose and Application », publié le 8 août 2022.

système d'intelligence artificielle évaluée, conformément aux règlements, si le système est un système à incidence élevée. »⁴⁶⁸. Cette évaluation sera guidée, plus ou moins précisément, par les règlements élaborés par le Cabinet du ministre qui indiqueront les critères de détermination d'une incidence élevée⁴⁶⁹.

Cette stratégie consistant à confier la réalisation de l'évaluation de la criticité d'une pratique (traitement de données, utilisation d'un système d'IA, etc.) à l'auteur de cette activité entraîne deux conséquences. Tout d'abord, et conformément à la stratégie juridique de conformité, le débiteur de l'obligation doit déterminer concrètement à quoi correspond le niveau de criticité énoncé par la réglementation lorsque ce-dernier n'est pas défini. Une fois qu'il a défini tant bien que mal ce seuil, il doit réfléchir à une méthodologie pour identifier ce seuil de criticité ainsi que les différents niveaux de risques qui le constituent. Il doit donc affronter une double incertitude à la fois sur la détermination de l'objectif qu'est le seuil de criticité et sur la méthodologie pour identifier cet objectif. La flexibilité permise par cette délégation d'évaluation des risques suppose donc une charge de travail importante pour les entités débitrices de cette obligation.

L'importance du risque élevé dans le droit des systèmes d'IA. Dans son livre blanc sur l'IA, la Commission européenne établissait le champ d'application de son futur cadre légal en proposant deux catégories de systèmes d'IA : ceux présentant un risque élevé et ceux ne présentant pas un risque élevé⁴⁷⁰. La proposition visait alors à ne poser de contraintes juridiques que sur les systèmes d'IA à haut risque, faisant de la notion de risque élevé une condition d'application stricte du texte européen. Cette logique binaire apparaît réductrice et potentiellement limitante pour lutter contre la diversité des enjeux des technologies d'IA⁴⁷¹. En comparaison, la directive canadienne sur la prise de décision automatisée aborde le sujet de façon plus nuancée en proposant quatre niveaux de risques⁴⁷². Toutefois, la Commission européenne a entendu les critiques émises à l'encontre des textes préparatoires et offre désormais une gradation de quatre niveaux de risques : inacceptables, élevés, spécifiques et autres risques.

⁴⁶⁸ *Loi sur l'intelligence artificielle et les données*, 16 juin 2022., article 7.

⁴⁶⁹ Teresa Scassa blog – « *Canada's Proposed AI & Data Act – Purpose and Application* ».

⁴⁷⁰ COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020.

⁴⁷¹ CASTETS-RENARD Céline, « Le Livre blanc de la Commission européenne sur l'intelligence artificielle : vers la confiance ? », *Recueil Dalloz Sirey*, vol. 15, 23 avril 2020, p. 837-842.

⁴⁷² « Directive sur la prise de décision automatisée (Canada) », 2019.

Les niveaux de risques non élevés. Tout d'abord, lorsque le recours à un système d'IA présente des risques trop élevés, son utilisation devient inacceptable et le système d'IA est donc interdit par le règlement⁴⁷³. Ces utilisations interdites sont listées limitativement par le texte européen. Par exemple, la mise sur le marché de systèmes d'IA qui « *créent ou développent des bases de données de reconnaissance faciale par le moissonnage non ciblé d'images faciales provenant de l'internet ou de la vidéosurveillance* »⁴⁷⁴ est une pratique strictement interdite par le règlement. De la même manière, le recours à un système d'IA pour « *inférer les émotions d'une personne physique sur le lieu de travail et dans les établissements d'enseignement [...]* »⁴⁷⁵ est interdit, sauf si cela poursuit un objectif médical ou de sécurité⁴⁷⁶. Ces usages prohibés illustrent que tous les risques ne peuvent pas être mis en balance avec les bénéfices que les systèmes d'IA pourraient apporter à la société. Le législateur fait ici le choix fort d'interdire strictement l'utilisation de systèmes d'IA pour ces pratiques sans essayer une minimisation des risques associés. Par conséquent, le règlement européen ne dispose d'aucune autre exigence pour ces finalités de systèmes d'IA, ni sur leur conception ni sur leur utilisation.

Ensuite, certains systèmes d'IA présentent des risques spécifiques et font ainsi l'objet de dispositions particulières les concernant au titre IV de la proposition de règlement⁴⁷⁷. Ces risques spécifiques sont explicitement décrits par le règlement européen et représentent en réalité trois situations, à savoir : « *[...] les systèmes d'IA destinés à interagir avec des personnes physiques [...]* », les « *système de reconnaissance des émotions ou [...] système de catégorisation biométrique* » et enfin « *un système d'IA qui génère ou manipule des images ou des contenus audio ou vidéo [...]* ». L'article 50 du règlement prévoit des obligations de transparence particulières pour les fournisseurs des systèmes d'IA dans la première situation et aux utilisateurs de ces systèmes d'IA dans les deux autres cas. Dans ces trois scénarii, les acteurs ne sont soumis à aucune des autres obligations énoncées dans le règlement.

Enfin, les autres systèmes d'IA, à savoir ceux exposant à des risques moyens ou faibles, représentent le dernier niveau de risque mis en place par le règlement européen. Considéré comme un niveau de risque négligeable, les logiciels d'IA concernés ne se voient imposer aucune exigence contraignante spécifique à la conception d'un tel système⁴⁷⁸. Le législateur

⁴⁷³ « Règlement sur l'intelligence artificielle », 2024. Article 5.

⁴⁷⁴ *Ibid.* Article 5.1.e)

⁴⁷⁵ *Ibid.* Article 5.1.f).

⁴⁷⁶ *Ibid.*

⁴⁷⁷ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final.

⁴⁷⁸ Le fournisseur d'un système d'IA non critique n'est pas pour autant exonéré de toute obligation juridique, le droit de l'IA ne venant que se superposer aux régimes juridiques déjà existants.

européen se limite à inciter les fournisseurs de ces systèmes d'IA à mettre en place des codes de conduites :

« Le comité européen de l'IA et les États membres encouragent et facilitent l'élaboration de codes de conduite, y compris les mécanismes de gouvernance associés, destinés à favoriser l'application volontaire aux systèmes d'IA autres que les systèmes d'IA à haut risque de tout ou partie des exigences énoncées au titre III, chapitre 2, du présent règlement, en tenant compte des solutions techniques disponibles et des meilleures pratiques de l'industrie permettant l'application de ces exigences. »⁴⁷⁹

Sans fournir plus de précisions quant aux mesures qui pourraient être mises en place pour « encourager et favoriser » ces codes de conduite, la Commission européenne renvoie à la bonne volonté des acteurs dans la mise en place d'un environnement de production des systèmes d'IA non critiques digne de confiance⁴⁸⁰. Imposer des obligations contraignantes à tous les systèmes d'IA indépendamment du risque, a été envisagé par l'Union européenne mais rapidement abandonné en raison des coûts économiques trop importants que cette option faisait peser sur les entreprises et la perte de compétitivité au niveau mondial des entreprises européennes⁴⁸¹. Par conséquent, les systèmes d'IA à moyens et faibles risques se révèlent bien moins essentiels à étudier que les systèmes d'IA à haut risque.

L'articulation entre la notion de risque élevé et les systèmes d'IA à usage général. Initialement absents de la proposition de règlement européen sur l'IA communiquée par la Commission européenne en 2021, les systèmes d'IA à usage général désignent :

« Un système d'IA basé sur un modèle d'IA à usage général, capable de répondre à des besoins divers, tant pour une utilisation directe que pour l'intégration dans d'autres systèmes d'IA. »⁴⁸²

Ainsi, la définition d'un système d'IA à usage général est intimement liée à la notion de modèle d'IA à usage général, qui est également définie par le règlement européen comme :

« Un modèle d'IA, y compris lorsqu'il est entraîné à l'aide d'une grande quantité de données en utilisant l'auto-supervision à grande échelle, qui fait preuve d'une grande généralité et qui est capable d'exécuter avec compétence un large éventail de tâches distinctes, quelle que soit la manière dont le

⁴⁷⁹ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final.

⁴⁸⁰ « EU would also promote industry-led initiatives for AI in an effort to advance 'soft law' but would not establish any framework for such voluntary codes. » - COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021. P.38

⁴⁸¹ *Ibid.*

⁴⁸² « Règlement sur l'intelligence artificielle », 2024. Version du compromis final.

*modèle est mis sur le marché, et qui peut être intégré dans une variété de systèmes ou d'applications en aval. Cela ne couvre pas les modèles d'IA qui sont utilisés avant leur mise sur le marché pour des activités de recherche, de développement et de prototypage. »*⁴⁸³

Les systèmes d'IA, et les modèles, à usage général n'ont pas de finalité spécifique associée. Ils se caractérisent par « *la capacité d'exécuter de manière compétente un large éventail de tâches distinctes.* »⁴⁸⁴. L'exemple le plus connu de modèle à usage général est GPT-3 (et GPT-4 pour la version suivante payante) et son système de discussion associé : Chat-GPT. Du fait de leur polyvalence, l'approche par les risques contenue dans le règlement européen devient alors beaucoup moins pertinente face à ce type de logiciel. En effet, un même système d'IA peut être utilisé pour différentes finalités et donc engendrer des risques de différents niveaux. Par conséquent, le législateur européen a décidé d'élaborer des obligations spécifiques pour les fournisseurs de ces systèmes d'IA, notamment l'établissement de documentations relatives à la procédure de conception du système⁴⁸⁵ ou aux modalités de réutilisation du système d'IA à usage général par d'autres fournisseurs⁴⁸⁶. Contrairement au risque élevé, la qualification de système d'IA à usage général n'entraîne que peu d'obligations et aucune ne concerne le fonctionnement du logiciel. Pour autant, s'intéresser aux systèmes d'IA à risque élevé n'implique pas d'exclure de l'analyse les systèmes d'IA à usage général. La première catégorie est caractérisée par le niveau de risque lié à sa finalité tandis que la seconde se définit au travers de sa capacité à « *répondre à des besoins divers, tant pour une utilisation directe que pour l'intégration dans d'autres systèmes d'IA* »⁴⁸⁷. Par ailleurs, le règlement reconnaît la possibilité pour un système d'IA à usage général d'être utilisé en tant que système d'IA à haut risque :

*« Les systèmes d'IA à usage général qui peuvent être utilisés comme systèmes d'IA à haut risque ou comme composants de systèmes d'IA à haut risque au sens de l'article 6 doivent se conformer aux exigences établies au titre III, chapitre 2, du présent règlement. »*⁴⁸⁸

⁴⁸³ Ibid. Version du compromis final.

⁴⁸⁴ Ibid. Considérant 97.

⁴⁸⁵ « 1. Les fournisseurs de modèles d'IA à usage général doivent : (a) établir et tenir à jour la documentation technique du modèle, y compris son processus de formation et d'essai et les résultats de son évaluation, qui contient, au minimum, les éléments figurant à l'annexe [...], afin de la fournir, sur demande, au Comité européen sur l'IA et aux autorités nationales compétentes ; » - Ibid. Version du compromis final. Article 52c.1.a)

⁴⁸⁶ « (b) élaborer, tenir à jour et mettre à disposition des informations et de la documentation aux fournisseurs de systèmes d'IA qui ont l'intention d'intégrer le modèle d'IA à usage général dans leur système d'IA. » - Ibid. Version du compromis final. Article 52c.1.b)

⁴⁸⁷ Ibid.

⁴⁸⁸ Ibid.

L'application du régime des systèmes d'IA à risque élevé sur des systèmes d'IA à usage général, le cas échéant, illustre encore une fois l'étendue et l'importance de la notion de risque élevé au sein du règlement européen sur l'IA.

Les risques systémiques entraînés par certains systèmes d'IA à usage général. La catégorie des modèles d'IA à usage général possède une échelle de risques qui lui est propre. Ces logiciels sont divisés en deux sous-catégories : ceux présentant des risques systémiques lors de leur utilisation et ceux n'en présentant pas. Contrairement au règlement européen sur les services numériques, le règlement européen sur l'IA ne décrit pas explicitement les risques systémiques mais donne une définition des « risques systémiques au niveau européen » :

« Un risque spécifique aux systèmes d'IA à usage général avec des capacités à fort impact, ayant un impact significatif sur le marché intérieur en raison de sa portée, et ayant des effets négatifs réels ou raisonnablement prévisibles sur la santé, la sûreté, la sécurité publique, les droits fondamentaux ou la société dans son ensemble, pouvant se propager à grande échelle dans la chaîne de valeur. »⁴⁸⁹

Il existe donc une double condition cumulative pour considérer qu'un système d'IA à usage général présente des risques systémiques. Premièrement, les risques auxquels expose l'utilisation du logiciel doivent pouvoir « se propager à grande échelle » et avoir « des effets négatifs réels ou raisonnablement prévisibles sur la santé, la sûreté, la sécurité publique, les droits fondamentaux ou la société dans son ensemble ». Cette condition est liée aux caractéristiques des risques produits par le logiciel. Deuxièmement, la notion de risques systémiques ne concerne que les systèmes d'IA à usage général qui possèdent des « capacités à fort impact ». De façon étonnante, cette condition n'est pas liée aux risques en eux-mêmes mais au système d'IA auxquels ils sont associés. Plus précisément, les « capacités à fort impact » désignent un système d'IA à usage général dont « les capacités correspondent ou dépassent les capacités enregistrées par les modèles d'IA à usage général les plus avancés »⁴⁹⁰. Le type de capacités dont il est question dans cette définition n'est pas clairement indiqué par le législateur européen. Cependant, le règlement sur l'IA énonce une présomption de capacités à fort impact lorsque le nombre de calculs effectués lors de l'entraînement du système d'IA à

⁴⁸⁹ Ibid. Traduction réalisée par l'auteur.

⁴⁹⁰ Ibid. Version du compromis final. Traduction réalisée par l'auteur.

usage général dépasse un certain seuil⁴⁹¹. L'identification des risques systémiques est donc conditionnée en partie par des caractéristiques de conception du logiciel. La notion n'est donc pas incompatible avec celle de risque élevé. En d'autres termes, un système d'IA à usage général utilisé comme système d'IA à haut risque peut présenter aussi des risques systémiques. À nouveau, la qualification de risque élevé est cumable avec cette nouvelle catégorie juridique de systèmes d'IA. La décision de concentrer notre analyse sur les systèmes d'IA à risque élevé inclut par conséquent une partie des systèmes d'IA à usage général présentant des risques systémiques. De plus, cela conforte la place centrale occupée par la notion de risque élevé dans le règlement européen sur l'IA.

⁴⁹¹ « Un modèle d'IA à usage général est présumé avoir des capacités d'impact élevées conformément au paragraphe 1, point a), lorsque la quantité cumulée de calcul utilisée pour son entraînement, mesurée en opérations à virgule flottante (FLOP), est supérieure à 10^{25} . » - Ibid. Version du compromis final. Traduction réalisée par l'auteur.

CONCLUSION DE LA SECTION I :

La proposition réalisée en 2021 par la Commission européenne fait le choix audacieux d'encadrer spécifiquement l'IA. Au regard des difficultés définitionnelles rencontrées par la communauté scientifique, la délimitation du champ d'application matériel d'un tel règlement est une tâche particulièrement complexe. L'Union européenne élabore ainsi la notion de système d'IA comme objet juridique central de son texte et comme condition d'application matérielle de son règlement. Au cours des travaux préparatoires, la définition du système d'IA a été significativement élargie et éloignée du lexique scientifique afin de limiter les risques d'obsolescence juridique du texte. Le législateur européen choisit alors d'imposer une obligation de certification européenne aux systèmes d'IA à haut risque préalablement à toute entrée sur le marché européen. Le risque élevé est le niveau de risque le plus important autorisé par l'Union européenne. En comparaison avec les autres niveaux de risques, les systèmes d'IA présentant un risque élevé font donc l'objet des contraintes d'utilisation les plus sévères⁴⁹². Apparaissant ainsi comme le cœur du règlement sur l'IA, il convient donc d'appliquer notre grille d'analyse aux exigences que le texte impose aux acteurs impliqués dans la conception de ces systèmes d'IA à haut risque.

Section II : Une conformité juridique ex ante pour les systèmes d'IA à haut risque

Annonce de plan. Au regard des besoins soulevés par les textes éthiques et par le droit positif, le texte de la Commission européenne concentre de grandes attentes de la part des acteurs de la communauté. Bien qu'il ne soit pas le seul outil de la stratégie européenne sur l'IA⁴⁹³, l'approche horizontale du règlement sur l'IA en fait un symbole de l'intégration par le droit des différents enjeux normatifs associés à ces technologies. L'encadrement des systèmes d'IA impose ainsi une conformité juridique préalable à la mise sur le marché ou la mise en service de tous les systèmes d'IA à haut risque. Au travers de son règlement, l'Union européenne

⁴⁹² Les systèmes d'IA présentant des risques inacceptables pour les individus n'entrent pas dans cette comparaison car leur utilisation est tout simplement interdite par le texte européen.

⁴⁹³ Voir notamment la *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, (2022) ainsi que la réforme de la directive européenne relative à la responsabilité du fait des produits défectueux : COMMISSION EUROPÉENNE, *Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux*, 28 septembre 2022.

aborde et encadre tous les enjeux de fond liés à l'IA par le biais d'exigences essentielles (**Sous-section 1**). Cependant, la stratégie choisie par l'Union européenne conduit à déléguer à des normes complémentaires les problématiques de fond de l'encadrement de l'IA, à savoir la précision et l'articulation des exigences (**Sous-section 2**).

Sous-section 1 : Les exigences essentielles relatives aux systèmes d'IA

Annonce de plan. Fondé principalement sur l'article 114 du TFUE⁴⁹⁴, le règlement européen met en avant une certaine vision de l'IA. En effet, l'Union européenne perçoit et traite l'IA comme un produit à encadrer sur le marché européen. Le choix du système d'IA comme objet juridique central de la réglementation en est une des conséquences majeures. Au sein du règlement, le législateur européen s'inspire des nombreux enjeux relevés par les textes éthiques et le droit positif pour élaborer un régime de conformité des systèmes d'IA à haut risque (**§1**). Une seconde conséquence déterminante du choix de considérer le système d'IA comme un produit est la stratégie d'encadrement choisie par l'Union européenne. Celle-ci a en effet recours à une approche traditionnelle, mais non sans conséquences, de la réglementation des produits en Europe : la « nouvelle approche » (**§2**).

§1. Les obligations juridiques posées pour les systèmes d'IA à haut risque

Annonce de plan. Les nouvelles obligations relatives aux systèmes d'IA à haut risque sont contenues au sein des chapitres 2 et 3 du titre III du règlement européen sur l'IA. Le chapitre 2 concerne les exigences imposées directement aux systèmes d'IA tandis que le chapitre 3 dispose des obligations auxquelles les différents acteurs concernés sont soumis (fournisseurs, dépoyeurs, importateurs, etc.). Bien qu'elles soient toutes constitutives de la conformité d'un système d'IA, ces exigences seront étudiées selon deux catégories distinctes. Dans un premier

⁴⁹⁴ De façon très secondaire et spécifique, le règlement européen sur l'IA se fonde également sur l'article 16 du TFUE relatif à la protection des données à caractère personnel : « En outre, étant donné que la présente proposition contient certaines règles spécifiques sur la protection des personnes physiques en ce qui concerne le traitement des données à caractère personnel [...], il convient de fonder le présent règlement, dès lors que ces règles spécifiques sont concernées, sur l'article 16 du TFUE. » - « Règlement sur l'intelligence artificielle », 2024. Exposé des motifs de la Commission européenne.

temps, l'analyse portera sur les exigences relatives au fonctionnement du système d'IA. Ces exigences seront appelées ci-après « exigences substantielles » **(A)**. Dans un second temps, nous nous attarderons sur les exigences relatives aux processus de mise en conformité, dénommées ci-après « exigences processuelles » **(B)**.

A - Les exigences substantielles pour un système d'IA conforme.

La recherche d'un risque résiduel acceptable. Dès le début de la section II du chapitre 3, le règlement européen sur l'IA aborde la question des risques liés à l'utilisation du système d'IA. L'obligation de départ est la gestion de tous les risques liés aux systèmes d'IA avec pour objectif leur suppression. Mais de façon assez lucide, le législateur européen considère que ce but ne pourra pas être atteint dans toutes les situations. Les risques qui résisteraient à toute mesure de suppression sont alors appelés des « risques résiduels ». Ainsi, les fournisseurs de systèmes d'IA se voient imposer une obligation de minimisation des risques résiduels⁴⁹⁵.

Selon le texte européen, les risques résiduels doivent être envisagés sous deux aspects : individuel et global. Premièrement, le fournisseur du système d'IA à haut risque doit identifier chaque danger associé avec les utilisations conforme « *à sa destination [du système d'IA] ou dans des conditions de mauvaise utilisation raisonnablement prévisible.* »⁴⁹⁶. Chaque danger se voit associer un niveau de risque que le fournisseur doit tenter de supprimer. Une fois les mesures de gestion des risques mises en place, le fournisseur estime le niveau de risque résiduel attaché à chacun des dangers. Le niveau de risque résiduel est alors évalué individuellement, danger par danger. Le second aspect des risques résiduels est leur estimation au niveau du système d'IA. Les fournisseurs doivent s'interroger sur le niveau de « *risque résiduel global lié au système d'IA à haut risque [...]* ». En d'autres termes, le risque résiduel global est composé de la somme des risques résiduels associés à chaque danger auquel exposera l'utilisation du système d'IA.

Le règlement européen impose que la minimisation des risques résiduels, pris individuellement et globalement, atteigne un niveau « *acceptable* ». Ce niveau acceptable de risques résiduels est un seuil que le fournisseur d'un système d'IA à haut risque doit atteindre pour satisfaire son obligation. Par exemple, un fournisseur de système d'IA parvient à réduire son niveau de risque

⁴⁹⁵ *Ibid.* Article 9.

⁴⁹⁶ *Ibid.* Article 19.4.

pour atteindre un seuil acceptable de risques résiduels. Malgré cela, il est possible que les risques résiduels, pris dans leur globalité, ne soient pas réduits à un niveau d'acceptable. Cela peut s'expliquer, par exemple, par le nombre de dangers auxquels expose le système d'IA évalué. Il est alors nécessaire de distinguer la notion de risque résiduels entre le niveau individuel et le niveau global. Cependant, les modalités d'établissement de ce seuil d'acceptabilité ne sont pas détaillées par le règlement européen. Il semble donc que cette tâche incombe au fournisseur lui-même ou aux organismes de normalisation. Les détails des mesures de gestion des risques et de la définition du seuil d'acceptabilité sont donc encore inconnus.

L'énonciation de critères de qualité pour les données utilisées. Ensuite, la Commission européenne s'attaque aux données et à leur gouvernance lors de la conception des systèmes d'IA⁴⁹⁷. Les technologies d'IA ont mis en avant une utilisation nouvelle des données, ces dernières pouvant être utilisées aussi lors de la conception du logiciel et plus seulement lors de son utilisation. Les données de conception sont un élément central pour certaines approches de l'IA, notamment celles relevant de l'apprentissage automatique. La présence de biais discriminatoires dans le fonctionnement d'un système d'IA est causée en majeure partie par des problèmes relatifs aux données de conception⁴⁹⁸. Désormais, les données de conception sont envisagées sous un prisme quantitatif, c'est-à-dire en tant qu'ensemble global, plus que sous un prisme qualitatif, à savoir la qualité de chaque donnée pris individuellement⁴⁹⁹. Cette évolution dans la manière de percevoir les données pour la conception apparaît dans le règlement sur l'IA qui préfère le terme de « *jeux de données* »⁵⁰⁰ à celui de « données »⁵⁰¹. Les obligations relatives aux données et à la gouvernance des données portent ainsi sur les jeux de données et non sur les données en elles-mêmes.

L'expression « jeux de données de conception » est employée ici pour désigner trois catégories de jeux de données désignées par l'article 10 du règlement européen sur l'IA, à savoir les « *jeux de données d'entraînement, de validation et de test* »⁵⁰². Dans un premier temps, les jeux de

⁴⁹⁷ Ibid. Article 10.

⁴⁹⁸ À titre d'exemple : MATTU Jeff Larson Julia Angwin, Lauren Kirchner, Surya, *How We Analyzed the COMPAS Recidivism Algorithm*, [<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=dvecePGfOFm-2Axng71w4PbbV6wY27tC>].

⁴⁹⁹ ZUBOFF Shoshana, « Big other: surveillance capitalism and the prospects of an information civilization », *Journal of Information Technology*, vol. 30, n° 1, 1 mars 2015, p. 75-89. ZUBOFF Shoshana, *L'Âge du capitalisme de surveillance*, Zulma, 2020.

⁵⁰⁰ « Règlement sur l'intelligence artificielle », 2024. Article 10.

⁵⁰¹ En informatique, le terme « jeu » est un synonyme de celui « d'ensemble ». Par exemple, un jeu de caractères désigne un « ensemble convenu et fini de caractères reconnus par les circuits d'un ordinateur ou d'une unité d'entrée-sortie » : <https://www.cnrtl.fr/definition/jeu>

⁵⁰² « Règlement sur l'intelligence artificielle », 2024. Article 10.

données d'entraînement désignent les ensembles des données utilisés lors du « processus d'apprentissage automatique pendant lequel le système d'IA construit un modèle à partir de ces données »⁵⁰³. Dans un second temps, les jeux de données de validation sont les ensembles de données utilisés pendant la phase de validation, c'est-à-dire le « processus consistant à expérimenter, observer et optimiser le comportement du système lors de son exécution »⁵⁰⁴. Dans un dernier temps, le processus de validation est défini comme suivant :

« Le test est un processus consistant à évaluer les performances d'un système et à rechercher des erreurs liées à l'exécution d'un algorithme ou d'un programme en s'appuyant sur des jeux de données d'entrée n'ayant pas été utilisés lors de la phase d'entraînement. »

Au premier abord, la phase de validation et la phase de test peuvent être difficiles à différencier. La phase de validation a lieu pendant le processus de conception tandis que la phase de test a lieu à la fin de cette phase. Pour illustrer autrement cette distinction, il est possible de faire une analogie avec un élève et un enseignant. Dans cet exemple, l'élève est le logiciel que l'on souhaite entraîner afin qu'il devienne compétent dans une matière. La phase d'entraînement représente les cours qui lui sont dispensés. La phase de validation représente le contrôle continu qui permet à l'enseignant d'évaluer si l'apprentissage se déroule correctement ou non (et modifier les cours le cas échéant). La phase de test est l'examen final de l'étudiant où l'on va évaluer sa performance dans la maîtrise de la matière.

Au vu de des termes employés, le législateur européen a souhaité traiter particulièrement les nouvelles problématiques relatives aux données que l'on retrouve dans les approches de l'IA par apprentissage automatique. Pour autant, les dispositions de gouvernance des données ne sont pas limitées à ces approches et concernent l'ensemble des systèmes d'IA⁵⁰⁵. De façon intéressante, la formulation de ces exigences relatives aux données et à leur gouvernance rappelle que le règlement européen sur l'IA a été pensé au départ pour répondre à une nécessité d'encadrer les spécificités des technologies d'IA. Par conséquent, les exigences ne sont pas

⁵⁰³ Pour la définition du jeu de données d'entraînement, voir : <https://www.cnil.fr/fr/definition/ensemble-dentrainementdapprentissage>. Pour la définition de la phase d'entraînement, voir : <https://www.cnil.fr/fr/definition/entrainement-ou-apprentissage>

⁵⁰⁴ <https://www.cnil.fr/fr/definition/validation-ia>

⁵⁰⁵ Les systèmes d'IA utilisant d'autres technologies que l'apprentissage automatique n'échappent pas pour autant aux exigences sur les données : « Des pratiques appropriées en matière de gouvernance et de gestion des données s'appliquent au développement de systèmes d'IA à haut risque autres que ceux qui font appel à des techniques impliquant l'entraînement de modèles [...] » - « Règlement sur l'intelligence artificielle », 2024. Article 10.6.

pensées pour l'ensemble des technologies contenues dans la nouvelle définition élargie du système d'IA.

Dans un premier temps, les fournisseurs de systèmes d'IA se voient ainsi imposer explicitement des « *critères de qualité* »⁵⁰⁶ qui prennent la forme d'une obligation de mettre en place des « *pratiques appropriées en matière de gouvernance et de gestion des données* ». Le texte met en avant, au travers de l'expression « *en particulier* », une liste non-exhaustive d'opérations déterminantes sur les données pour satisfaire ces critères de qualité :

- « (a) les choix de conception pertinents ;
- (b) la collecte de données ;
- (c) les opérations de traitement pertinentes pour la préparation des données, telles que l'annotation, l'étiquetage, le nettoyage, l'enrichissement et l'agrégation ;
- (d) la formulation d'hypothèses pertinentes, notamment en ce qui concerne les informations que les données sont censées mesurer et représenter ;
- (e) une évaluation préalable de la disponibilité, de la quantité et de l'adéquation des jeux de données nécessaires ;
- (f) un examen permettant de repérer d'éventuels biais ;
- (g) la détection d'éventuelles lacunes ou déficiences dans les données, et la manière dont ces lacunes ou déficiences peuvent être comblées. »⁵⁰⁷

Ces pratiques peuvent se diviser en deux catégories. D'un côté, le règlement énonce des opérations déjà courantes dans le cycle de conception d'un système d'IA (la collecte de données, la formulation d'hypothèses pertinentes, les choix de conceptions, etc.). De l'autre côté, le texte européen formule des mesures moins courantes qu'ils souhaitent voir appliquées dans la phase de conception (l'examen d'identification des biais, la détection d'éventuelles lacunes dans les données, l'évaluation de la disponibilité des jeux de données etc.).

L'ensemble de ces pratiques a pour objectif de respecter les exigences de qualité du règlement relatives aux jeux de données de conception, à savoir que les « *« jeux de données d'entraînement, de validation et de test soient pertinents, représentatifs, exempts d'erreurs et*

⁵⁰⁶ *Ibid.* Article 10.1.

⁵⁰⁷ *Ibid.* Article 10.2.

complets »⁵⁰⁸. Les niveaux acceptables d'erreur, de pertinence, d'erreurs et de complétude des jeux de données de conception ne sont pas déterminés par le texte. Cependant, ces niveaux d'acceptabilité doivent être établis au regard du contexte d'utilisation du système d'IA⁵⁰⁹. À nouveau, les exigences sur les jeux de données doivent être satisfaites aussi bien pour chaque jeu de donnée pris individuellement que sur l'ensemble des jeux de données pris au global⁵¹⁰.

L'obligation d'examiner les biais est une avancée plus que bienvenue au regard de l'immense problématique qu'est la discrimination algorithmique⁵¹¹. La lutte contre ce phénomène souffrait d'un problème majeur : l'accès aux données de discrimination. Pour identifier, mesurer ou même corriger de potentiels biais dans les données, avoir accès aux variables responsables de ces biais est un avantage significatif⁵¹². Or, s'il s'agit d'un biais sur la distinction homme/femme, il est nécessaire d'obtenir l'information sur le sexe des individus en question. Ces informations sont, a minima, des données à caractère personnel au sens de l'article 4 du règlement (UE) 2016/679 relatif à la protection des données⁵¹³, voire des catégories particulières de données au sens de l'article 9 dudit texte européen⁵¹⁴. Leur traitement, aussi bien intentionné soit-il, est par principe interdit. L'article 10.5 du règlement sur l'IA ouvre alors une possibilité dans le traitement de ces données. La lutte contre les biais algorithmiques permet aux fournisseurs de systèmes d'IA de traiter ces données sensibles dès lors que ce traitement est nécessaire et qu'il est accompagné de « *garanties appropriées pour les droits et libertés fondamentaux des personnes physiques [...]* ». L'objectif est de réduire les biais présents dans les données jusqu'à un niveau acceptable.

⁵⁰⁸ *Ibid.* Article 10.3

⁵⁰⁹ « *Les jeux de données d'entraînement, de validation et de test tiennent compte, dans la mesure requise par la destination, des caractéristiques ou éléments propres au contexte géographique, comportemental ou fonctionnel spécifique dans lequel le système d'IA à haut risque est destiné à être utilisé.* » - Article 10.4 RIA.

⁵¹⁰ « *Ces caractéristiques des jeux de données peuvent être présentes au niveau des jeux de données pris individuellement ou d'une combinaison de ceux-ci.* » - Article 10.3 RIA.

⁵¹¹ Voir le paragraphe 1 de la section 1 du chapitre 1 du présent titre.

⁵¹² XIANG Alice, *Reconciling Legal and Technical Approaches to Algorithmic Bias*, Rochester, NY, Social Science Research Network, 2021.

⁵¹³ « *toute information se rapportant à une personne physique identifiée ou identifiable* » - « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 4.1).

⁵¹⁴ « *des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique* » - *Ibid.* Article 9.1.

Les exigences relatives aux données de conception ainsi que la disposition spécifique aux biais dans les données sont des points centraux pour cette réglementation. D'une part, les problèmes liés à l'utilisation des données est une des spécificités disruptives des technologies d'IA que la Commission européenne a utilisé pour justifier son texte. D'autre part, les données de conception sont directement liées aux enjeux de fiabilité technique et de *fairness* des systèmes d'IA. La décision d'encadrer ces jeux de données est donc cruciale car elle permet de répondre, en partie, aux besoins de fond ayant émergé de l'environnement normatif de l'IA.

L'exigence de transparence envers les déployeurs des systèmes d'IA. Alors que la transparence algorithmique est depuis des années une contrainte réservée aux déployeurs des logiciels, l'article 13 du règlement sur l'IA impose ici une obligation de transparence aux fournisseurs du système d'IA à haut risque :

*« La conception et le développement des systèmes d'IA à haut risque sont tels que le fonctionnement de ces systèmes est suffisamment transparent pour permettre aux déployeurs d'interpréter les sorties d'un système et de les utiliser de manière appropriée. Un type et un niveau adéquats de transparence sont garantis afin de veiller au respect des obligations pertinentes incombant au fournisseur et au déployeur énoncées à la section 3. »*⁵¹⁵

Directement destinée à la phase de conception, cette exigence crée toutefois un pont explicite avec la phase d'utilisation du système d'IA à haut risque et ses utilisateurs. L'obligation poursuit alors un triple enjeu. Premièrement, elle traite de l'opacité caractéristique des technologies d'IA en imposant au fournisseur du logiciel de garantir un niveau suffisant de transparence. Deuxièmement, elle garantit la possibilité au déployeur de réaliser ses exigences imposées par le règlement, telle que l'obligation de surveillance du fonctionnement du système d'IA à haut risque⁵¹⁶. Troisièmement et dernièrement, l'obligation de transparence répond à un besoin d'articulation entre le règlement sur l'IA et les potentielles exigences de transparence déjà imposées en droit positif au déployeur lors de l'utilisation d'un logiciel⁵¹⁷.

⁵¹⁵ « Règlement sur l'intelligence artificielle », 2024. Article 13.1.

⁵¹⁶ « Les déployeurs surveillent le fonctionnement du système d'IA à haut risque sur la base de la notice d'utilisation [...] » - Ibid. Article 26.5.

⁵¹⁷ Pour une description des règles de transparence applicables aux systèmes d'IA en droit positif, voir §1. Les mécanismes juridiques de la transparence.

La facilitation du contrôle du déployeur sur la machine. Ensuite, le règlement sur l'IA impose au fournisseur du système d'IA à haut risque l'obligation de faciliter le contrôle humain sur le logiciel lors de son utilisation :

« La conception et le développement des systèmes d'IA à haut risque permettent, notamment au moyen d'interfaces homme-machine appropriées, un contrôle effectif par des personnes physiques pendant leur période d'utilisation. »

Le contrôle humain vise à prévenir ou à réduire au minimum les risques pour la santé, la sécurité ou les droits fondamentaux qui peuvent apparaître lorsqu'un système d'IA à haut risque est utilisé conformément à sa destination ou dans des conditions de mauvaise utilisation raisonnablement prévisible, en particulier lorsque de tels risques persistent malgré l'application d'autres exigences énoncées dans la présente section. »⁵¹⁸

Le contrôle humain, comme la transparence, est une notion directement inspirée par les principes éthiques de l'IA élaborés ces dernières années⁵¹⁹. Cette inspiration s'observe au travers de la fonction accordée au contrôle humain. Celui-ci est un moyen pour atteindre un objectif, en l'espèce la protection de la santé, la sécurité et les droits fondamentaux. De façon similaire à « l'intervention humaine » de l'article 22.3 du RGPD⁵²⁰, le règlement européen sur l'IA désigne l'humain comme une mesure de protection. Le texte tire cependant des leçons du RGPD en précisant que l'objectif des mesures mises en place doit être un contrôle humain effectif, un oubli critiqué dans le texte européen sur les données personnelles⁵²¹. Le règlement sur l'IA illustre ces mesures de contrôle humain par des « *interfaces homme-machine appropriées* »⁵²² mais qui doivent être réalisées spécifiquement par le fournisseur. Néanmoins, dans le cas où la mise en application technique serait trop complexe, le fournisseur peut

⁵¹⁸ « Règlement sur l'intelligence artificielle », 2024. Articles 14.1 et 14.2.

⁵¹⁹ Voir C– Le principe de contrôle humain.

⁵²⁰ « Dans les cas visés au paragraphe 2, points a) et c), le responsable du traitement met en oeuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement [...] » - « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 22.3. Emphase ajoutée par l'auteur.

⁵²¹ G29, « Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679 », 2018. Le désormais CEPD recommande d'ailleurs, de façon non contraignante, ce niveau de contrôle humain : « Pour qu'il y ait intervention humaine, le responsable du traitement doit s'assurer que tout contrôle de la décision est significatif et ne constitue pas qu'un simple geste symbolique. Le contrôle devrait être effectué par une personne qui a l'autorité et la compétence pour modifier la décision. Dans le cadre de l'analyse, il convient de tenir compte de toutes les données pertinentes. ». Emphase ajoutée par l'auteur.

⁵²² « Règlement sur l'intelligence artificielle », 2024. Article 14.1.

également se limiter à identifier les mesures appropriées, laissant alors à l'utilisateur lui-même ou à un tiers la charge de leur réalisation⁵²³. En d'autres termes, le législateur européen laisse la porte ouverte aussi bien aux solutions intrinsèques qu'aux mesures *ex post*.

La performance du système d'IA à haut risque, une préoccupation centrale. Enfin, le dernier article du règlement européen sur l'IA prévoit une triple exigence relative à l'exactitude, la robustesse et la cybersécurité. Ces trois objectifs sont envisagés sous le prisme de la performance du logiciel. La cybersécurité y est davantage discutée comme un outil de préservation de la performance et du maintien du fonctionnement continu du système d'IA que comme une mesure de protection des droits et libertés des individus, notamment la protection de leurs données. Par exemple, l'alinéa 5 de l'article 15 indique : « *Les systèmes d'IA à haut risque résistent aux tentatives de tiers non autorisés visant à modifier leur utilisation ou leurs performances en exploitant les vulnérabilités du système.* ». Toutefois, cette concentration sur le fonctionnement technique du système d'IA est contrebalancée de deux manières. Premièrement, le droit européen contient plusieurs obligations de sécurité des systèmes informatiques, le règlement sur l'IA ne faisant qu'ajouter de nouvelles exigences spécifiques sans déroger à celles existantes. Le RGPD⁵²⁴, la directive européenne SRI 2⁵²⁵ sont deux exemples de textes imposant une obligation de sécurité à certains utilisateurs d'un système informatique. Secondement, la performance est directement liée au taux d'erreur du système

⁵²³ « *Les mesures de contrôle sont proportionnées aux risques, au niveau d'autonomie et au contexte d'utilisation du système d'IA à haut risque, et sont assurées au moyen d'un ou des deux types de mesures suivants: a) des mesures identifiées et, lorsque cela est techniquement possible, intégrées par le fournisseur dans le système d'IA à haut risque avant la mise sur le marché ou la mise en service de ce dernier; b) des mesures identifiées par le fournisseur avant la mise sur le marché ou la mise en service du système d'IA à haut risque et qui se prêtent à une mise en oeuvre par le déployeur.* » - Ibid. Article 14.3.

⁵²⁴ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) ». Article 32 « Sécurité du traitement » ; Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil », 2016. Article 29 « Sécurité du traitement » ; « Règlement (UE) 2018/1725 du Parlement européen et du Conseil du 23 octobre 2018 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union et à la libre circulation de ces données, et abrogeant le règlement (CE) n° 45/2001 et la décision n° 1247/2002/CE (Texte présentant de l'intérêt pour l'EEE.) ». Article 33 « Sécurité du traitement »

⁵²⁵ « Directive (UE) 2022/2555 du Parlement européen et du Conseil du 14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement (UE) no 910/2014 et la directive (UE) 2018/1972, et abrogeant la directive (UE) 2016/1148 (directive SRI 2) (Texte présentant de l'intérêt pour l'EEE) », 2022.

d'IA et donc permet indirectement de prévenir le plus possible les dommages que pourra causer un résultat algorithmique erroné dans ces utilisations critiques.

B - Les exigences processuelles pour un système d'IA conforme.

Au-delà des exigences de qualité attendues des systèmes d'IA produits, le règlement sur l'IA dispose également de nombreuses obligations découlant directement de sa logique de conformité. Les exigences processuelles remplissent deux fonctions. Dans un premier temps, elles sont des obligations à part entière destinées à surveiller le fonctionnement des systèmes d'IA afin de garantir leur conformité sur la durée, à savoir le cycle de vie d'un système d'IA. Dans un second temps, elles ont une finalité probatoire, permettant de constituer des preuves du bon fonctionnement du système d'IA et ainsi d'attester de la surveillance et de la conformité des systèmes d'IA lors des futurs contrôles réalisés avant et après l'entrée sur le marché européen.

La création de systèmes de gestion. Le premier système de gestion établi par le règlement européen sur l'IA porte sur la question des risques. L'objectif de minimisation des risques jusqu'à un niveau résiduel acceptable doit être atteint à l'aide d'un processus itératif appelé « système de gestion des risques »⁵²⁶. Énoncé à l'article 9, ce système est composé de quatre étapes :

« a) l'identification et l'analyse des risques connus et raisonnablement prévisibles associés à chacun des risques que le système d'IA à haut risque peut présenter pour la santé, la sécurité ou les droits fondamentaux lorsque le système d'IA à haut risque est utilisé conformément à sa destination ;

b) l'estimation et l'évaluation des risques susceptibles d'apparaître lorsque le système d'IA à haut risque est utilisé conformément à sa destination et dans des conditions de mauvaise utilisation raisonnablement prévisible ;

c) l'évaluation d'autres risques susceptibles d'apparaître, sur la base de l'analyse des données recueillies au moyen du système de surveillance après commercialisation visé à l'article 61 ;

⁵²⁶ « Les mesures de gestion des risques visées au paragraphe 2, point d), sont telles que le risque résiduel pertinent associé à chaque danger ainsi que le risque résiduel global lié aux systèmes d'IA à haut risque sont jugés acceptables. » - « Règlement sur l'intelligence artificielle », 2024. Article 9.5.

*d) l'adoption de mesures de gestion des risques appropriées et ciblées, destinées à faire face aux risques identifiés en vertu du point a) du présent paragraphe conformément aux dispositions des paragraphes suivants. »*⁵²⁷

A la fin du processus, tous les risques résiduels présents doivent alors être communiqués à l'utilisateur afin qu'il puisse en tenir compte lors de l'exploitation du système d'IA en question. La gestion des risques permet également de contribuer à la mise en conformité du système d'IA vis-à-vis des exigences de qualité.

Ce système de gestion des risques s'inscrit lui-même dans un système de gestion plus large encore qu'est le « système de gestion de la qualité⁵²⁸ ». Imposé aux fournisseurs de systèmes d'IA à haut risque, ce système de gestion se situe à un niveau plus global et prend la forme de stratégies, de procédures d'évaluation des systèmes d'IA, des référentiels techniques choisis (normes, normes techniques, etc.) et toutes autres procédures nécessaires au respect de l'ensemble des obligations prévues dans règlement européen. À nouveau, cette obligation impose la documentation afin de pouvoir vérifier *a posteriori* la bonne application des processus de mise en conformité des systèmes d'IA à haut risque. Ce mode de fonctionnement est loin d'être nouveau pour les entreprises car cela reprend très largement les logiques des normes ISO, en particulier la série des ISO 9000, déjà très largement appliquées dans le secteur informatique⁵²⁹.

L'apparition d'une obligation de documentation technique pour les systèmes d'IA. Si les systèmes de gestion permettent de conserver des preuves de l'application des processus réglementaires par le fournisseur du système d'IA à haut risque, la documentation technique⁵³⁰ quant à elle est la preuve de la conformité du système d'IA à haut risque au moment de la mise sur le marché européen et/ou la mise en service. Imposée aux fournisseurs de systèmes d'IA à haut risque⁵³¹, cette documentation technique sert alors de référentiel pour la vérification de la

⁵²⁷ *Ibid.* Version du compromis final. Traduction réalisée par l'auteur. Article 9.2.

⁵²⁸ *Ibid.* Article 17.

⁵²⁹ VEALE Michael et BORGESIU Frederik Zuiderveen, « Demystifying the Draft EU Artificial Intelligence Act », *arXiv:2107.03721 [cs]*, , 26 novembre 2021.

⁵³⁰ « Règlement sur l'intelligence artificielle », 2024. Article 11.

⁵³¹ *Ibid.* Article 18.

conformité du logiciel par tous les autres acteurs de l'écosystème énoncés par le règlement, des importateurs⁵³² aux autorités de contrôle en passant par les distributeurs⁵³³.

La recherche de la responsabilité facilitée par l'exigence de traçabilité. Dernière exigence processuelle imposée dans cette recherche de conformité, la traçabilité des actions réalisées par les systèmes d'IA à haut risque se traduit par la création de mécanisme de journaux informatiques, plus connus sous sa traduction anglaise de *logs*. Associés à un système d'IA, ces journaux sont des recueils d'enregistrement de données liées au fonctionnement du système d'IA en question. Les informations minimales constituant la traçabilité attendue ne sont pas énoncées globalement par le règlement européen sur l'IA. Les systèmes d'IA d'identification biométrique à distance, en « temps réel » et « a posteriori » des personnes physiques se voient précisées les informations à enregistrer dans leurs *logs*. Toutefois, les données énoncées par le texte sont trop spécifiques à leur finalité pour qu'une généralisation soit pertinente⁵³⁴. Les enregistrements obtenus devront être conservés par la personne en charge de ces *logs*, soit en principe le fournisseur du système d'IA, pendant une « *pendant une période adaptée à la finalité du système d'IA à haut risque, d'au moins six mois, sauf disposition contraire du droit de l'Union ou du droit national [...]* »⁵³⁵. Cette exigence de traçabilité ajoute à son tour d'autres traces du fonctionnement des systèmes d'IA à haut risque, des éléments enregistrés qui contribuent à la mise en place d'un système probatoire de conformité.

§2. La concentration sur les exigences essentielles : le recours à la nouvelle approche

Historique de la nouvelle approche. La « nouvelle approche » est un terme faisant référence une stratégie politique originale de l'Union européenne officialisée par une résolution du

⁵³² « Avant de mettre sur le marché un système d'IA à haut risque, les importateurs de ce système s'assurent que [...] (b) le fournisseur a établi la documentation technique conformément à l'Annexe IV ; [...] ; » - Ibid. Article 26 – Obligation des importateurs

⁵³³ « Avant de mettre un système d'IA à haut risque à disposition sur le marché, les distributeurs vérifient que le système d'IA à haut risque porte le marquage de conformité CE requis, qu'il est accompagné de la documentation et de la notice d'utilisation requises [...] » - Ibid. Article 27 – Obligation des distributeurs

⁵³⁴ La seule information présente dans l'article 12.4.d) dont l'enregistrement pourrait être exigée pour tous les systèmes d'IA à haut risque est « l'enregistrement de la période de chaque utilisation de chaque utilisation du système ».

⁵³⁵ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final. Traduction réalisée par l'auteur.

Conseil de 1985⁵³⁶. Avec ce texte, le Conseil des communautés européennes souhaitait s'approcher de son objectif d'établissement d'un marché unique européen qui connaissait alors deux difficultés majeures. Dans un premier temps, des organismes de normalisation nationaux publiaient chaque année de nombreuses normes techniques, sans aucune coordination particulière entre les États membres⁵³⁷. Cette production hétérogène multipliait les contradictions entre les normes techniques de chaque État, rendant alors risqué et difficile les échanges transfrontaliers de produits soumis à ces normes techniques nationales. Or, favoriser les échanges entre les États membres est précisément l'objectif poursuivi par l'UE. Ce premier problème était exacerbé par un second : ces normes techniques sont des normes privées. De ce fait, elles se situaient, d'une part, hors du champ de compétence de l'article 34 du TFUE⁵³⁸ et elles échappaient, d'autre part, aux mesures d'harmonisation de la Communauté européenne. « L'ancienne approche », par opposition à la « nouvelle approche », aurait conduit à remplacer chaque standard national par des spécifications techniques au niveau européen. La durée d'un tel chantier, la complexité du résultat, la priorité réelle de certains sujets standardisés ou encore le risque très important d'obsolescence de textes à l'élaboration aussi lente ont conduit à abandonner cette idée vouée à l'échec pour une solution différente⁵³⁹.

En 1983, le Conseil des communautés européennes décida donc de renforcer les normes européennes, et par la même occasion les organismes de normalisation européens que sont le CEN⁵⁴⁰ et le CENELEC⁵⁴¹, par le biais de la directive 83/189/EEC⁵⁴². Désormais, les organismes de normalisation nationaux doivent informer le CEN et le CENELEC de leurs projets de normes nationales, ces derniers devenant alors responsable de ce système de notification de normes. Ensuite, les organismes nationaux, tel que l'AFNOR⁵⁴³ en France, se sont vus imposer une obligation de suspendre tout projet de norme qui viendrait empiéter sur le

⁵³⁶ « Résolution du Conseil, du 7 mai 1985, concernant une nouvelle approche en matière d'harmonisation technique et de normalisation ».

⁵³⁷ SCHEPEL Harm, JOERGES Christian, TEUBNER Gunther et COLLINS Hugh, *The Constitution of Private Governance: Product Normes techniques in the Regulation of Integrating Markets*, Oxford ; Portland, Or, Hart Publishing, 2005.

⁵³⁸ À l'époque, il s'agissait de l'article 28 du Traité instituant la Communauté européenne.

⁵³⁹ SCHEPEL Harm, « The New Approach to the New Approach: The Juridification of Harmonized Normes techniques in EU Law », *Maastricht Journal of European and Comparative Law*, vol. 20, n° 4, 1 décembre 2013, p. 521-533.

⁵⁴⁰ Comité européen de normalisation.

⁵⁴¹ Comité européen de normalisation en électronique et en électrotechnique.

⁵⁴² Council Directive 83/189/EEC of 28 March 1983 laying down a procedure for the provision of information in the field of technical normes techniques and regulations

⁵⁴³ Association française de normalisation : <https://www.afnor.org/>

champ d'application d'une norme européenne future⁵⁴⁴. Enfin, la directive a consacré une procédure permettant à la Commission européenne de passer commande de normes techniques européennes aux organismes de normalisation européens. Le problème de normes techniques nationaux était donc résolu. Quant au problème de compétence de l'UE, c'est la résolution de 1985⁵⁴⁵ qui vient le régler en disposant que les normes techniques européennes valent désormais présomption de conformité aux réglementations produites par le législateur européen. La nouvelle approche était lancée. Malgré leur pouvoir incitatif considérable, ces normes techniques demeurent volontaires et les textes de l'Union européenne permettaient toujours aux autres entités de produire des alternatives conformes aux normes techniques européennes⁵⁴⁶.

Présentation de la stratégie actuelle *nouvelle approche*. La nouvelle approche désigne une stratégie européenne de normalisation qui prend la forme d'un renvoi aux normes techniques. Les directives européennes fondées sur cette approche n'établissent alors que des « exigences essentielles » de la réglementation⁵⁴⁷. Ensuite, la Commission européenne va confier un mandat aux organismes de normalisation européens (désormais au nombre de trois : CEN, CENELEC et ETSI⁵⁴⁸) pour réfléchir aux « spécifications techniques » et les produire sous la forme de normes harmonisées. Les normes techniques sont alors pensées comme des normes complémentaires, facultatives, aux exigences essentielles contenues dans le règlement européen⁵⁴⁹. Depuis 2012, les normes harmonisées entretiennent des liens plus étroits avec le droit européen que ce qui était prévu au début de cette stratégie⁵⁵⁰. Désormais, la CJUE détient un pouvoir de contrôle sur toute norme qui « *a pour effet d'entraver la commercialisation de*

⁵⁴⁴ Article 7 de la directive 93/189/CEE : « *Les États membres prennent toutes mesures utiles pour que leurs organismes de normalisation n'établissent pas ou n'introduisent pas de normes dans le domaine en cause pendant l'élaboration de la norme européenne [...]* »

⁵⁴⁵ « Résolution du Conseil, du 7 mai 1985, concernant une nouvelle approche en matière d'harmonisation technique et de normalisation ».

⁵⁴⁶ SCHAPEL Harm, « The New Approach to the New Approach », *op. cit.*

⁵⁴⁷ MAZEAU Laurène, « Responsabilité. Les enjeux de la normalisation technique dans le domaine de l'intelligence artificielle », *Cahiers Droit, Sciences & Technologies*, n° 8, 13 mars 2019, p. 225-231.

⁵⁴⁸ *European Telecommunications Normes techniques Institute* ou Institut européen des normes de télécommunications

⁵⁴⁹ Par principe, les normes européennes harmonisées ne sont pas obligatoires. Cependant, un récent arrêt de la Cour de justice de l'Union européenne indique que la présomption de conformité associée au respect de ces normes techniques harmonisées, ainsi que la quasi-impossibilité de recourir à des alternatives pour les opérateurs économiques, les rendent « *manifestement obligatoires* ». Elles font donc partie intégrante du droit de l'Union européenne et doivent donc respecter l'article 2 du TFUE imposant un accès libre au droit de l'Union : Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P. Par. 65 à 87.

⁵⁵⁰ « Règlement (UE) n° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012.

produits qui ne sont pas certifiés par ledit organisme. »⁵⁵¹. Le statut juridique des normes harmonisées a longtemps été incertain mais leur publication au Journal officiel de l'UE et la présomption de conformité qui découle de leur respect ont permis à la CJUE de les considérer comme des éléments appartenant au droit de l'Union⁵⁵². Cette nouvelle qualification rend ainsi compétente la CJUE pour interpréter les normes en question. La cour européenne détermine ainsi dans l'arrêt *Anstar*⁵⁵³ les règles d'interprétation de ces nouveaux textes européens⁵⁵⁴.

C'est donc dans cette approche nouvellement judiciairisée que le règlement sur l'IA vient s'inscrire. La Commission européenne élabore au travers de son règlement sur l'IA des exigences qu'elle considère comme essentielles à appliquer aux systèmes d'IA à haut risque⁵⁵⁵. Associées aux procédures de mise en conformité et autres obligations, ces exigences essentielles sont les éléments principaux souhaités par l'Union européenne⁵⁵⁶. Elles doivent être suffisamment claires pour permettre l'élaboration de normes techniques complémentaires, aussi bien celles produites par les organismes de normalisation européens⁵⁵⁷ que celles produites par les entreprises ou les autres organismes de normalisation privés. Le recours à la nouvelle approche pour la conformité technique des systèmes d'IA indique, d'une part, l'existence d'une sélection dans les exigences essentielles présentes au sein du règlement et, d'autre part, le choix d'un niveau de précision considéré comme acceptable pour concevoir le régime *ex ante* spécifique aux systèmes d'IA à haut risque.

La force contraignante des normes européennes harmonisées. Bien qu'elles « *jouent un rôle très important dans le marché intérieur [...]* »⁵⁵⁸, les normes techniques publiées au Journal officiel de l'UE ne sont pas obligatoires en droit⁵⁵⁹. En pratique toutefois, elles sont considérées

⁵⁵¹ Voir CJUE, 12 juillet 2012, aff., C-171/11, Fra.bo, ECLI:EU:C:2012:453, point 32

⁵⁵² *James Elliott Construction Limited contre Irish Asphalt Limited*, 2016. Point 40

⁵⁵³ *Procédure engagée par Anstar Oy*, 2017.

⁵⁵⁴ Elles sont au nombre de quatre : 1) « *se reporter au contenu de ladite norme, y compris à ses annexes, relatif à son champ d'application.* » ; 2) « *[...] interpréter une norme harmonisée à la lumière du mandat qui en est à l'origine.* » ; 3) « *[...] rechercher si la norme la plus récente n'a pas eu pour effet d'abroger la plus ancienne.* » ; 4) « *Les documents d'orientation publiés par des organismes nationaux ou internationaux de normalisation [...] peuvent constituer un guide utile aux fins de la mise en œuvre de cette norme.* »

⁵⁵⁵ « *Les exigences ex ante essentielles applicables aux systèmes d'IA à haut risque définies dans la présente proposition [...]* » - « *Règlement sur l'intelligence artificielle* », 2024. P. 5.

⁵⁵⁶ « *Règlement sur l'intelligence artificielle* », 2024. Objectif spécifique n° 2. P.100.

⁵⁵⁷ *Frabo SpA contre Deutsche Vereinigung des Gas- und Wasserfaches eV (DVGW) — Technisch-Wissenschaftlicher Verein*, 2012.

⁵⁵⁸ « *Règlement (UE) n° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE* », OJ L, 2012. Considérant 5.

⁵⁵⁹ *Ibid.* Article 2.

comme obligatoires en raison des avantages économiques et juridiques extrêmement importants qu'elles apportent :

*« Les normes techniques restent (en droit) volontaires, mais les acteurs économiques bénéficient de la présomption de conformité avec le droit (fondée sur les directives européennes) si les produits et les services sont conformes aux normes européennes harmonisées auxquelles la directive se réfère. Cela signifie que, dans la pratique, ces normes harmonisées sont presque obligatoires pour la plupart des acteurs économiques. »*⁵⁶⁰

Les normes européennes harmonisées sont *de facto* généralement la seule garantie de conformité juridique d'un produit acceptée par le marché, les solutions de remplacement étant irréalistes⁵⁶¹. La quasi-impossibilité de recourir à des normes techniques alternatives a été reconnue par le Tribunal⁵⁶² et la Cour de justice de l'Union européenne⁵⁶³. Dans une récente décision du 5 mars 2024, la CJUE a indiqué que les normes européennes harmonisées sont « *manifestement obligatoires* », notamment lorsqu'elles confèrent une présomption de conformité aux réglementations auxquelles elles sont associées⁵⁶⁴. Cette décision reconnaît non seulement la force contraignante réelle des normes européennes harmonisées sur les opérateurs économiques mais en tire enfin les conséquences. En effet, les juges européens considèrent que cette force obligatoire *de facto* fait de ces normes une partie intégrante du droit de l'Union. Par conséquent, elles doivent respecter l'exigence d'accès libre au droit de l'Union prévue par l'article 2 du TFUE. Attendue depuis longtemps, l'ouverture des normes européennes harmonisées est une avancée certaine pour les opérateurs économiques et les justiciables européens. Dans notre contexte, l'arrêt contribue à mettre en lumière l'importance de ces normes techniques, et des spécifications qu'elles contiennent, dans l'encadrement d'un produit, en l'espèce les systèmes d'IA à haut risque.

⁵⁶⁰ « « Standards are still (de jure) voluntary, but economic players get the presumption of compliance with the law (based on European directives) if products and services are in line with the European standards to which the directive refers. This implies that in practical terms these harmonised standards are almost obligatory for most economic players. » - EIM BUSINESS & POLICY RESEARCH, « Access to Standardisation - Study for the European Commission », Enterprise and Industry [DG], 2010. P.17. Traduction de l'auteur.

⁵⁶¹ « elles constituent en général la seule méthode acceptée sur le marché pour garantir la conformité au droit dérivé concerné de l'Union » - Cour de justice de l'Union européenne, 22 juin 2023, n° C-588/21 P. Par. 45.

⁵⁶² Tribunal de l'Union européenne (Deuxième chambre), 26 janvier 2017, n° T-474/15. Par. 67.

⁵⁶³ « il ressort du dossier que, en pratique, la quasi-totalité des consommateurs allemands n'achètent que des raccords en cuivre certifiés » - Cour de justice de l'Union européenne, 12 juillet 2012, n° C-171/11. Par. 30.

⁵⁶⁴ Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P. Par. 79. La décision porte sur quatre normes harmonisées différentes. Trois d'entre elles renvoient à la réglementation des jouets tandis que la quatrième porte sur la réglementation des produits chimiques.

Sous-section 2 : La responsabilisation des acteurs lors de la phase de conception

Annonce de plan. L'analyse des différentes obligations imposées aux systèmes d'IA et aux acteurs ciblés dans le règlement sur l'IA révèle plusieurs points quant au fonctionnement de la conformité technique attendue par l'Union européenne. Tout d'abord, la proposition européenne contient de nombreuses obligations imposant aux acteurs la constitution de documents sur la réalisation de la mise en conformité de leurs systèmes d'IA. Ces différentes obligations sont le symbole du mécanisme de redevabilité des acteurs lors de la phase de conception du système d'IA qu'il convient d'examiner (§1). Ensuite les exigences de qualité imposées aux systèmes d'IA à haut risque représentent le cœur de la proposition européenne. C'est autour d'elles que s'articulent les obligations processuelles et probatoires. Pour autant, il appert que la détermination des critères de qualité énoncés par la Commission européenne est renvoyée globalement aux normes techniques, les éléments contenus dans le texte se révélant relativement abstraits (§2).

§1. La mise en place d'une logique de redevabilité

Annonce de plan. Le système de mise en conformité mis en place par le règlement sur l'IA nécessite l'attribution de rôle et de responsabilité envers les différents acteurs concernés. Parmi les obligations imposées, le fournisseur du système d'IA se voit attribuer plusieurs obligations documentaires qui indiquent qu'il porte la charge de la preuve de la conformité du système d'IA produit (A). Cette responsabilisation des acteurs lors de la phase de conception s'inscrit dans le mouvement de l'Union européenne de mise en place d'un mécanisme de redevabilité *ex ante* (B).

A- Les obligations détaillées de documentation à la charge du fournisseur

La fonction probatoire du système de gestion de la qualité. Tout d'abord, le système de gestion de qualité, mécanisme déjà bien connu des acteurs privés, voit son contenu détaillé au

sein d'une liste constituée d'un total de treize éléments⁵⁶⁵. Le fournisseur de systèmes d'IA à haut risque est notamment contraint d'y indiquer les référentiels qu'il a employé pour établir la qualité de son système d'IA⁵⁶⁶, les procédures utilisées pour tester la qualité des systèmes d'IA⁵⁶⁷ ou encore le système de surveillance après commercialisation existant⁵⁶⁸. Au-delà de sa fonction d'aide à la « gestion de la qualité », ce système remplit clairement une fonction probatoire pour le cadre juridique des systèmes d'IA. Par exemple, le fournisseur de systèmes d'IA est dans l'obligation de documenter « *de manière méthodique et ordonnée sous la forme de politiques, de procédures et d'instructions écrites [...]* »⁵⁶⁹. En contenant l'entière des obligations et informations imposées, le document n'est plus seulement un recueil destiné à renseigner les tiers au développement du logiciel, il est aussi une référence sur laquelle se fondent lesdits tiers. Le système de gestion de la qualité est un des documents centraux dans les procédures d'évaluation de conformité⁵⁷⁰. En effet, la procédure de contrôle de conformité effectuée par un tiers notifié repose exclusivement sur un examen de deux documents, la documentation technique et le système de gestion de la qualité⁵⁷¹. Il en va de même pour la procédure de contrôle interne prévue à l'annexe VI du règlement européen⁵⁷². Par conséquent, le système de gestion de qualité est un élément de preuve crucial dans l'établissement de la conformité des systèmes d'IA à haut risque.

Le rôle des journaux d'enregistrement. Dans la même logique, les journaux d'enregistrement imposés à l'article 12 du règlement européen sur l'IA ne se limitent pas au suivi du bon fonctionnement du système d'IA par son fournisseur ou ses déployeurs. La traçabilité technique du fonctionnement se met au service d'une traçabilité juridique de la conformité du logiciel. Les *logs* deviennent des éléments de preuve du caractère continu de la conformité du logiciel pendant son cycle de vie en cas de contrôle et/ou de survenance d'un dommage⁵⁷³.

⁵⁶⁵ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final. Article 17.1.

⁵⁶⁶ « *des spécifications techniques, notamment des normes, à appliquer et, lorsque les normes harmonisées pertinentes ne sont pas appliquées intégralement, les moyens à utiliser pour faire en sorte que le système d'IA à haut risque satisfasse aux exigences énoncées au chapitre 2 du présent titre ;* » - Ibid. Article 11.1(e).

⁵⁶⁷ « *des procédures d'examen, de test et de validation à exécuter avant, pendant et après le développement du système d'IA à haut risque, ainsi que la fréquence à laquelle elles doivent être réalisées ;* » - Ibid. Article 10.1(d).

⁵⁶⁸ « *l'élaboration, la mise en oeuvre et le maintien d'un système de surveillance après commercialisation conformément à l'article 61;* » - Article 11.1(h)

⁵⁶⁹ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final. Article 17.1.

⁵⁷⁰ Pour plus d'informations sur la question du contrôle de conformité des systèmes d'IA à haut risque, voir Chapitre 2 : Le contrôle des preuves de conception.

⁵⁷¹ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final. Annexe VII : Conformité fondée sur l'évaluation du système de gestion de la qualité et l'évaluation de la documentation technique.

⁵⁷² Ibid. Annexe VI.

⁵⁷³ A - Les journaux d'enregistrement des systèmes d'IA.

Le rôle de la documentation technique. Enfin, la documentation technique du système d'IA à haut risque est également une obligation révélant une fonction probatoire essentielle. Elle fait l'objet d'une annexe dédiée qui permet à la Commission européenne d'entrer encore plus dans les détails de son contenu⁵⁷⁴ ainsi que de modifier plus facilement ce dernier en fonction de l'évolution des technologies⁵⁷⁵. En comparaison des obligations substantielles sur le fonctionnement du système d'IA, la documentation technique présente un niveau de précision exceptionnel. Sans pour autant être exhaustive, la description détaillée contenue dans la documentation technique impose aux fournisseurs de systèmes d'IA d'indiquer par exemple les « *éventuelles incidences discriminatoires* »⁵⁷⁶ une fois les procédures de tests réalisées. Face aux exigences de qualité diverses et encore floues, l'annexe IV précise aussi l'importance de renseigner les compromis que le fournisseur a fait entre les différentes contraintes qu'il a rencontré lors de la conception. La liste des précisions est encore longue et il n'est pas pertinent de la citer de façon complète. La documentation technique permet de s'apercevoir que les documents à établir par les acteurs de l'écosystème des technologies d'IA sont bien mis en place et précisés. La documentation technique est d'ailleurs le second élément central du contrôle de la conformité du système d'IA avec le système de gestion de la qualité⁵⁷⁷.

La durée de conservation de la documentation. La fonction probatoire de cette documentation est renforcée par l'indication d'une durée de conservation des documents produits. Les documents établis dans le cadre du système de gestion de qualité et la documentation technique doivent être conservés pendant une période de « *10 ans après la mise sur le marché ou la mise en service du système d'IA [...]* » en cas de demande par les autorités nationales compétentes⁵⁷⁸. De la même manière, les journaux générés automatiquement doivent être conservés pendant une période variable, d'au moins six mois, en fonction de la destination du système d'IA et des obligations légales auxquelles sont soumis les utilisateurs⁵⁷⁹.

⁵⁷⁴ *Ibid.* Annexe IV « Documentation technique ».

⁵⁷⁵ « *La Commission est habilitée à adopter des actes délégués conformément à l'article 97 pour modifier l'annexe IV, lorsque cela est nécessaire, afin de garantir que, compte tenu du progrès technique, la documentation technique fournit toutes les informations requises pour évaluer la conformité du système avec les exigences énoncées dans la présente section.* » - *Ibid.* Article 11.3.

⁵⁷⁶ *Ibid.* Annexe IV, alinéa 2.g).

⁵⁷⁷ « *Règlement sur l'intelligence artificielle* », 2024. Annexes VII et VIII. Pour plus de détails, voir Chapitre 2 : Le contrôle des preuves de conception.

⁵⁷⁸ *Ibid.* Article 18.1.

⁵⁷⁹ « *Sans préjudice du droit de l'Union ou du droit national applicable, les journaux sont conservés pendant une période adaptée à la destination du système d'IA à haut risque, d'au moins six mois, sauf disposition contraire dans le droit de l'Union ou le droit national applicable, en particulier dans le droit de l'Union sur la protection des données à caractère personnel.* » - *Ibid.* Article 19.1.

B- Le passage à une redevabilité juridique ex ante.

L'encadrement de la phase de conception des systèmes d'IA. Jusqu'à présent, le droit positif relatif aux systèmes algorithmiques contenait quasi-exclusivement des obligations à la charge des déployeurs des logiciels⁵⁸⁰. L'exploitation de tels outils n'était pas encadrée en soi et les obligations n'apparaissaient qu'après la survenance d'un problème ou d'un dommage. L'utilisation d'un système d'IA doit alors simplement produire des résultats conformes aux exigences légales imposées pour tout processus de prise de décision dans un contexte similaire. Comme le RGPD avant lui, le règlement sur l'IA change le paradigme pour passer à une logique de redevabilité *ex ante*. Les acteurs concernés sont principalement ceux qui interviennent avant la commercialisation ou la mise sur le marché du produit. Ce sont notamment les fournisseurs qui ont la charge de la majorité des obligations présentées précédemment. Ils doivent attester de la conformité de leurs systèmes d'IA par principe, avant tout préjudice subi par un utilisateur ou une personne affectée par un résultat algorithmique.

Les approches de la logique de redevabilité. La logique de redevabilité peut être utilisée de deux manières en fonction des objectifs poursuivis par le législateur : une manière proactive et une manière réactive⁵⁸¹. Dans un premier temps, la redevabilité proactive renvoie à un objectif déterminé. Le législateur souhaite montrer le niveau à atteindre en donnant une liste de critères clairs qui caractérise l'objectif souhaité. Dans le cadre du règlement sur l'IA, l'objectif est l'élaboration d'un système d'IA redevable⁵⁸². L'accent est alors mis sur la détermination de la conformité et des mécanismes de surveillance de cette conformité pendant une période donnée, le cycle de vie d'un système d'IA par exemple. Dans un second temps, la redevabilité réactive se concentre davantage sur la mise en œuvre, à savoir les mesures qui sont exécutées pour gérer un évènement particulier⁵⁸³. Dans ce cas, on se concentre davantage sur la communication et la correction, par les débiteurs des obligations, des problèmes rencontrés par l'objet de la redevabilité. À cela s'ajoute alors une importance accrue accordée à l'application de sanctions par les autorités compétentes. L'approche proactive conçoit davantage la redevabilité comme

⁵⁸⁰ Exception faite pour les fournisseurs de produits réglementés déjà visés par des obligations juridiques de conformité (les dispositifs médicaux par exemple)

⁵⁸¹ NOVELLI Claudio, TADDEO Mariarosaria et FLORIDI Luciano, « Accountability in artificial intelligence », *AI & Society*, 39, 2023.

⁵⁸² *Ibid.*

⁵⁸³ *Ibid.*

une vertu que doit posséder l'objet régulé tandis que l'approche réactive se veut plus proche du mécanisme de *reporting*⁵⁸⁴.

Malgré leurs différences, ces deux manières sont complémentaires et souvent utilisées de concert dans les textes ayant recours à une logique de redevabilité. Cette utilisation conjointe n'empêche cependant pas une des deux formes de redevabilité de prédominer sur l'autre. Dans le cas du règlement sur l'IA, c'est la redevabilité proactive qui est privilégiée par la Commission européenne. En effet, l'importance des exigences relatives au fonctionnement des systèmes d'IA et non aux acteurs ainsi que l'emphase instaurée sur le contrôle de la conformité pré-commercialisation et sa surveillance pendant son existence sur le marché confirment cela. De plus, les sanctions et les mécanismes en réaction à un manque de conformité sont moins mis en avant dans le texte⁵⁸⁵.

Les intérêts de la logique de redevabilité. Le choix par l'Union européenne⁵⁸⁶ d'une redevabilité principalement proactive est assez cohérent au regard des notions déjà abordées jusqu'alors. Dans un premier temps, l'influence des principes éthiques de l'IA sur l'élaboration de cette réglementation s'illustre, au-delà des obligations présentes, par cette idée d'une redevabilité intégrée aux systèmes d'IA comme une valeur à respecter. Dans un deuxième temps, la décision de réglementer l'IA spécifiquement permet d'orienter plus facilement l'innovation vers une direction souhaitée par le régulateur. Le choix de l'objet réglementé permet donc de faciliter l'objectif qu'une redevabilité proactive instaure. Troisièmement, le but premier du règlement sur l'IA est l'harmonisation des règles concernant l'intelligence artificielle⁵⁸⁷ vers un objectif de protection des droits fondamentaux, d'une part, et, d'autre part, afin d'éviter la fragmentation du marché européen⁵⁸⁸. Or, la stratégie de la redevabilité remplit elle aussi une fonction d'harmonisation des valeurs et des croyances au sein d'une société⁵⁸⁹. Dans le cadre de la réglementation sur les systèmes d'IA, cette approche harmonisée a vocation

⁵⁸⁴ BOVENS Mark, « Two Concepts of Accountability », *West European Politics*, 33, Routledge, 2010.

⁵⁸⁵ À titre d'exemple, l'obligation du distributeur d'informer les autorités nationales en cas de risque national produit par un système d'IA à haut risque – « Règlement sur l'intelligence artificielle », 2024. Article 27.2.

⁵⁸⁶ Cette approche proactive étant déjà présente au sein des lignes directrices produites par le HLEG IA en 2019, il est difficile d'attribuer ce choix exclusivement à la Commission européenne.

⁵⁸⁷ Le nom complet de la législation sur l'IA est : Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828.

⁵⁸⁸ L'un des objectifs spécifiques énoncé par la Commission européenne dans le préambule du règlement sur l'IA est : « faciliter le développement d'un marché unique pour des applications d'IA légales, sûres et dignes de confiance, et empêcher la fragmentation du marché. »

⁵⁸⁹ DUBNICK Melvin et FREDERICKSON H. George, « Public Accountability », National Academy of Public Administration et Kettering Foundation, 2011.

à améliorer la confiance des acteurs dans l'écosystème⁵⁹⁰. Plus la sécurité juridique proposée par le régime juridique est forte, plus la confiance des acteurs est importante. Par conséquent, l'approche davantage proactive de la redevabilité s'insère elle aussi dans l'atteinte de l'objectif de sécurité juridique prônée par la Commission européenne.

§2. La délégation des objectifs de qualité aux normes techniques

Annnonce de plan. Le règlement sur l'IA impose une certification préalable à tout système d'IA à haut risque avant son entrée sur le marché européen. Cette certification est conditionnée au respect par le système d'IA des exigences de qualité énoncées par la proposition européenne. La logique de certification met naturellement en avant l'importance des exigences de qualité attendues. Toutefois l'analyse de ce type d'exigences présentes dans le règlement sur l'IA montre un manque significatif de précision dans leur description **(A)**. Les spécifications des exigences substantielles sont alors déléguées à de futures normes techniques qui viendront en complément du texte européen **(B)**.

A- Des nouvelles exigences de qualité des systèmes d'IA imprécises.

Tout d'abord, il est essentiel de constater que les diverses obligations substantielles présentées précédemment représentent une avancée considérable dans la lutte contre les atteintes sur la sécurité, la santé et les droits fondamentaux des personnes affectées par les systèmes d'IA à haut risque. Leur présence dans la proposition européenne est aussi une preuve de l'influence significative des initiatives éthiques sur les réflexions politiques et juridiques. Équité, transparence, contrôle humain, performance, quatre enjeux de fond majeurs de l'IA se retrouvent dans le règlement sur l'IA. Et pour ce qui est des principes de vie privée et de responsabilité, le premier est déjà pris en compte par le droit positif tandis que le second est traité à part dans des propositions réglementaires spécifiques à la problématique⁵⁹¹. L'existence des nouvelles exigences imposées directement aux systèmes d'IA permet d'envisager la

⁵⁹⁰ TADDEO Mariarosaria, « Trusting Digital Technologies Correctly », *Minds & Machines*, 27, 2017.

⁵⁹¹ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022. COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux*, 28 septembre 2022.

possibilité d'une sécurité juridique nouvelle pour les personnes affectées. En effet, le respect de leurs droits fondamentaux apparaît aujourd'hui incertain face aux problèmes apportés par les technologies d'IA.

L'apparition de termes juridiques imprécis. Ensuite, la prise en compte de ces principes éthiques par le règlement entraîne alors la juridicisation de nouvelles notions. Ainsi, la notion de biais énoncé à l'article 10 du règlement sur l'IA, outil au service du principe de *fairness*, est un concept juridique original. La référence du vocabulaire juridique qu'est le *Vocabulaire juridique* du professeur de droit Gérard Cornu⁵⁹², pris dans sa 12^e édition, ne définit pas ce qu'est un biais au sens du droit. De plus, si l'on retire des recherches l'expression « par le biais de », le terme de biais n'est jamais utilisé dans l'ouvrage, que ce soit dans une signification similaire ou non de celle entendue dans le contexte des technologies d'IA. En recherchant sur l'ensemble des règlements de l'Union européenne, le terme « biais » n'est toutefois par inconnu. Déjà utilisé à plusieurs reprises dans le contexte d'utilisation de données⁵⁹³, il n'est cependant pas défini par les textes. Cruciale pour leur identification, la définition du biais n'est pas non plus énoncée par le texte de la Commission européenne. Cependant, la version finale du règlement sur l'IA ajoute désormais explicitement, parmi les pratiques appropriées sur les données de conception, un lien entre les biais et leurs conséquences⁵⁹⁴ :

« (f) un examen permettant de repérer d'éventuels biais qui sont susceptibles de porter atteinte à la santé et à la sécurité des personnes, d'avoir une incidence négative sur les droits fondamentaux ou de se traduire par une discrimination interdite par le droit de l'Union, en particulier lorsque les données de sortie influencent les entrées pour les opérations futures ; »⁵⁹⁵

Un autre exemple de ce lien explicite entre les biais et les conséquences que le législateur cherche à prévenir se trouve en droit fédéral canadien dans la proposition de loi sur l'IA⁵⁹⁶. Le texte proposé par le gouvernement utilise de la notion de « *résultat biaisé* » qu'il définit à l'aide d'un renvoi à la réglementation canadienne sur la discrimination :

« Contenu généré, prédiction ou recommandation faite ou décision prise par un système d'intelligence artificielle qui défavorisent, directement ou indirectement et sans justification, un individu sur le fondement d'un ou

⁵⁹² CORNU Gérard, *Vocabulaire juridique*, Presses Universitaires de France - PUF, 2022.

⁵⁹³ Recherche du terme « biais » effectuée à l'aide du moteur « Eur-lex » pour les règlements européens de 2022 à 2017. L'expression « par le biais » a été exclue des résultats de recherche.

⁵⁹⁴ « *examination in view of possible biases that are likely to affect health and safety of persons Or, lead to discrimination prohibited by Union law* ; » - Texte de compromis de la France, p.8

⁵⁹⁵ *Législation sur l'intelligence artificielle*, supra note 246 - Article 10.5.2.f).

⁵⁹⁶ *Loi sur l'intelligence artificielle et les données*, 2022.

*plusieurs motifs de distinction illicite prévus à l'article 3 de la Loi canadienne sur les droits de la personne, ou de leur effet combiné. [...] »*⁵⁹⁷

La disposition du texte européen et la proposition de loi canadienne tentent d'intégrer, plus ou moins directement, ce nouveau concept dans le cadre juridique existant. Sans contexte, le terme de biais peut renvoyer à une multitude de problématiques techniques. Le rapprochement avec les objectifs de protection de la sécurité, de la santé et des droits fondamentaux permet de contextualiser le sens que les législateurs souhaitent donner à ce terme. Toutefois, il faut remarquer que la création de trop nombreux liens avec d'autres notions juridiques connues limite significativement l'intérêt du mécanisme. Dans le règlement européen sur l'IA, le terme de biais est désormais rattaché à quatre objectifs différents dont la protection des droits fondamentaux qui regroupe à elle seule une multitude d'intérêts divers. À titre de comparaison, la proposition de loi canadienne sur l'IA et les données associe la problématique des biais uniquement à la notion juridique de discrimination. Malgré la pluralité des critères de discrimination, le législateur canadien propose donc une contextualisation juridique du terme de « biais » bien plus précise que celle contenue dans le règlement européen sur l'IA.

Le constat est identique pour les autres exigences posées par le règlement sur l'IA. Au-delà de l'exemple précédent sur le terme « biais », les critères de qualité des jeux de données fournis par l'article 10, à savoir que ces jeux de données doivent être « *pertinents, suffisamment représentatifs et, dans la mesure du possible, exempts d'erreurs et complets [...] »*⁵⁹⁸, ne sont pas non plus détaillés dans le texte. À titre de comparaison, la déclaration de Toronto, publiée en mai 2018, qui traitait spécifiquement des risques de discrimination algorithmique, énonçait alors ces derniers de la manière suivante :

*« Lorsqu'ils cartographient les risques, les acteurs du secteur privé doivent prendre en compte les risques communément associés aux systèmes reposant sur l'apprentissage automatique – par exemple, l'alimentation des systèmes avec des données incomplètes ou non représentatives, ou les jeux de données reflétant des préjugés historiques ou systémiques. »*⁵⁹⁹

Si la reprise des éléments produits par les textes éthiques est pertinente, l'on constate que le niveau de précision et de détails entre la déclaration et la proposition européenne sont similaires.

⁵⁹⁷ Loi sur l'intelligence artificielle et les données, 2022. Article 5(1) Définitions.

⁵⁹⁸ « Règlement sur l'intelligence artificielle », 2024. Version du compromis final. Traduction réalisée par l'auteur. Article 10.3.

⁵⁹⁹ « The Toronto Declaration : Protecting the right to equality and non-discrimination in machine learning systems ».Paragraphe 46/ Emphase ajoutée par l'auteur.

Par conséquent, ce niveau d'abstraction, déjà identifié comme une lacune applicative des textes éthiques, risque de poser la même difficulté avec le texte européen, d'autant plus qu'il s'agit d'une exigence technique de qualité du système d'IA à haut risque.

De façon semblable, l'exigence de qualité du système d'IA que représente l'article 15 concernant l'exactitude, la robustesse et la cybersécurité propose un seuil de qualité minimal sans pour autant le décrire. Ainsi, il n'est pas fait mention de concept comme « le taux de faux positif » ou « le taux de faux négatif » ou tout autre concept qui permettrait d'identifier les critères de qualités sous-jacents à cette obligation.

Au contraire des exigences sur la performance et sur les données utilisées par les systèmes d'IA, le règlement sur l'IA offre des critères plus précis dans la description des exigences de transparence et de contrôle humain. La transparence du fonctionnement d'un système d'IA à haut risque est satisfaisante dès lors qu'elle permettra :

« aux déployeurs d'interpréter les sorties d'un système et de les utiliser de manière appropriée. Un type et un niveau adéquats de transparence sont garantis afin de veiller au respect des obligations pertinentes incombant au fournisseur et au déployeur énoncées à la section 3. »⁶⁰⁰

Il est alors possible d'identifier les différents sous-critères que pourra prendre en compte le fournisseur pour mettre en conformité ses systèmes d'IA concernés. La production d'une notice d'utilisation détaillée, imposée au fournisseur de systèmes d'IA, éclaire également les informations considérées comme nécessaires pour atteindre ce niveau de qualité en termes de transparence.

B- La place des normes techniques dans le règlement européen sur l'IA

Une délégation aux normes techniques assumée. Le caractère incomplet des exigences imposées aux systèmes d'IA avant leur mise en service et/ou leur commercialisation est une démarche consciente du législateur européen. En se rattachant à la stratégie de la « nouvelle approche », il accepte de n'intégrer que les exigences essentielles à son texte, éléments qui seront ensuite complétés par les « spécifications techniques ». Cette logique est

⁶⁰⁰ « Règlement sur l'intelligence artificielle », 2024. Article 13.1.

particulièrement présente concernant les obligations substantielles relatives au fonctionnement des systèmes d'IA à haut risque.

La détermination des mesures appropriées. Dans un premier temps, c'est le choix des mesures nécessaires à la satisfaction de ces obligations de qualité qui est subordonné à l'état de la technique, symbolisé par les normes techniques. Ainsi, les fournisseurs de systèmes d'IA à haut risque doivent satisfaire les exigences relatives à leur logiciel « *en tenant compte de leur destination ainsi que de l'état de la technique généralement reconnu en matière d'IA et de technologies liées à l'IA* »⁶⁰¹. Le renvoi à l'état de l'art est ici généralisé à l'ensemble des obligations applicables aux systèmes d'IA à haut risque. L'utilisation des normes techniques comme référence de qualité doit ainsi être indiquée explicitement dans le système de gestion de qualité, permettant aux autorités de contrôler facilement les résultats avec le référentiel utilisé⁶⁰².

La définition des nouvelles notions imprécise. Dans un second temps, c'est la définition même des exigences qui est associée à l'aspect technique. Les quatre critères de qualité des données que sont la pertinence, la représentativité, l'absence d'erreurs et la complétude sont explicitement mises en lien avec des référentiels techniques. Les jeux de données d'entraînement, de validation et de tests « *possèdent les propriétés statistiques appropriées* »⁶⁰³. Cette phrase renforce le besoin de transformation d'exigences littérales en indicateurs mathématiques. Une tendance confirmée par l'article 9 du règlement sur l'IA qui dispose que les tests utilisés pour identifier et évaluer les risques présents dans chaque système d'IA à haut risque doivent être effectués « *sur la base d'indicateurs et de seuils probabilistes préalablement définis, qui sont adaptés à la destination du système d'IA à haut risque.* »⁶⁰⁴. En d'autres termes, le besoin de mesures et le choix de ces mesures seront déterminés au regard d'indicateurs mathématiques. Ces indicateurs viendraient alors préciser dans la pratique les exigences littérales encore trop floues concernant la qualité imposée aux systèmes d'IA à haut risque. Une autre information significative de ce texte est l'existence de « *seuils probabilistes* » qui viennent faire écho aux termes « appropriés » ou « suffisant » inscrits dans les dispositions du règlement sur l'IA. La Commission européenne propose ici une complémentarité entre la norme juridique et la norme technique pour parvenir à une conformité des systèmes d'IA. Par

⁶⁰¹ *Ibid.* Article 8.1

⁶⁰² *Ibid.* Article 17.1(e)

⁶⁰³ *Ibid.* Article 10.3.

⁶⁰⁴ *Ibid.* Article 9.8.

conséquent, les besoins de forme de l'encadrement de l'IA, à savoir la précision des termes et leur articulation, sont des problématiques confiées aux normes techniques complémentaires, en particulier les normes européennes harmonisées.

CONCLUSION DE LA SECTION II :

La phase de conception des systèmes d'IA à haut risque fait l'objet d'exigences à la fois sur la qualité de fonctionnement dudit logiciel (les exigences substantielles), sur les procédures et stratégies à mettre en place (les exigences processuelles) et sur la documentation liée à la réalisation de ces exigences (les exigences documentaires). Par cet ensemble d'obligations, le législateur européen aborde l'ensemble des besoins de fond de l'encadrement de l'IA mis en avant par les textes éthiques et le droit positif ; à savoir la transparence, l'équité, le contrôle humain, la responsabilité, la fiabilité technique et la protection de la vie privée. Toutefois, le choix traditionnel par l'Union européenne de la stratégie de la nouvelle approche limite le niveau de détails des dispositions. Les spécifications techniques des exigences essentielles sont alors déléguées à de futures normes techniques complémentaires, en particulier les normes européennes harmonisées. Les besoins de précision et d'articulation ne sont donc pas traités par le règlement européen. Dès lors, ces problématiques sont désormais à la charge de nouveaux acteurs que sont les organismes de normalisation dont les principaux sont les organismes de normalisation européens. Outre l'importance de ces besoins de fond pour l'encadrement de l'IA, ils portent ici sur l'ensemble des exigences substantielles relatives au fonctionnement des systèmes d'IA à haut risque. Par conséquent, les normes techniques mentionnées par le texte européen détiennent autant, voire davantage, d'importance dans l'établissement de ce nouveau cadre européen sur l'IA.

CONCLUSION DU CHAPITRE 2 :

Avec le règlement sur l'IA, l'Union européenne surmonte l'obstacle de fournir une définition de l'IA et délimite un cadre juridique spécifique aux systèmes d'IA. Le législateur a alors concentré ses exigences principalement sur les systèmes d'IA à haut risque qui représentent la catégorie la plus critique de systèmes d'IA dont l'utilisation est autorisée sur le marché européen. En effet, les systèmes d'IA à haut risque sont la seule catégorie de systèmes d'IA dont la conception fait l'objet d'exigences relatives à la qualité de leur fonctionnement. De ce fait, ils ont été considérés comme le point critique du règlement européen et le choix a donc été fait de leur appliquer notre grille d'analyse. Parmi les exigences relatives aux systèmes d'IA à haut risque, les obligations imposées aux fournisseurs de ces logiciels quant à leur fonctionnement sont au centre de l'attention. Si elles peuvent également poursuivre des objectifs propres, les exigences processuelles et documentaires du règlement servent principalement à atteindre, et à prouver, la qualité de fonctionnement imposée par l'Union européenne. Par ces exigences substantielles, le législateur européen reprend tous les principaux besoins de fond de l'encadrement de l'IA qui ont été mis en lumière par les textes éthiques et le droit positif. Cependant, les exigences substantielles portant sur la conception des systèmes d'IA à haut risque sont constituées de nouvelles notions juridiques que le règlement ne définit pas. De plus, le législateur européen a recours à des notions floues dans les obligations imposées aux fournisseurs. Usuelles en droit, leur manque de précision risque toutefois d'entraîner des difficultés importantes pour le fournisseur du système d'IA qui aura la charge de les mettre en application. L'interprétation de ces différentes notions est un élément crucial pour l'articulation du règlement sur l'IA avec le droit positif mais également pour l'articulation de ses exigences entre elles. Face au manque de précision de ces exigences, le législateur européen a choisi de confier cette tâche aux organisations de normalisation lors de l'élaboration des spécifications techniques.

CONCLUSION DU TITRE 1 :

Le règlement européen sur l'IA est un projet ambitieux de l'Union européenne visant à encadrer de façon transversale la conception de l'ensemble des systèmes d'IA destinés à être commercialisés ou utilisés sur le marché européen. Ce texte a notamment pour objectif de résoudre les nombreuses difficultés que rencontre le droit positif lorsqu'il s'applique aux technologies d'IA ainsi que les problématiques de l'IA mises en lumière par les textes éthiques. Parmi les différentes catégories de niveaux de risque élaborées dans le règlement européen, les obligations les plus importantes à respecter sont celles relatives aux systèmes d'IA à haut risque. Non-discrimination algorithmique, protection des données utilisées, fiabilité, transparence, responsabilité et contrôle humain, l'Union européenne parvient à réunir l'ensemble des principaux principes éthiques de l'IA dans les exigences imposées aux fournisseurs de systèmes d'IA à haut risque. Néanmoins, notre analyse révèle que le règlement ne répond pas aux besoins de forme que sont la précision et l'articulation des exigences. En traitant le système d'IA comme un simple produit, le législateur a refusé de détailler précisément ces exigences ainsi que d'aborder les problèmes d'articulation entre elles et avec le droit positif. Par conséquent, les fournisseurs de systèmes d'IA à haut risques, parce qu'ils sont débiteurs de ces obligations, devront déterminer eux-mêmes les détails de ces obligations et les preuves adéquates pour attester de leur respect. Cependant, cette tâche ne repose pas exclusivement sur ces derniers. En effet, la Commission européenne a délégué de façon directe l'élaboration des spécifications des exigences aux organismes de normalisation européens, et de façon plus indirecte aux acteurs chargés du contrôle de ces exigences.

TITRE II : LA PLACE DE LA PREUVE DANS LES EXIGENCES DE CONCEPTION D'UN SYSTEME D'IA

Introduction. Au sens du règlement européen 1025/2012 relatif à la normalisation européenne, une spécification technique est « *un document qui prescrit les exigences techniques à respecter par un produit [...]* »⁶⁰⁵, notamment « *les caractéristiques requises d'un produit, dont les niveaux de qualité, de performance, d'interopérabilité, de protection de l'environnement, de santé, de sécurité [...]* »⁶⁰⁶. Parmi la multitude d'éléments pouvant être contenus dans ces normes techniques, notre analyse se concentrera sur les preuves, à savoir les éléments utilisés afin d'emporter la conviction du destinataire quant à l'exactitude d'une situation juridique⁶⁰⁷. En l'espèce, il s'agit donc des éléments permettant d'attester de la conformité du fonctionnement du système d'IA à haut risque aux exigences substantielles de conception énoncées dans le règlement européen sur l'IA. Ces preuves particulières seront désignées par la formulation de « preuves de conception » pour la suite de la thèse. Comme cela sera démontré ci-après, les preuves de conception sont au cœur des garanties de qualité que souhaite imposer le législateur européen aux systèmes d'IA à haut risque. Dès lors, il est indispensable de s'interroger sur les acteurs en charge de la détermination⁶⁰⁸ de ces preuves, à savoir le choix des éléments qui seront considérés comme pertinents pour attester de la conformité au règlement sur l'IA. En effet, les travaux sur l'éthique de l'IA ont révélé qu'un même principe peut aboutir à des exigences très différentes selon les personnes chargées de son interprétation.

Annonce de plan. La détermination des preuves de conception pertinentes au regard des exigences du règlement sur l'IA est le résultat d'un processus complexe. Tout d'abord, ces preuves sont réfléchies et élaborées, au moins en partie, au sein des spécifications techniques commandées par la Commission européenne aux organismes européens de normalisation (**Chapitre 1**). Cependant, l'application de ces potentielles propositions de preuves de conception par les fournisseurs de systèmes d'IA à haut risque devront ensuite être validées lors

⁶⁰⁵ « Règlement (UE) n ° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n ° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012. Article 4.

⁶⁰⁶ *Ibid.* Article 4.a).

⁶⁰⁷ Voir III.La preuve juridique comme outil d'analyse du droit des systèmes d'IA, P. 20.

⁶⁰⁸ « Préciser, fixer, circonscrire ce qui est incertain. » - « DÉTERMINER : Définition de DÉTERMINER », sur *Cnrtl.fr*.

des différents contrôles prévus par le règlement européen. Les acteurs en charge de ces évaluations contribueront alors eux aussi à déterminer les preuves de conception considérées comme pertinentes (**Chapitre 2**).

CHAPITRE 1 : L'ELABORATION DES PREUVES DE CONCEPTION

Introduction. Le 5 décembre 2022, la Commission européenne notifiait⁶⁰⁹ les organisations européennes de normalisation d'une requête pour l'élaboration de normes techniques relatives à certaines exigences contenues dans le règlement européen sur l'IA. La majeure partie des normes techniques commandées concernent les exigences substantielles relatives à la qualité de fonctionnement pour les systèmes d'IA à haut risque⁶¹⁰. Ces obligations représentent le cœur de la réglementation européenne sur l'IA⁶¹¹. En effet, les obligations imposées aux différents acteurs sont toutes liées d'une manière ou d'une autre aux exigences de qualité du chapitre 2 titre III du règlement sur l'IA. Par exemple, le fournisseur du système d'IA doit mettre en place diverses mesures dans le but de satisfaire lesdites exigences⁶¹² tandis que l'importateur doit s'assurer que le fournisseur a correctement évalué la conformité de son système⁶¹³. Le distributeur ne peut se limiter à vérifier la présence du marquage CE sur le produit. Il doit aussi être vigilant sur les indices de non-conformité effective du produit⁶¹⁴. Les obligations processuelles énoncées dans le règlement⁶¹⁵ ont donc pour fondement partiel de permettre de vérifier formellement que le fournisseur a produit un système conforme aux attentes européennes. En d'autres termes, le système de certification des systèmes d'IA à haut risque mis en place par l'Union européenne n'a de sens qu'en ce qu'il permet de rendre compte d'une qualité européenne des systèmes d'IA mis sur le marché. Or, ce sont les normes techniques qui détermineront préciseront le contenu de cette qualité indispensable à l'obtention du marquage CE.

⁶⁰⁹ « La Commission établit un système de notification pour toutes les parties prenantes, y compris les organisations européennes de normalisation et les organisations des parties prenantes européennes recevant un financement de l'Union conformément au présent règlement, afin de garantir une consultation adéquate et l'adéquation au marché avant: [...] b) d'adopter les demandes de normalisation visées à l'article 10 ; » - « Règlement (UE) n° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012. Article 12.

⁶¹⁰ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*. Annexe I.

⁶¹¹ VEALE Michael et ZUIDERVEEN BORGESIOUS Frederik, « Demystifying the Draft EU Artificial Intelligence Act », *Computer Law Review International*, 22, 2021.

⁶¹² « Règlement sur l'intelligence artificielle », 2024. Article 16 du règlement européen sur l'IA

⁶¹³ *Ibid.* Article 26.

⁶¹⁴ *Ibid.*

⁶¹⁵ En particulier l'obligation d'établir une documentation technique énoncée aux articles 11 et 18 du règlement européen sur l'IA. L'obligation de générer des journaux de façon automatique des articles 12 et 20 garantit la traçabilité des actions prises par le système d'IA afin de vérifier qu'il fonctionne conformément à la qualité européenne.

Annonce de plan. Au regard de la technicité informatique des systèmes d'IA, l'élaboration d'exigences relatives à la qualité de leur fonctionnement nécessite l'existence de règles particulièrement précises. Les organisations européennes de normalisation sont ainsi chargées de produire les spécifications techniques des exigences essentielles contenues dans le règlement sur l'IA. Sous la forme de normes techniques, ces spécifications déclinent de façon plus détaillée les obligations juridiques auxquelles elles sont associées. Au cours de cette tâche, les organisations de normalisation doivent tenter de conserver la signification de chaque exigence malgré le recours à des termes plus précis informatiquement. Au sein des normes techniques, l'on peut retrouver les objectifs précis de fonctionnement que le système d'IA doit atteindre pour être certifié, de même que les modalités exactes de contrôle de ces objectifs. Ces tests techniques et les résultats qui en découlent sont autant de preuves permettant d'attester de la conformité du logiciel. Ces preuves occupent une place fondamentale dans le processus de certification européen car elles garantissent in fine un fonctionnement qualitatif des systèmes d'IA (**Section I**). La reconnaissance de l'importance de ces preuves dans la conformité juridique des systèmes d'IA à haut risque implique de s'intéresser au processus d'élaboration de ces preuves juridiques (**Section II**).

Section I : L'importance des preuves au sein des normes techniques

Annonce de plan. Parmi les nombreuses informations et spécifications que contiennent les normes techniques, notre analyse se concentrera sur les éléments techniques de preuves aux exigences substantielles des systèmes d'IA à haut risque. Ces éléments techniques de preuves, aussi appelés des preuves techniques ci-après, désignent ici les éléments constitutifs des tests (procédures de tests, résultats obtenus, etc.) réalisés sur les systèmes d'IA afin d'attester de leur respect des exigences de fonctionnement contenus dans le règlement européen sur l'IA. Ce choix fort se justifie par le rôle important qu'elles occupent dans le système de certification européen des systèmes d'IA. Dans un premier temps, cette importance est reconnue par le législateur européen lui-même qui accorde directement et indirectement une certaine place à la question des preuves techniques dans son règlement (**Sous-section 1**). Dans un second temps, l'étude des preuves techniques des systèmes d'IA révèle que, au-delà de leur complexité technique, ces preuves contiennent des enjeux juridiques de taille. De la précision des notions à leur articulation entre elles, les preuves techniques de la conformité des systèmes d'IA

viennent répondre aux besoins non satisfaits par le règlement européen sur l'IA (**Sous-section 2**).

Sous-section 1 : La place accordée aux preuves techniques dans le règlement

Annonce de plan. Au sein du règlement européen sur l'IA, l'intérêt accordé par l'Union européenne concernant l'importance des preuves techniques se manifeste directement au travers des dispositions à leur sujet (**§1**). De façon plus indirecte, cette importance peut aussi être constatée par le biais des contreparties que le législateur accorde en échange de preuves techniques satisfaisantes (**§2**).

§1. La reconnaissance partielle de l'importance de preuves techniques adéquates

Une prise de conscience du législateur européen. Dans la première version du règlement européen sur l'IA proposé par la Commission européenne, le sujet des preuves techniques, abordé au travers des notions de « métriques » et de « seuils probabilistes », n'était mentionné qu'à deux reprises⁶¹⁶. Tout d'abord, l'article 9.7 énonçait que « *les tests sont effectués sur la base de métriques et de seuils probabilistes préalablement définis, qui sont adaptés à la destination du système d'IA à haut risque.* »⁶¹⁷. Ensuite, l'article 15.2 réutilisait la notion de métrique appliquée cette fois-ci à l'évaluation de l'exactitude des systèmes d'IA à haut risque : « *Les niveaux d'exactitude et les métriques pertinents en matière d'exactitude des systèmes d'IA à haut risque sont indiqués dans la notice d'utilisation jointe.* »⁶¹⁸. Le caractère adapté ou pertinent des preuves techniques est mentionné dans chacune de ces deux dispositions mais le législateur ne s'y attarde pas.

⁶¹⁶ Dans cette première version de la proposition de règlement, les preuves techniques n'apparaissent pas dans les considérants du texte.

⁶¹⁷ « Règlement sur l'intelligence artificielle », 2024. Version de la Commission européenne.

⁶¹⁸ *Ibid.* Version de la Commission européenne.

A la lecture de la version finale du règlement européen, il apparaît plusieurs nouveaux⁶¹⁹ paragraphes dédiés au sujet de l'adéquation des preuves techniques. Tout d'abord, la question est abordée dès les considérants :

*« Dans ce contexte [l'objectif de garantie de l'exactitude des mesures], en coopération avec les parties prenantes et organisations concernées, telles que les autorités de métrologie et d'étalonnage des performances, la Commission devrait encourager, le cas échéant, l'élaboration de critères de référence et de méthodes de mesure pour les systèmes d'IA. Ce faisant, la Commission devrait prendre note des partenaires internationaux travaillant sur la métrologie et les indicateurs de mesure pertinents relatifs à l'IA et collaborer avec ceux-ci. »*⁶²⁰

Cet intérêt pour la pertinence des preuves techniques utilisées lors des tests sur les systèmes d'IA à haut risque est renouvelé à l'article 15.2 :

*« Pour examiner les aspects techniques de la manière de mesurer les niveaux appropriés d'exactitude et de robustesse visés au paragraphe 1 et tout autre indicateur de performance pertinent, la Commission, en coopération avec les parties prenantes et organisations concernées, telles que les autorités de métrologie et d'étalonnage des performances, encourage, le cas échéant, l'élaboration de critères de référence et de méthodes de mesure. »*⁶²¹

Ces deux paragraphes mettent en lumière une prise de conscience de l'importance des preuves techniques dans le règlement européen sur l'IA. D'une part, le texte impose à la Commission européenne de s'intéresser précisément au sujet. La présence de cette disposition se pose alors légèrement à contre-courant de la logique derrière la stratégie de la « nouvelle approche ». Celle-ci sépare, d'un côté, les exigences essentielles présentes dans le règlement et, de l'autre côté, les spécifications techniques contenues dans les normes techniques européennes. La volonté d'impliquer la Commission européenne, un acteur intervenant principalement pour déterminer les exigences essentielles du texte, sur des sujets aussi techniques que « *les niveaux appropriés d'exactitude* » ou le choix de « *l'indicateur de performance pertinent* » n'est donc pas anodin. Ainsi, il nous semble que ce paragraphe illustre la reconnaissance par l'Union européenne d'une certaine limite à séparer hermétiquement droit et technique dans l'encadrement de l'IA. D'autre part, l'article enjoint la Commission européen à s'entourer

⁶¹⁹ Les paragraphes sont considérés comme nouveaux par rapport à la version initiale. Les dispositions mentionnées ici ont été élaborées pour la première fois au sein de la version du règlement proposé par le Parlement européen.

⁶²⁰ « Règlement sur l'intelligence artificielle », 2024. Considérant 74.

⁶²¹ *Ibid.*

d'autorités spécialisées dans la métrologie⁶²² et l'étalonnage⁶²³. Cette précision est bienvenue car elle reconnaît, à juste titre, que le choix de ces éléments est une tâche extrêmement complexe, nécessitant une expertise spécifique.

Une prise de conscience aux conséquences réduites. L'apport de ces paragraphes se révèle toutefois assez restreint. Par sa nature juridique, le considérant 74 du règlement européen ne sert qu'à éclairer l'interprétation du corps du texte et n'a aucune force contraignante. En ce qui concerne l'article 15.1.a), il est limité aux indicateurs relatifs à la performance des systèmes d'IA à haut risque. Les biais discriminatoires, la transparence et autres enjeux de fond du droit de l'IA ne sont pas soumis à des dispositions similaires alors qu'ils seront eux aussi soumis à des tests de conformité, dont les résultats serviront d'éléments techniques de preuve pour attester du respect des exigences réglementaires. Le faible rôle accordé à la Commission européenne est également très critiquable. En effet, cette-dernière doit seulement « *encourager, le cas échéant, l'élaboration de critères de référence et de méthodes de mesure pour les systèmes d'IA.* »⁶²⁴. En comparaison, l'amendement du Parlement européen, dont cette disposition est inspirée, imposait à la Commission européenne la production de documents non contraignants dans le but d'éclairer les acteurs dans leur choix des indicateurs (métriques, seuils probabilistes, etc.) adéquats⁶²⁵. La place accrue de la problématique des preuves techniques dans le texte ne présage donc pas un changement radical par rapport à la version initiale. Par conséquent, si l'Union européenne accorde désormais plus de place aux preuves techniques dans son règlement sur l'IA, les conséquences concrètes de cette évolution devraient être relativement faibles.

§2. La contrepartie juridique à l'obtention des preuves techniques

La déduction de l'importance des preuves par leur contrepartie juridique. Le règlement européen sur l'IA est un texte juridique spécifique aux systèmes d'IA qui vient se superposer à

⁶²² La métrologie désigne la science des mesures. « MÉTROLOGIE : Définition de MÉTROLOGIE » LAROUSSE Éditions, « Définitions ».

⁶²³ « *Vérification de la graduation et du réglage d'un appareil de mesure par comparaison avec l'étalon* » - « ÉTALONNAGE : Définition de ÉTALONNAGE », sur Cnrtl.fr.

⁶²⁴ « Règlement sur l'intelligence artificielle », 2024.

⁶²⁵ « *To address the technical aspects of how to measure the appropriate levels of accuracy and robustness set out in paragraph 1 of this Article, the AI Office shall bring together national and international metrology and benchmarking authorities and provide non-binding guidance on the matter as set out in Article 56, paragraph 2, point (a).* » - Ibid. Version du Parlement européen. Emphase ajoutée par l'auteur.

toutes les branches du droit applicables à ces technologies. Dès lors, le nouveau texte européen peut à la fois poursuivre ses propres objectifs et venir combler les lacunes que certaines dispositions de droit positif rencontrent lors de leur application aux technologies d'IA. Cependant, les objectifs du droit positif et ceux du règlement européen sur l'IA peuvent ne pas toujours être convergents. Ainsi, la satisfaction d'une exigence du règlement peut imposer de porter atteinte à une exigence de droit positif. De la même manière, le respect de principes existants en droit positif peut venir limiter la capacité des fournisseurs d'IA à remplir leurs obligations. Par conséquent, il est parfois nécessaire d'opérer une conciliation entre différents intérêts. L'atteinte portée aux droits et libertés doit être proportionnée à l'objectif poursuivi. Pour déterminer l'importance des preuves de la conformité juridique des systèmes d'IA, la première méthode est d'observer les phénomènes de mise en balance impliquant les preuves de conception. Plus l'atteinte aux droits concédée est importante, plus l'objectif satisfait en contrepartie peut être considéré comme important. En d'autres termes, le prix à payer est acceptable aux yeux de législateur. Dans notre contexte, cet objectif est l'obtention des éléments de preuves permettant d'attester de la qualité acceptable de fonctionnement d'un système d'IA à haut risque.

L'exemple de la lutte contre les biais algorithmiques. Ce compromis s'illustre particulièrement dans les obligations relatives aux biais algorithmiques. Dans le règlement européen sur l'IA, la lutte contre les biais algorithmiques se divise en deux étapes : leur détection et leur correction⁶²⁶. Pour rappel, les biais algorithmiques discutés au sein du droit européen des systèmes d'IA sont les biais « *susceptibles de porter atteinte à la santé et à la sécurité des personnes, d'avoir une incidence négative sur les droits fondamentaux ou de se traduire par une discrimination interdite par le droit de l'Union* »⁶²⁷. Si la sécurité et la santé des personnes physiques apparaissent comme une évidence car ce sont deux des trois objectifs poursuivis globalement par le règlement, la discrimination est une précision importante. Les biais recherchés par le texte sont en particulier des biais discriminatoires, c'est-à-dire la prise en compte, directe ou indirecte, de critères de discrimination énoncés par les textes de lois dans la prise de décision. Celle-ci étant *in fine* défavorable pour une ou plusieurs sous-catégories de population. Pour surveiller et détecter le potentiel impact d'un ou plusieurs critères discriminants sur le raisonnement d'un système d'IA, il faut tout d'abord savoir si les données de conception utilisées proviennent de personnes concernées qui possèdent ces critères. En

⁶²⁶ « Dans la mesure où cela est strictement nécessaire aux fins de la détection et de la correction des biais en ce qui concerne les systèmes d'IA à haut risque [...] » Ibid. Article 10.5

⁶²⁷ Ibid. Article 10.2(f).

d'autres termes, pour savoir si un système d'IA est défavorable envers les femmes, il faut déjà savoir si les individus composant le jeu de données et les individus affectés par le résultat sont ou non des femmes. Il est donc indispensable que les concepteurs en charge de la lutte contre les biais algorithmiques aient accès à la variable « sexe ». D'un point de vue informatique, la communauté scientifique s'accorde aussi sur l'importance pour les concepteurs de systèmes d'IA d'avoir connaissance des « variables sensibles » liées aux critères de discrimination indiqués par la loi pour pouvoir lutter contre ces biais⁶²⁸.

Toutefois, bon nombre de critères de discrimination correspondent à des catégories particulières de données à caractère personnel au sens du règlement (UE) 2016/679 relatif à la protection des données (ci-après RGPD)⁶²⁹. L'article 9 du règlement définit les données, aussi appelées « données sensibles », contenues dans cette catégorie :

*« Le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits. »*⁶³⁰

Les données sensibles sont donc toutes les données liées à ces onze critères. Or, la totalité de ces critères peut être rattachée aux critères de discrimination contenus en droit européen⁶³¹. Le RGPD impose l'interdiction de principe du traitement de ce type de données, un principe qui connaît dix exceptions⁶³². Néanmoins, seules deux des dix exceptions sont envisageables

⁶²⁸ ŽLIOBAITĖ Indrė et CUSTERS Bart, « Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models », *Artif Intell Law*, 24, 2016.

⁶²⁹ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

⁶³⁰ *Ibid.* Article 9.1

⁶³¹ « Directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en œuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique », 180, 2000 « Directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail », 303, 2000 « Directive 2004/113/CE du Conseil du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans l'accès ` des biens et services et la fourniture de biens et services », 373, 2004 « Directive 2006/54/CE du Parlement européen et du Conseil du 5 juillet 2006 relative à la mise en œuvre du principe de l'égalité des chances et de l'égalité de traitement entre hommes et femmes en matière d'emploi et de travail (refonte) », 204, 2006.

⁶³² « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016, *supra* note 55, article 9.2.a) à j)

lorsque les données sensibles sont traitées dans le cadre d'une décision individuelle automatisée au sens de l'article 22 du RGPD. Dans cette situation, les deux exceptions maintenues par le texte européen⁶³³ sont le consentement de la personne⁶³⁴ concernée et la présence de motifs d'un intérêt public important⁶³⁵. Toutefois, il n'est pas certain que la détection, la surveillance ou la correction de biais algorithmiques satisfassent systématiquement les conditions pour être qualifiée de décision individuelle automatisée⁶³⁶. Pour cette raison, les dix exceptions seront conservées dans la présente analyse et non les deux exceptions spécifiques à l'article 22 du RGPD. De plus, les données sensibles au sens du RGPD font également l'objet d'un régime spécifique au sein de la directive 2016/680⁶³⁷ (ci-après directive police-justice) ainsi que le règlement 2018/1725 concernant le traitement de données personnelles par les institutions, organes et organismes de l'Union⁶³⁸. Si ce dernier texte conserve les dix exceptions mentionnées dans l'article 9 du RGPD, la directive police-justice renverse le principe d'interdiction pour un principe d'utilisation limité à trois situations :

« a) lorsqu'ils sont autorisés par le droit de l'Union ou le droit d'un État membre;

b) pour protéger les intérêts vitaux de la personne concernée ou d'une autre personne physique ; ou

⁶³³ Ibid. Article 22.4.

⁶³⁴ Ibid. « la personne concernée a donné son consentement explicite au traitement de ces données à caractère personnel pour une ou plusieurs finalités spécifiques » - Article 9.2.a)

⁶³⁵ Ibid. « le traitement est nécessaire pour des motifs d'intérêt public important, sur la base du droit de l'Union ou du droit d'un État membre qui doit être proportionné à l'objectif poursuivi, respecter l'essence du droit à la protection des données et prévoir des mesures appropriées et spécifiques pour la sauvegarde des droits fondamentaux et des intérêts de la personne concernée; » - Article 9.2.g)

⁶³⁶ Pour plus d'informations sur l'article 22 du RGPD, voir §1. Les mécanismes juridiques de la transparence ainsi que PONS Ronan, *Le droit et la discrimination algorithmique*, 2018. EDWARDS Lilian et VEALE Michael, « Enslaving the Algorithm », 2018. WACHTER Sandra, MITTELSTADT Brent et FLORIDI Luciano, « Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation », *International Data Privacy Law*, 1, 2017.

⁶³⁷ « Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil », OJ L, 2016.

⁶³⁸ « Règlement (UE) 2018/1725 du Parlement européen et du Conseil du 23 octobre 2018 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union et à la libre circulation de ces données, et abrogeant le règlement (CE) n° 45/2001 et la décision n° 1247/2002/CE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2018.

*c) lorsque le traitement porte sur des données manifestement rendues publiques par la personne concernée. »*⁶³⁹

Pour permettre la lutte contre les biais algorithmiques, la proposition européenne de règlement sur l'IA doit donc s'articuler ou déroger à ces trois textes. En l'espèce c'est la première option qui a été privilégiée par le législateur européen :

*« Dans la mesure où cela est strictement nécessaire aux fins de la surveillance, de la détection et de la correction des biais en ce qui concerne les systèmes d'IA à haut risque, les fournisseurs de ces systèmes peuvent traiter des catégories particulières de données à caractère personnel visées à l'article 9, paragraphe 1, du règlement (UE) 2016/679, à l'article 10 de la directive (UE) 2016/680 et à l'article 10, paragraphe 1, du règlement (UE) 2018/1725, sous réserve de garanties appropriées pour les droits et libertés fondamentaux des personnes physiques, y compris des limitations techniques relatives à la réutilisation ainsi que l'utilisation des mesures les plus avancées en matière de sécurité et de protection de la vie privée, telles que la pseudonymisation, ou le cryptage lorsque l'anonymisation peut avoir une incidence significative sur l'objectif poursuivi. »*⁶⁴⁰

Au préalable, il faut garder à l'esprit que l'articulation entre le règlement européen sur l'IA et le droit européen sur la protection des données personnelles est encore floue. S'il semble que la lutte contre les biais algorithmiques se rapproche de l'exception pour « motif d'intérêt public important », ce n'est pour l'instant qu'une hypothèse attendant d'être confirmée explicitement par le législateur européen. Actuellement, le règlement européen sur l'IA n'est pas encore suffisamment claire pour affirmer qu'il s'agira de la base légale à venir pour un tel traitement de ces catégories particulières de données personnelles⁶⁴¹. De façon similaire, les « *garanties appropriées pour les droits et libertés fondamentaux des personnes physiques* » sont encore vagues et l'établissement de leur nature et du degré de protection exigée n'a pas été confié aux organisations européennes de normalisation⁶⁴². Il est donc possible que ce travail revienne plutôt aux autorités de contrôle nationales. À l'image du Comité européen de la protection des

⁶³⁹ Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection Or, prosecution of criminal offences Or, the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, *supra* note 487. Article 10.c)

⁶⁴⁰ *Législation sur l'intelligence artificielle*, *supra* note 1, article 10.5

⁶⁴¹ « the current provisions do not seem sufficiently clear to create a legal basis for the processing of special categories of data » - « Avis conjoint 05/2021 de l'EDPB et du CEPD sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) », 2021.

⁶⁴² Il ne s'agit pas d'une des dix exigences listées dans la requête de normalisation formulée par la Commission européenne :

données, le règlement européen sur l'IA met en place un Comité européen de l'IA dont la mission est de garantir la coopération des autorités nationales et l'application cohérente des dispositions⁶⁴³. Bien que le traitement ait lieu, ces mesures de sauvegarde ont pour finalité de minimiser les risques encourus par les personnes concernées. Toutefois, ces mesures devront s'articuler correctement avec celles déjà exigées en droit des données à caractère personnel. La tâche est d'autant plus complexe que ces mesures varient selon les législations nationales, en particulier concernant la directive 2016/680⁶⁴⁴. Le coût juridique de l'obtention des preuves techniques de conception ne pourra être précisément évalué qu'après l'entrée en vigueur du règlement et, très probablement, la publication de lignes directrices de la part des autorités compétentes. Toutefois, il est certain que la possibilité de traiter des données sensibles par les fournisseurs de systèmes d'IA à haut risque est une contrepartie significative à l'obligation de minimiser les biais algorithmiques dans les jeux de données de conception. Par conséquent, l'étendue du risque auquel cette disposition expose les personnes concernées permet d'affirmer que l'obtention des preuves du niveau acceptable de biais est un élément important dans le règlement européen.

Sous-section 2 : La place concrète des preuves techniques au sein du règlement

Annonce de plan. Au sein du règlement sur l'IA, le législateur européen insiste sur l'importance de s'appuyer « *l'état de la technique généralement reconnu* »⁶⁴⁵ pour le choix des mesures de gestions des risques liés aux systèmes d'IA. En particulier, le texte impose le recours

⁶⁴³ « Le Comité fournit des conseils et une assistance à la Commission afin : a) de contribuer à la coopération efficace des autorités de contrôle nationales et de la Commission en ce qui concerne les matières relevant du présent règlement ; b) de coordonner les orientations et analyses de la Commission et des autorités de contrôle nationales et d'autres autorités compétentes sur les questions émergentes dans l'ensemble du marché intérieur en ce qui concerne les matières relevant du présent règlement, et de contribuer à ces orientations et analyses ; c) d'aider les autorités de contrôle nationales et la Commission à assurer une application cohérente du présent règlement. » - Article 56 du règlement européen sur l'IA.

⁶⁴⁴ « Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil », OJ L, 2016.

⁶⁴⁵ « Règlement sur l'intelligence artificielle », 2024. Article 8.1.

à des objets techniques que sont les « *indicateurs* »⁶⁴⁶ et les « *seuils probabilistes* »⁶⁴⁷ afin d'évaluer et attester du respect, par les systèmes d'IA, des exigences substantielles du règlement. L'utilisation de ces deux outils dans le règlement européen sur l'IA leur confère la qualité d'objet juridique. Comme indiqué par le règlement européen, le processus de test est composé de deux éléments majeurs avec, d'un côté, une métrique, qui représente l'objet de la preuve (§1), et de l'autre un seuil statistique, qui définit le niveau quantitatif d'exigence que l'on souhaite atteindre (§2).

§1. La détermination de l'objet de la preuve : l'indicateur

Annonce de plan. L'indicateur est un « *dispositif fournissant des repères et servant à mesurer* »⁶⁴⁸. Le recours à cet outil de quantification n'a rien de surprenant dans le règlement européen sur l'IA. En effet, la logique de redevabilité choisie par le législateur européen s'articule particulièrement bien avec la quantification dont la première des quatre fonctions normatives est de rendre compte⁶⁴⁹. Le choix de l'indicateur pour attester de la conformité du fonctionnement d'un système d'IA au règlement européen est donc un choix important réalisé par l'Union européenne (**A**). La présomption de conformité attachée aux normes européennes harmonisées confère aux indicateurs qu'elles contiennent une valeur légale, ils deviendront une représentation quantitative d'une ou plusieurs exigences essentielles. Dès lors, l'indicateur choisi doit être cohérent avec l'exigence à laquelle il est associé (**B**).

A- Le choix de l'indicateur

La notion d'indicateur. Le terme de « indicateur » est employé à plusieurs reprises dans le règlement européen sur l'IA. Ces indicateurs doivent être utilisés dans les tests de qualité lors de la conception⁶⁵⁰ mais également communiquées par le fournisseur du système d'IA aux

⁶⁴⁶ *Ibid.* Article 9.7. Le terme « indicateur » a remplacé celui de « métrique » choisi par la Commission européenne dans la version initiale du règlement sur l'IA.

⁶⁴⁷ *Ibid.* Article 9.7.

⁶⁴⁸ « INDICATEUR : Définition de INDICATEUR », sur *Cnrtl.fr*.

⁶⁴⁹ Alain Supiot, *La Gouvernance par les nombres*, Nantes : Paris, Fayard, 2015, p. 91 et suivantes. C'est d'ailleurs la comptabilité, en anglais *accountability*, qui a donné son nom à la logique juridique éponyme.

⁶⁵⁰ « Règlement sur l'intelligence artificielle », 2024. Version compromise tchèque du 03/11/22, articles 9.7 et 15.2.

déployeurs⁶⁵¹ du produit. L'importance de ces outils est confirmée dans l'annexe IV qui liste les éléments à inclure dans la documentation technique imposée par l'article 11 dudit règlement européen :

« 2. Une description détaillée des éléments du système d'IA et de son processus de développement, y compris:

[...]

*(g) les procédures de validation et d'essai utilisées, y compris les informations sur les données de validation et d'essai utilisées et leurs principales caractéristiques; les indicateurs utilisés pour mesurer l'exactitude, la robustesse et le respect des autres exigences pertinentes énoncées au chapitre III, section 2, ainsi que les éventuelles incidences discriminatoires [...] »*⁶⁵²

En ce qui concerne les exigences de qualité énoncées dans le chapitre 2 du titre III du règlement sur l'IA⁶⁵³, les indicateurs sont des dispositifs d'évaluation et de contrôle du système d'IA concerné⁶⁵⁴. Il symbolise alors deux enjeux : que mesure-t-on exactement et comment le mesure-t-on ? Elaborer un indicateur revient à répondre à ces deux questions. Au préalable, il est important de signaler que le recours à un indicateur de qualité a d'ores et déjà des conséquences sur la nature de l'objectif poursuivi : il doit être mesurable. En effet, quel que soit l'objectif d'origine, il est indispensable de trouver un moyen de le quantifier. C'est une condition d'utilisation *sine qua non* des indicateurs.

Exemples d'éléments mesurables dans un système d'IA. Dans un système algorithmique, de nombreux éléments peuvent être pris en compte par les indicateurs. À titre d'illustration, prenons le cas d'un système d'IA destiné à accorder des prêts bancaires à des demandeurs. Ce logiciel a pour finalité d'évaluer précisément le risque de défaut de chaque demandeur grâce à une analyse de leur dossier. Un autre élément à prendre en compte peut être le taux de faux positifs. Les faux positifs sont les cas où le système s'est prononcé favorablement à l'accord d'un prêt, c'est le résultat positif, mais les emprunteurs ont finalement fait défaut au remboursement de leur prêt, c'est l'erreur transformant le résultat positif en « faux positif ». Il

⁶⁵¹ *Ibid.* Considérant 49 (« *The level of accuracy should and accuracy metrics should be communicated to the users* »), articles 13.3.(b)(ii), 15.2.

⁶⁵² *Ibid.* Annexe IV.2.g).

⁶⁵³ Voir chapitre 2 section 2.

⁶⁵⁴ Prendre les normes techniques et les indicateurs au sérieux, Benoit FRYDMAN, chapitre 1 dans « Gouverner par les normes techniques et les indicateurs. De Hume aux rankings / Frydman Benoît et van Waeyenbergen Arnaud (dir.) », en ligne: *Droit & Société* <<https://ds.hypotheses.org/959>>.

est également possible d'observer le taux de faux négatif. À l'inverse de l'élément précédent, les faux positifs sont les situations où le système s'oppose à l'octroi d'un prêt, le résultat négatif, alors que les nouveaux emprunteurs arriveront parfaitement à rembourser la banque. Encore un autre élément peut être la précision globale du système, c'est-à-dire le nombre de fois où le système a évalué correctement la situation, par exemple le nombre de fois où le système a correctement prédit le risque de défaut de remboursement des dossiers évalués, que ce soit des décisions d'octroi ou de refus du prêt.

Le choix contextuel de la métrique. Les indicateurs exigés, mais non choisis ni conseillés, par le législateur européen représentent donc un enjeu très important dans le processus de certification des systèmes d'IA à haut risque. En tant qu'objets techniques, le choix du ou des indicateurs est particulièrement complexe car il doit être cohérent avec une obligation juridique. Cette tâche est particulièrement difficile lorsque l'exigence de référence n'est pas particulièrement précise. De façon générale, la priorité dans le choix des indicateurs est leur « *pertinence et son efficacité par rapport aux objectifs pour lesquels ils seront utilisés* »⁶⁵⁵. Dans le contexte du règlement sur l'IA, l'adéquation entre les indicateurs et les exigences doit tenir compte du secteur d'utilisation du système d'IA. Au-delà de la distinction entre les systèmes d'IA présentant un risque élevé et les autres systèmes d'IA, la finalité a un impact sur les critères de sélection de l'indicateur. À ce titre, le règlement européen sur l'IA indique que le fournisseur du système d'IA doit préciser dans ses instructions d'utilisation « *le cas échéant, sa performance [du système d'IA] en ce qui concerne des personnes ou groupes de personnes spécifiques à l'égard desquels le système est destiné à être utilisé;* »⁶⁵⁶. Le terme « *approprié* » renvoie à cet aspect de la pertinence liée à l'utilisation finale du système d'IA. Par exemple, la vision de l'exactitude pour un système d'IA d'analyse de radiographie médicale peut raisonnablement mettre la priorité sur un taux faible de faux négatifs plutôt que sur le taux de faux positifs. En effet, il est préférable qu'un individu soit à tort considéré comme nécessitant un examen médical plus complet (faux positif) plutôt qu'il soit renvoyé chez lui à tort et que son cas s'aggrave (faux négatif). En revanche, un système d'IA responsable de l'octroi de prêt bancaire peut accorder une importance différente à ces éléments, la banque préférant refuser des prêts par erreur que de s'exposer au risque de prêter de l'argent à un individu qui ne pourra pas la rembourser *in fine*. La priorisation d'un faible taux de faux positifs ou d'un faible taux

⁶⁵⁵ « Chapitre 10 - Indicateurs de conformité avec les instruments internationaux sur les droits de l'homme », *Revue de l'OCDE sur le développement*, 10, 2009.

⁶⁵⁶ « Règlement sur l'intelligence artificielle », 2024. Article 13.3.b)v)

de faux négatifs aura des conséquences très différentes selon le contexte d'utilisation du système d'IA.

La préexistence d'un indicateur vis-à-vis de l'exigence légale. L'utilisation d'un indicateur pour évaluer le niveau de respect à une exigence nécessite de quantifier ladite exigence. Parce qu'elles sont formulées en langage naturel, toutes les exigences sont par principe difficiles à traduire sous la forme d'indicateur. En dépit de cette problématique intrinsèque, toutes les obligations relatives au fonctionnement des systèmes d'IA, contenues dans le règlement européen, ne présentent pas la même difficulté à être associées à un indicateur d'évaluation. Par exemple, le choix d'un indicateur pour évaluer l'exactitude d'un système d'IA⁶⁵⁷ sera beaucoup plus aisé que pour tester sa transparence⁶⁵⁸ ou la présence de biais discriminatoires⁶⁵⁹. Cette différence s'explique par la familiarité du secteur de l'IA avec chaque exigence contenue dans le règlement européen sur l'IA. Par exemple, l'exactitude d'un système d'IA est une caractéristique cruciale aussi bien du point de vue juridique qu'informatique. En revanche, la détection de biais discriminatoires ou la transparence du logiciel est une préoccupation bien plus importante dans le domaine du droit qu'en informatique. Par conséquent, le travail de réflexion sur les métriques d'exactitude est plus avancé que celui sur les métriques de transparence, d'explicabilité ou de biais discriminatoires. Dès lors, il est plus probable que le fournisseur d'un système d'IA possède davantage de connaissances et de choix pour sélectionner la métrique pertinente afin d'évaluer la précision de son logiciel que pour d'autres caractéristiques. En d'autres termes, l'habitude des concepteurs de systèmes d'IA vis-à-vis d'une contrainte du règlement européen sur l'IA accroît leur facilité à choisir une métrique d'évaluation. Cette aisance est en partie justifiée par la préexistence de la contrainte du point de vue informatique. Les exigences importantes seulement du point de vue juridique auront plus de risques d'être évaluées de façon moins pertinente. En effet, les lacunes du concepteur et/ou la faible diversité des métriques existantes complexifiera la sélection d'un indicateur approprié au regard du contexte d'utilisation du système d'IA.

⁶⁵⁷ *Ibid.* Article 15.

⁶⁵⁸ *Ibid.* Article 13.

⁶⁵⁹ *Ibid.* Article 10.5.

B-Le choix d'un indicateur adéquat à l'exigence légale

Le choix des indicateurs de qualité des systèmes d'IA à haut risque sera très probablement une tâche de sélection des métriques existantes les plus adéquates par rapport aux objectifs juridiques à représenter. Toutefois, la reprise d'un indicateur existant n'exonère pas les différents acteurs (organismes de normalisation, fournisseur du système d'IA, autorités de contrôle, etc.) du travail de traduction de l'exigence juridique en langage mathématique. La légitimité des indicateurs couramment utilisés dans la communauté scientifique facilite la sélection par les acteurs mais n'est en rien une garantie de satisfaction de l'exigence juridique. Ainsi le travail de sélection des indicateurs par les acteurs, en premier lieu par les organismes de normalisation, s'expose à deux obstacles d'adéquation.

L'obstacle disciplinaire. Premièrement, il existe un obstacle disciplinaire. Les indicateurs existants ont été conçus principalement pour servir un but scientifique et non juridique. L'objectif scientifique pour lequel l'indicateur a été conçu peut apparaître similaire sur la forme à un concept juridique mais s'avérer en réalité fort éloigné sur le fond de celui-ci. L'explicabilité est un exemple représentatif de cet obstacle. Tout d'abord, la définition de l'explicabilité ne fait déjà pas l'objet d'un consensus au sein même de la communauté scientifique. Ensuite, le droit ne fournit pas non plus de définition de la notion d'explicabilité⁶⁶⁰. Par conséquent, parvenir à élaborer une définition commune aux deux disciplines représente un défi colossal, voire impossible à l'heure actuelle. Dès lors, il n'est pas étonnant que les métriques développées par la communauté scientifique ne puissent pas être réutilisées aisément pour les besoins juridiques.

Une particularité de cet obstacle disciplinaire tient à son caractère inévitable. En d'autres termes, il est nécessaire qu'au moins un individu procède à ce travail de traduction, qu'il soit réalisé globalement au sein des normes européennes harmonisées ou spécifiquement par les débiteurs des obligations relatives à la qualité de fonctionnement des systèmes d'IA à haut risque. La force contraignante des textes juridiques impose le respect des obligations de qualité de fonctionnement du système d'IA. La démonstration de cette conformité passe par la réalisation de tests et l'obtention de résultats acceptables, basés sur des métriques qui devront nécessairement être sélectionnées et définies. La cohérence entre les différentes significations nécessite un dialogue interdisciplinaire important. S'il se démocratise progressivement dans le

⁶⁶⁰ Les termes « explicable », « explicabilité » et « explication » sont présents neuf fois dans le règlement sur l'IA, sans jamais être associés à une définition. Parmi les exigences contenues dans le texte européen, l'explicabilité peut être associée à la fois à l'obligation de transparence (prévue à l'article 13) et au droit à l'explication de décisions individuelles (prévu à l'article 86). *Ibid.*

secteur de la recherche, cet échange doit être garanti dans les différentes étapes de sélection des indicateurs.

L'obstacle culturel. L'appropriation par le droit des métriques scientifiques rencontre un second obstacle : l'obstacle culturel. Outre l'obstacle disciplinaire, les indicateurs sont développés par une communauté internationale. De la même manière que la conception d'un jeu de données ou d'un algorithme renseigne sur la culture (et les biais) de son concepteur, l'élaboration d'un indicateur, d'un standard et plus généralement d'une norme technique symbolise une vision de la société⁶⁶¹. Les divergences culturelles possibles sont encore plus importantes depuis que les normes techniques se sont immiscées dans des thématiques sociales⁶⁶². Ainsi lors des discussions relatives à l'élaboration de la norme ISO 26000 sur la responsabilité sociale des entreprises (ci-après RSE)⁶⁶³. De plus, les cadres réglementaires possèdent également, entre eux, des différences de vision et de logique qui viennent renforcer l'obstacle culturel. Or, une majorité d'indicateurs existants dans la littérature scientifique et les travaux de rapprochement entre ces indicateurs et le droit sont d'origine américaine, et non européenne⁶⁶⁴. L'adéquation entre des indicateurs empreints de la culture, voire de la réglementation, nord-américaine et les exigences du règlement européen sur l'IA est donc un travail complexe nécessitant une grande prudence.

L'exemple des indicateurs de biais algorithmiques. Pour illustrer ces phénomènes autour de l'adoption des indicateurs scientifiques par le droit, il est proposé d'approfondir le cas des métriques dédiées à l'exigence de biais discriminatoires. La question des biais est une sous-problématique de l'équité algorithmique qui concerne le fonctionnement des systèmes d'IA⁶⁶⁵.

⁶⁶¹ *supra* note 7. « Chapitre 3 : Les normes ISO, CEN et celles issues de consortium privés : bric à brac ou système pour l'Union européenne ? » - Arnaud Van Waeyenberge, pp. 108-109.

⁶⁶² « Avec le passage d'objet de la normalisation des produits vers les services, la normalisation s'est de plus en plus portée sur le terrain social » « Chapitre 3 : Les normes ISO, CEN et celles issues de consortium privés : bric à brac ou système pour l'Union européenne ? » - Arnaud Van Waeyenberge, p. 108

⁶⁶³ Jean-Christophe Graz, « Quand les normes font loi : Topologie intégrée et processus différenciés de la normalisation internationale » (2004), pp. 252 et suivantes

⁶⁶⁴ WACHTER Sandra, MITTELSTADT Brent et RUSSELL Chris, « Why Fairness Cannot Be Automated », *Computer Law & Security Review* 41, 41, 2021.

⁶⁶⁵ L'équité algorithmique ne peut pas et ne doit pas être réduite à la seule lutte contre les biais algorithmiques discriminatoires. Cette réduction de la problématique s'explique ici par le périmètre de notre étude, qui est délimité aux exigences relatives au fonctionnement des systèmes d'IA.

Les nombreux travaux et cas réels sur ce sujet offrent des connaissances suffisantes pour réaliser une illustration détaillée de l'incidence du choix d'indicateurs⁶⁶⁶.

La première difficulté rencontrée par le choix d'un indicateur d'évaluation des biais est la compréhension de la notion de « biais ». Le biais est défini communément comme « *une déformation, un travers* »⁶⁶⁷. Il s'agit d'un dysfonctionnement du schéma de pensée, du raisonnement, qui influence nos choix ou notre vision de la réalité⁶⁶⁸. Il se différencie de l'aléa en ce qu'il n'est ni une déviation du raisonnement due au seul hasard ni du bruit⁶⁶⁹. Bien que très souvent associé au principe d'équité, un biais peut se révéler purement technique et porter seulement sur la performance d'un outil, sans aucun préjudice pour un ou plusieurs individus⁶⁷⁰. En l'absence de précisions sémantiques, un biais peut donc renvoyer à de nombreuses significations techniques. Ainsi, le règlement européen sur l'IA associe désormais explicitement⁶⁷¹ la notion de biais aux risques pour la santé, la sécurité et plus particulièrement la discrimination au sens du droit de l'Union européenne⁶⁷². Ce lien avec le droit européen de la non-discrimination est aussi visible plus indirectement dans le contenu de la notice d'utilisation que les fournisseurs de système d'IA à haut risque doivent constituer :

« La notice d'utilisation contient au moins les informations suivantes :

[...]

(b) les caractéristiques, les capacités et les limites de performance du système d'IA à haut risque, notamment:

⁶⁶⁶ Pour n'en citer qu'une infirme partie des travaux de cette communauté : O'NEIL Cathy, *Weapons of Math Destruction*, Crown, 2017. ZUIDERVEEN BORGESIU Frederik, « Discrimination, Artificial Intelligence and Algorithmic Decision-Making », Conseil de l'Europe, 2018. CRAWFORD Kate, « The Hidden Biases in Big Data », *Harvard Business Review*, 1 avril 2013. DÉFENSEUR DES DROITS et COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Algorithmes : prévenir l'automatisation des discriminations », Défenseur des Droits, 2020. SÁNCHEZ-MONEDERO Javier, DENCİK Lina et EDWARDS Lilian, What does it mean to « solve » the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2020.

⁶⁶⁷ Définition du biais : <https://www.cnrtl.fr/definition/biais>

⁶⁶⁸ Ronan Pons, *Le droit et la discrimination algorithmique*, 2018

⁶⁶⁹ Ronan Pons & Laurent Risser, « Preuve juridique et systèmes d'intelligence artificielle » dans C. Castets-Renard & J. Eynard (dir.), *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général ; Perspectives comparées* (Bruylant Edition, 2023). P. 752. Le bruit désigne une variabilité statistique de la mesure effectuée.

⁶⁷⁰ Un système d'IA destiné à analyser des ondes pourrait tout à fait être biaisé de façon accorder plus d'importance à certaines longueurs d'onde qu'à d'autres. Cela n'aurait aucun impact sur une personne physique ou morale mais seulement sur la précision de l'appareil.

⁶⁷¹ La version du règlement européen révélée le 21 avril 2021 par la Commission européenne ne faisait pas le lien entre les biais et les risques pour la sécurité, la santé et le droit à la non-discrimination. Il s'agit d'un ajout proposé dans le texte de compromis par la présidence tchèque le 3 novembre 2022.

⁶⁷² « Règlement sur l'intelligence artificielle », 2024. Version texte de compromis tchèque, article 10.2(f)

[...]

(iv) le cas échéant, sa performance en ce qui concerne des personnes ou groupes de personnes spécifiques à l'égard desquels le système est destiné à être utilisé ; »⁶⁷³

Le texte prévoit ainsi la possibilité de distinguer les performances obtenues par le système d'IA lors de la phase de test en fonction du groupe d'individus sur lequel les résultats sont produits. La problématique des biais est donc abordée à la fois dans les jeux de données de conception et dans les résultats produits par le système d'IA. Ces dispositions ont pour objectif de lutter contre les biais discriminatoires, c'est-à-dire les biais qui peuvent entraîner un résultat algorithmique biaisé, puis mener potentiellement à une décision discriminatoire pour la personne affectée⁶⁷⁴.

D'un point de vue technique, la problématique des biais informatiques est un sujet connu depuis de nombreuses années⁶⁷⁵ et plusieurs indicateurs ont été proposées pour mesurer ce type de dysfonctionnement⁶⁷⁶. Leur objectif est de mesurer la *fairness*, que l'on traduit ici par le concept « d'équité », mais qui ne peut pas être considérée comme un simple synonyme de non-discrimination⁶⁷⁷. Élaborées par les communautés mathématiques et informatiques, les métriques portent une signification spécifique de l'équité qui n'est pas nécessairement proche de la notion juridique de non-discrimination.

Au sens juridique du terme, la discrimination désigne la différenciation de traitement considérée comme illégitime⁶⁷⁸. La notion de discrimination peut alors être interprétée de façon restrictive ou extensive. Dans sa conception restrictive, la discrimination renvoie au traitement défavorable dont sont victimes certaines personnes particulièrement vulnérables en raison de leur appartenance à un groupe défini par un attribut particulier⁶⁷⁹ (la religion, l'orientation sexuelle, l'origine ethnique, le handicap, etc.). Cette approche restrictive est notamment appliquée en droit français par le code du travail⁶⁸⁰ et le code pénal⁶⁸¹, et en droit européen par

⁶⁷³ *Ibid.* Article 13.

⁶⁷⁴ La production d'un résultat biaisé sur la base de critères discriminants n'entraîne pas nécessairement la réalisation d'une décision discriminatoire. Pour plus d'informations sur cette distinction, voir Sous-section 2 : Les interactions entre le fonctionnement du système d'IA à haut risque et les droits et libertés fondamentaux.

⁶⁷⁵ Voir notamment : BATYA Friedman et HELEN Nissenbaum, *Bias in computer systems, Proceedings of the 1996 Conference on Fairness, Accountability, and Transparency*, 1996.

⁶⁷⁶ WACHTER Sandra, MITTELSTADT Brent et RUSSELL Chris, « Bias Preservation in Machine Learning », *West Virginia Law Review*, 123, 2021. Dans cet article, les auteurs recensent vingt métriques de *fairness*.

⁶⁷⁷ Sur le concept de *fairness*, sa traduction en « équité » et la distinction avec le concept juridique de non-discrimination, voir A- Le principe de *fairness*.

⁶⁷⁸ LOCHAK Danièle, « La notion de discrimination », *Confluences Méditerranée*, 48, 2004. P. 15.

⁶⁷⁹ *Ibid.* P. 16.

⁶⁸⁰ Article L. 1132-1 du code du travail.

⁶⁸¹ Article 225-1 du code pénal.

le Traité d'Amsterdam⁶⁸². Interprétée extensivement, la discrimination est entendue comme tout atteinte arbitraire et injustifiée au principe d'égalité⁶⁸³. Cette conception extensive est présente dans la jurisprudence du Conseil d'État, du Conseil constitutionnel et de la Cour européenne des droits de l'Homme⁶⁸⁴. L'existence de deux conceptions différentes d'une même notion juridique est une difficulté supplémentaire dans la recherche d'une métrique adéquate pour la représenter.

De manière générale, cet obstacle disciplinaire est abordé par des équipes pluridisciplinaires de chercheurs qui tentent de faire le lien entre cette notion scientifique de *fairness* et le concept juridique de non-discrimination⁶⁸⁵. Pour rappel, la notion de « discrimination » pris dans un contexte informatique n'a que peu d'intérêt au regard de la tâche à accomplir. En effet, la définition informatique de la discrimination est identique à son sens originel, à savoir une simple différence de traitement, sans aucune connotation négative. Pris dans ce sens, la très grande majorité des systèmes d'IA discrimine. Par conséquent, cette signification informatique n'a que peu d'intérêt dans notre cas. Le terme informatique d'équité est celui qui se rapproche le plus de la notion de non-discrimination en droit. Cependant, l'équité fait l'objet d'une grande diversité de conceptions et chacune d'entre elles est formalisée au travers d'un indicateur. Chacun des indicateurs mesure donc un objectif différent, une conception propre de l'équité. Par conséquent, l'instrument probatoire, à savoir l'indicateur, représente donc une interprétation de l'objet de la preuve, ici le biais algorithmique discriminatoire.

Pour illustrer ce phénomène, il n'est pas nécessaire de lister la totalité des métriques d'équité existantes. Il sera traité en l'espèce des trois métriques les plus usuelles dans le domaine⁶⁸⁶ : la parité statistique, l'égalité des opportunités et la balance des opportunités. Ces indicateurs sont de nature statistique et sont donc formulés sous la forme d'équations⁶⁸⁷.

⁶⁸² Traité d'Amsterdam modifiant le traité sur l'Union européenne, les traités instituant les Communautés européennes et certains actes connexes (97/C340/01). *Journal officiel* n° C 340 du 10/11/1997. Article 13.

⁶⁸³ LOCHAK Danièle, « La notion de discrimination », *Confluences Meditteranee*, 48, 2004. P. 16.

⁶⁸⁴ *Ibid.*

⁶⁸⁵ BAROCAS Solon et SELBST Andrew D., « Big Data's Disparate Impact », *California Law Review*, 104, 2016. HACKER Philipp, « Teaching Fairness to Artificial Intelligence », *55 Common Market Law Review* 1143-1186 (2018), 2018. CASTETS-RENARD Céline et BESSE Philippe, « Ex ante Accountability of the AI Act », 2022.

⁶⁸⁶ Alexandra Chouldechova, « Fair prediction with disparate impact: A study of bias in recidivism prediction instruments » (2016) arXiv:161007524 [cs, stat], en ligne: <<http://arxiv.org/abs/1610.07524>>, , arXiv: 1610.07524.

⁶⁸⁷ Pour une présentation des métriques en question sous leur forme mathématique, voir notamment *Ibid.*

En premier lieu, la parité statistique définit l'équité comme l'égalité des chances pour deux sous-populations d'obtenir une décision favorable. Dans le contexte d'un système d'IA de recrutement, si 80 candidats et 20 candidates se présentent pour 10 postes ouverts, le système d'IA devra sélectionner (ou recommander) 8 hommes et 2 femmes pour que la parité statistique soit respectée (10% de décisions positives pour chacune des deux sous-catégories de population). En second lieu, l'indicateur de l'égalité des opportunités définit l'équité comme l'égalité d'exactitude du système entre deux sous-populations. En d'autres termes, cela signifie que le système d'IA se trompe autant dans ses résultats pour un groupe d'individus que pour l'autre⁶⁸⁸. L'objectif recherché est une qualité de fonctionnement du logiciel égale entre les groupes d'individus, et non plus une égalité de résultats comme précédemment. En troisième et dernier lieu, la balance des taux d'erreurs est un approfondissement de la métrique d'égalité des opportunités. Cette métrique propose une conception de l'équité représentée par l'égalité entre les taux de faux positifs et l'égalité de faux négatifs entre les deux groupes d'individus. Plus clairement, cela signifie que le système d'IA se trompe défavorablement (un faux négatif) et favorablement (faux positif), autant pour une catégorie de personnes que pour une autre. À la différence de l'égalité des chances, la balance des taux d'erreurs prend également en compte les conséquences de l'erreur et permet de percevoir les conséquences des taux d'erreurs pour chaque catégorie comparée. Un candidat recruté alors qu'il ne le « méritait »⁶⁸⁹ pas est une erreur bénéfique pour l'individu alors que la situation où une candidate qui « méritait » d'être recrutée ne l'est pas est aussi une erreur du logiciel, mais celle-ci est défavorable pour l'individu. Si cet approfondissement peut sembler technique et subtil, il est d'importance cruciale dans la conception de l'équité que souhaite voir implémenter l'Union européenne dans les systèmes d'IA qui entreront dans son marché. À titre d'exemple, c'est sur cette différence de métrique que portait le débat concernant la tristement célèbre affaire COMPAS⁶⁹⁰. D'un côté, l'entreprise NorthPoint se défendait de toute attaque en discrimination car son logiciel possédait un taux d'erreur similaire entre les personnes affectées noires et les personnes affectées blanches. De l'autre côté, ProPublica défendait que ces erreurs étaient bien plus souvent défavorables aux personnes noires qu'aux personnes blanches et inversement⁶⁹¹. Le débat autour de l'outil qu'est l'indicateur est indissociable de celui autour de l'objectif à

⁶⁸⁸ Il est considéré dans cette métrique qu'un expert humain connaît la « vérité objective », aussi appelée « vérité terrain », sans quoi il n'est pas possible de savoir que le système d'IA en question a rendu un résultat erroné.

⁶⁸⁹ L'emploi de guillemets se justifie par les nombreux aspects du concept de mérite qui peuvent faire l'objet de débats, autant dans la société civile que dans la littérature scientifique.

⁶⁹⁰ JULIA ANGWIN Jeff Larson, « Machine Bias », sur *ProPublica*, ProPublica, publié le 23 mai 2016.

⁶⁹¹ *Ibid.* MATTU Jeff Larson Julia Angwin, Lauren Kirchner, Surya, « How We Analyzed the COMPAS Recidivism Algorithm », sur *ProPublica*.

mesurer. À nouveau, cela illustre que la métrique lie de façon intime l'instrument probatoire avec l'objet de la preuve.

Le dilemme des indicateurs d'équité. À première vue, la pluralité des indicateurs existants apparaît comme une opportunité dans la recherche d'un indicateur pour représenter les exigences de biais algorithmiques. Tout d'abord, la probabilité de trouver une version de l'équité compatible avec le principe de non-discrimination est plus probable. Ensuite, s'il n'existe aucun indicateur unique correspondant parfaitement à l'exigence de biais discriminatoires, il est tentant pour le législateur ou les organisations européennes de normalisation d'utiliser plusieurs métriques simultanément de façon complémentaire pour couvrir exhaustivement l'objectif poursuivi. Cependant, il s'avère que la différence entre les indicateurs existants dissimule également une incompatibilité entre eux, plus particulièrement entre la parité statistique et l'égalité ou la balance des opportunités⁶⁹². En effet, ces métriques appartiennent à deux catégories bien distinctes d'indicateurs statistiques énoncés par Wachter et al. en 2021⁶⁹³ : les indicateurs modifiant les biais, d'une part, et les indicateurs préservant les biais, d'autre part. La parité statistique appartient à la première catégorie : les biais présents au départ (les biais de société par exemple) seront remplacés par des biais insérés directement par les concepteurs de systèmes d'IA. Pour garantir une égalité de décisions favorables entre deux catégories d'individus, l'une favorisée et l'autre défavorisée socialement, il faut accepter de biaiser le système en faveur du groupe socialement discriminé dans le but d'essayer d'obtenir une égalité de résultats. L'exactitude du système d'IA passe alors au second plan, la priorité étant donnée à l'équité des résultats. De l'autre côté, l'égalité des chances et la balance des opportunités appartiennent à la catégorie des indicateurs préservant les biais existants dans la société. L'exactitude du système d'IA devient alors la priorité. Dans le cadre d'un recrutement, seuls les meilleurs profils sont théoriquement acceptés, nonobstant le fait que tous les candidats n'ont pas été traité sur un pied d'égalité pendant leur parcours respectif.

L'existence de ces deux catégories incompatibles entraîne deux conséquences. Premièrement, il n'est pas possible de supprimer totalement les biais dans les données de conception et/ou dans le raisonnement de l'algorithme. Le choix d'un indicateur comme objectif à atteindre implique donc de sacrifier le score que le système obtiendra lorsqu'il sera

⁶⁹² KLEINBERG Jon, MULLAINATHAN Sendhil et RAGHAVAN Manish, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, [s. n.], 2017.

⁶⁹³ WACHTER Sandra, MITTELSTADT Brent et RUSSELL Chris, « Bias Preservation in Machine Learning », *West Virginia Law Review*, 123, 2021.

évalué au regard d'un autre indicateur incompatible avec le premier. Dans un contexte d'utilisation où des inégalités sont présentes, il faut donc choisir entre conserver les biais de la société, présents dans les jeux de données de conception, ou biaiser le fonctionnement du système d'IA, au détriment de son exactitude. Deuxièmement, et cela découle du premier point, la sélection de la métrique dans les normes techniques et dans les normes européennes harmonisées impose la réalisation d'un compromis sur les biais devant être reproduit par le système d'IA⁶⁹⁴, et par conséquent de la vision des biais la plus respectueuse du principe de non-discrimination.

Du point de vue européen, ce travail essentiel souffre également de l'obstacle culturel. L'exigence de *fairness* dans le fonctionnement du système d'IA est mise en relation avec les notions de « *disparate impact* » et « *disparate treatment* », mais ces dernières sont tirées de la réglementation états-unienne relative à la non-discrimination⁶⁹⁵. Ainsi, les études outre-Atlantique indiquent que la parité statistique semble être la métrique la plus adéquate pour représenter les notions juridiques conformément à leur droit à la non-discrimination. Cependant, rien ne permet de conclure qu'il en va de même en droit de l'UE. Les chercheurs en droit européen sont encore peu nombreux à travailler sur la mise en relation entre les indicateurs mathématiques et les exigences du droit européen. La tâche est d'autant plus difficile que le principe de non-discrimination européen est flou et ne bénéficie pas d'un formalisme aussi arrêté qu'aux États-Unis⁶⁹⁶. À ce jour, le principe de non-discrimination en droit européen a été étudié en relation avec les métriques d'équité au travers de la jurisprudence de la Cour de justice de l'Union européenne et de celle de la Cour européenne des droits de l'Homme⁶⁹⁷. Ladite étude conclut que l'indicateur le plus pertinent dans ce contexte est la parité statistique conditionnelle, une version de la parité statistique prenant en compte plusieurs variables simultanément. Cette métrique permet ainsi de former et de comparer l'équité entre deux groupes d'individus identifiés par plusieurs attributs et par une caractéristique discriminante unique. Par exemple, la parité statistique conditionnelle permet de comparer le

⁶⁹⁴ À ce titre, de récentes recherches tentent de proposer des algorithmes qui peuvent s'adapter aux différentes familles de métriques selon le contexte : ZEHLIKE Meike, LOOSLEY Alex, JONSSON Håkan *et al.*, « Beyond Incompatibility », arXiv, 2022.

⁶⁹⁵ Titre VII de la loi états-unienne sur la non-discrimination → chercher le nom de la loi !!! Sur le travail de mise en relation entre les métriques statistiques et les concepts juridiques en question, voir BAROCAS Solon et SELBST Andrew D., « Big Data's Disparate Impact », *California Law Review*, 104, 2016 NACHBAR Thomas, « Algorithmic Fairness, Algorithmic Discrimination », *Florida State Law Review*, 509, 2021.

⁶⁹⁶ La règle des 4/5^{ème} utilisé en droit fédéral étatsunien possède une forme quantitative plus facile à associer avec des métriques mathématiques. BAROCAS Solon et SELBST Andrew D., *op. cit.*

⁶⁹⁷ WACHTER Sandra, MITTELSTADT Brent et RUSSELL Chris, « Bias Preservation in Machine Learning », *West Virginia Law Review*, 123, 2021.

pourcentage de candidates et de candidats sélectionnés par le système d'IA (taux de décisions favorables selon la caractéristique discriminante unique) à un niveau de diplômes et d'expériences égales (attributs autres que celui du sexe).

Ces premières réflexions devront enfin être accompagnées par des analyses au niveau de chaque législation nationale qui peuvent elles-aussi énoncer une conception un peu différente du principe de non-discrimination. De la même manière que cette étude, les organismes de normalisation doivent appréhender parfaitement le contexte d'utilisation et le contexte géographique pour lequel ils élaborent un indicateur pour garantir un fonctionnement suffisamment non-discriminant des systèmes d'IA à haut risque. L'indicateur sélectionné doit, pour chaque situation, dénaturer le moins possible la conception du principe de non-discrimination sous-jacent que l'on souhaite faire prévaloir.

§2. Le niveau de qualité acceptable : le seuil statistique

Annnonce du plan. L'expression de « *seuils probabilistes* »⁶⁹⁸ employée dans le règlement européen sur l'IA ne renvoie à aucune définition juridique connue. Il s'agit d'un terme informatique qui désigne un « *niveau à ne pas dépasser* »⁶⁹⁹. Il apparaît que la fonction d'un seuil est de matérialiser une limite quantitative entre deux espaces, avec, d'un côté, un espace désiré et de l'autre un espace à éviter. Par conséquent, le recours aux seuils statistiques permet de mettre leur fonction de délimitation quantitative au service des finalités du règlement européen sur l'IA, en particulier concernant les exigences de fonctionnement des systèmes d'IA à haut risque **(A)**. Par cette quantification, les seuils probabilistes facilitent la comparaison entre des exigences originellement qualitatives et révèlent le besoin d'articulation entre les exigences juridiques du texte européen **(B)**.

A- Le rôle du seuil statistique dans le cadre juridique des systèmes d'IA

Annnonce de plan. Malgré leur nature quantitative, les seuils statistiques sont déjà utilisés par les acteurs de l'IA pour poursuivre des finalités juridiques. Le recours à ces outils est

⁶⁹⁸ « Règlement sur l'intelligence artificielle », 2024. Article 9.7.

⁶⁹⁹ <https://www.cnrtl.fr/lexicographie/seuil>. Si la valeur à obtenir doit être inférieur au seuil, il s'agit d'un seuil maximal. À l'inverse, si la valeur doit nécessairement se situer au-dessus du seuil, il s'agit d'un seuil minimal.

généralement une initiative des fournisseurs et des déployeurs des systèmes d'IA (1). Cependant, dans le cadre du règlement sur l'IA, le législateur européen choisit explicitement de confier aux seuils statistiques une fonction juridique dans sa stratégie de mise en conformité des systèmes d'IA (2).

1) Le seuil statistique en droit

La fonction du seuil statistique. Par nature, une métrique est un élément quantitatif, c'est-à-dire qu'elle prend la forme d'une variable numérique. Dans le cas des métriques utilisées pour l'évaluation de la qualité des systèmes d'IA, ces variables sont dites « continues ». À l'inverse d'une variable dite « discrète⁷⁰⁰ », une variable continue « *accepte toutes les valeurs d'un intervalle fini ou infini*⁷⁰¹ » comme les variables représentant le poids ou la taille d'une personne. En science, une variable continue peut-être perçue comme un outil parmi d'autres pour étudier le monde. Toutefois, le processus de quantification aboutit généralement à la création de catégories. Ce phénomène est encore plus probable lorsque la quantification poursuit des finalités juridiques comme dans le cas des exigences de qualité des systèmes d'IA. La mesure devient un score et ce score est étiqueté en catégories. L'appartenance d'une mesure à une catégorie permet de la qualifier. Dans un premier temps, l'établissement des seuils et des catégories associées peut se fonder sur la réglementation. Par exemple, le code de la route fixe des seuils quantitatifs sur la vitesse maximale à laquelle peut rouler un conducteur d'un véhicule à moteur. Ainsi sur l'autoroute, 130 km/h est le seuil de légalité au-dessus duquel un conducteur commet une contravention (voire un délit). Le seuil permet de distinguer entre une vitesse de conduite réglementaire et une vitesse de conduite excessive. Par conséquent, le seuil permet de qualifier un conducteur comme étant respectueux de la loi ou comme un contrevenant. Dans un second temps, le seuil peut ne pas être fondé sur la loi. Par exemple, l'indice de masse corporelle (IMC) est une mesure de la corpulence d'un individu utilisé par l'Organisation mondiale de la santé (OMS)⁷⁰² et dans la loi française⁷⁰³. Sur la base d'études statistiques, l'OMS a établi cinq seuils et donc six catégories de corpulence : famine (inférieur à 16,5), maigre (entre 16,5 et

⁷⁰⁰ Une variable discrète est « *composée d'éléments discontinus, séparés, distincts* » : <https://www.cnrtl.fr/definition/discr%C3%A8te>. Par exemple, la couleur des yeux : marron, bleue, verte, grise, etc.

⁷⁰¹ <https://www.soft-concept.com/surveymag/definition-fr/definition-variable-continue.html>

⁷⁰² « Obésité et surpoids ».

⁷⁰³ « *L'exercice de l'activité de mannequin est conditionné à la délivrance d'un certificat médical. Ce certificat atteste que l'évaluation globale de l'état de santé du mannequin, évalué notamment au regard de son indice de masse corporelle, est compatible avec l'exercice de son métier.* » - Article L. 7123-2-1 du code du travail.

18,4), corpulence normale (entre 18,5 et 24,9), surpoids (entre 25 et 29,9), obésité modérée (entre 30 et 34,9), obésité sévère (entre 35 et 39,9) et obésité morbide (au-delà de 40). Même si elles ne sont pas initiées par le droit, ces catégories statistiques peuvent parfois être réutilisées à des fins juridiques. Cela ne signifie pas nécessairement que la catégorie statistique devient une nouvelle catégorie juridique. De manière générale, la qualification statistique d'un fait devient un élément contribuant à le qualifier juridiquement. Il s'agit de la fonction normative d'administration du processus de quantification⁷⁰⁴.

L'utilisation du seuil statistique en droit. Ce phénomène est illustré par l'affaire du logiciel COMPAS. Pour rappel, COMPAS est un système d'IA destiné à évaluer le risque de récidive d'un individu accusé. Utilisé par certains tribunaux états-uniens, le logiciel produit un score de récidive, la variable continue, allant de 1 à 10, 10 représentant le score de récidive le plus élevé possible⁷⁰⁵. Cette métrique est ensuite divisée en trois catégories. La première est étiquetée « risque faible » et regroupe les scores allant de 1 à 4. Les individus ayant obtenu des scores allant de 5 à 7 font partie de la deuxième catégorie intitulée « risque moyen ». Enfin, la catégorie « risque élevé » regroupe les scores entre 8 et 10⁷⁰⁶. La position du score d'un individu par rapport à ces seuils permet de le catégoriser et donc de le qualifier statistiquement. Il aurait été possible de réfléchir à cinq seuils afin de proposer une catégorisation plus nuancée. Il aurait été aussi possible de modifier la valeur du premier seuil à 6 plutôt que 4. Ces choix de conception ont été réalisés par l'entreprise Northpointe et peuvent être contestés. Toutefois, il s'agit des seuils aujourd'hui mis en oeuvre.

Dans un processus de quantification, le seuil statistique délimite les catégories statistiques. Au service du droit, il peut contribuer, tout ou partiellement, à distinguer entre deux catégories juridiques. Là où l'indicateur représente l'exigence que l'on souhaite évaluer, le seuil statistique définit les catégories de résultats pouvant être obtenues et par conséquent la qualification desdits scores.

⁷⁰⁴ SUPLOT Alain, *La Gouvernance par les nombres*, Fayard, 2015.

⁷⁰⁵ MATTU Jeff Larson Julia Angwin, Lauren Kirchner, Surya, « How We Analyzed the COMPAS Recidivism Algorithm », sur *ProPublica*.

⁷⁰⁶ *Ibid.*

2) Le seuil statistique dans le règlement sur l'IA

La place du seuil statistique dans la réglementation européenne sur l'IA. Dans le cadre du règlement sur l'IA, les seuils statistiques sont des outils invoqués à deux niveaux. Tout d'abord, la certification impose le recours à un seuil statistique pour chacune des exigences de fonctionnement afin de déterminer le niveau suffisant pour l'obtenir. Soit le logiciel ne présente pas une qualité suffisante pour être commercialisé et/ou utilisé sur le marché européen, soit il satisfait cet objectif. Par conséquent, le seuil statistique nécessaire est la délimitation de qualité technique entre un système d'IA à haut risque certifié et un système n'atteignant pas ce niveau de certification. À titre d'exemple, la société Microsoft, dans sa proposition de standard pour une IA responsable, propose de fixer un niveau de performance minimale comme objectif à atteindre pour ces systèmes d'IA⁷⁰⁷.

Ensuite, le règlement européen désigne des seuils statistiques dans le cadre des tests à réaliser sur les systèmes d'IA à haut risque. Tout d'abord l'article 9.8 énonce : « *Les tests sont effectués sur la base d'indicateurs et de seuils probabilistes préalablement définis, qui sont adaptés à la destination du système d'IA à haut risque.* »⁷⁰⁸. Le texte réitère ce besoin de façon un peu plus implicite et sur un aspect plus global dans l'article 9.5 : « *Les mesures de gestion des risques visées au paragraphe 2, point d), sont telles que le risque résiduel pertinent associé à chaque danger ainsi que le risque résiduel global lié aux systèmes d'IA à haut risque sont jugés acceptables.* »⁷⁰⁹. Cette disposition admet l'impossibilité d'imposer à un système d'IA à haut risque une absence totale de risque réels pour ses utilisateurs. Dès lors, elle met en lumière l'importance des seuils statistiques car ces derniers vont représenter le caractère acceptable de l'ensemble des risques résiduels que l'utilisation du système d'IA en question entraînera. Il apparaît alors clairement l'enjeu juridique de la fixation des seuils statistiques propres à chaque exigence technique de qualité pour les systèmes d'IA. En effet, la satisfaction d'un seuil de risque par un système d'IA permet de le considérer, ou du moins de contribuer à le considérer, comme présentant un niveau acceptable de risque. Le seuil statistique devient une condition centrale de la qualification juridique d'un risque comme étant acceptable. Malgré un besoin clairement exprimé par le législateur européen, celui-ci ne fournit ni ne propose de modalités d'établissement de ces seuils, qui représentent pourtant le niveau de risque légalement accepté

⁷⁰⁷ « *For each metric, document: 1) any target minimum performance level for all groups, [...]* » dans MICROSOFT, « Microsoft Responsible AI Standard, v2 "General Requirements" », 2022. P. 21.

⁷⁰⁸ « Règlement sur l'intelligence artificielle », 2024. Article 9.8.

⁷⁰⁹ *Ibid.* Article 9.5.

auxquels les utilisateurs et/ou les personnes affectées peuvent être exposés. Ce choix devra donc être réalisé soit au sein des spécifications techniques que sont les normes européennes harmonisées⁷¹⁰, soit par les fournisseurs de systèmes d'IA eux-mêmes.

B-L'équilibre entre les seuils ou la hiérarchisation des exigences de qualité

Annonce du plan. L'élaboration des seuils probabilistes pour satisfaire les différentes exigences de qualité imposées aux systèmes d'IA à haut risque est un défi complexe qui dépasse largement la question purement technique⁷¹¹. L'existence de plusieurs exigences entraînent nécessairement l'élaboration de plusieurs métriques et seuils quantitatifs. Cette mise en nombre des exigences de qualité facilite leur comparaison et pose la question de leur articulation. Ce besoin d'articulation expose alors inévitablement au développement d'une hiérarchisation, qu'elle soit entre les différents indicateurs d'une même exigence **(A)** ou entre les indicateurs d'exigences différentes **(B)**.

1) La hiérarchisation des seuils intra-exigences

L'identification et la mesure facilitées des biais discriminatoires algorithmiques. Le développement de seuils de qualité satisfaisante, au sens du processus de certification européen, offre un outil de comparaison du fonctionnement impossible à concevoir dans le cas d'un décideur humain. Cette différence humain-machine est par exemple particulièrement ressentie concernant l'exigence de minimisation des biais discriminatoires imposée pour la gouvernance des données de conception et pour la fiabilité du système d'IA. Dans le cas d'un décideur humain, s'il semble impossible de mesurer le niveau de biais négatifs ou positifs qu'il possède envers une catégorie de personnes, comparer ses différents niveaux de biais en fonction de l'appartenance d'un individu à tel ou tel groupe l'est encore plus. Dans le cas d'un système d'IA, il n'y a pas plus de difficultés à identifier et mesurer la présence de biais pour une sous-catégorie de population que pour une autre. Du point de vue technique, il ne s'agit que d'un

⁷¹⁰ COMMISSION EUROPÉENNE, « Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence », 2022.

⁷¹¹ CALO Ryan, « Artificial Intelligence Policy », *U.C. Davis Law Review*, 51, 2017. P.14.

changement de variable⁷¹². La méthode scientifique et les outils utilisés sont identiques quel que soit la population analysée. Par conséquent, les concepteurs de systèmes d'IA peuvent observer les conséquences de certains choix de conception à la fois sur le niveau du biais global de leur système d'IA et aussi sur les niveaux de biais pour chaque sous-catégorie de population identifiée.

La mise en balance entre le niveau de biais algorithmiques de différentes catégories de population. Malgré le fait qu'il s'agisse de la même exigence relative aux biais algorithmiques, rien n'impose que les niveaux mesurés des biais identifiés lors de la phase de conception soient égaux entre les catégories d'individus analysées. Lors des tests de détection de biais imposés par le règlement européen sur l'IA, il est ainsi parfaitement possible que les biais mesurés envers les femmes soient considérés comme acceptables, car très faibles, tandis que les personnes souffrant d'un handicap, par exemple, font l'objet d'un biais très important. La situation se complexifie davantage lorsque l'on comprend que toute action visant à modifier le niveau de biais pour une sous-catégorie de population affecte les niveaux de biais des autres populations. Dans l'hypothèse où des modifications sont apportées et que la seconde phase de test met en avant une amélioration importante de la performance sur les personnes souffrant d'un handicap mais une légère baisse pour les femmes, quelle version les concepteurs doivent-ils privilégier ? Qu'en est-il si le gain de performance se fait au détriment d'une perte significative de performance pour les femmes ? Le cas d'illustration choisi ici est volontairement simplifié. La gestion de l'équilibre entre un même indicateur appliqué à différentes populations s'avère en pratique bien plus complexe. Il ne s'agit pas simplement de sacrifier un seuil statistique pour en améliorer un autre tel des vases communicants. La gestion des compromis des métriques intra-exigences est une problématique qui dépasse largement les compétences d'une équipe technique. Certaines normes techniques témoignent déjà de cette prise de conscience en ne s'aventurant pas à fournir aux entreprises une solution à ce problème. Par exemple, le standard sur l'IA de Microsoft renvoie explicitement les entreprises à leur service juridique pour les décisions relatives aux compromis relatifs aux biais algorithmiques⁷¹³. Une possibilité proposée dans ledit standard est de définir au préalable un seuil de différence de performance maximal entre les différentes sous-catégories de

⁷¹² La détection des biais algorithmiques reste tout de même une étape significative en termes de temps et de ressources. Les outils de détection de biais sont encore au stade de recherche à l'heure actuelle.

⁷¹³ « Consult with your attorney to determine your approach to this, including how you will identify and document tradeoffs. » MICROSOFT, « Microsoft Responsible AI Standard, v2 "General Requirements" », 2022. P. 13.

population⁷¹⁴. L'objectif est de concentrer les ressources sur les sous-catégories de population dont les performances du système d'IA excèdent ce seuil maximum de différence de performance. Si l'idée est intéressante à introduire dans les normes techniques européennes, il convient d'être extrêmement vigilant sur la détermination de ce seuil car il représente l'idée d'une inégalité de performance, entre différentes sous-catégories de population, considérée comme acceptable juridiquement mais aussi socialement, ce qui est très contestable⁷¹⁵.

L'aspect multidimensionnel des indicateurs de biais algorithmiques. De plus, la gestion des différents seuils intra-indicateurs ne se limite pas à deux sous-catégories de population, aussi appelée « variables sensibles » une fois transcrites dans une métrique. En partant du postulat que les biais visés par le règlement européen sur l'IA sont les biais discriminatoires, il y a autant de sous-catégorie de population qu'il y a de critères de discrimination. Conformément au droit européen⁷¹⁶, le nombre de critères de discrimination est de quatorze. En outre, les législations nationales peuvent aussi accroître ce chiffre. Le droit français ajoute ainsi neuf autres critères dont les mœurs, l'apparence physique, la qualité de lanceur d'alerte, la particulière vulnérabilité résultant de la situation économique, etc⁷¹⁷. Cette liste de critères ne représente par ailleurs qu'une vision unidimensionnelle de la problématique. En effet, il est parfaitement possible que des sous-catégories de population puissent être caractérisées par plusieurs critères discriminatoires simultanés. Il s'agit alors de lutter contre la discrimination multiple⁷¹⁸, voire possiblement de mesurer les conséquences intersectionnelles des biais algorithmiques⁷¹⁹. Avec cette approche multidimensionnelle (ou intersectionnelle), il apparaît clairement que la recherche d'un équilibre parfait entre autant de seuils statistiques, même pour une seule

⁷¹⁴ « The target maximum (absolute Or, relative) performance difference between groups. » Ibid. P. 13.

⁷¹⁵ Cet exemple renvoie aux dispositions de l'article 9.5 du règlement européen sur l'IA qui dispose de l'idée de risque résiduels acceptables pour les systèmes d'IA à haut risque, voir paragraphe précédent.

⁷¹⁶ Directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en œuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique, 180 2000; Directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail, 303 2000; Directive 2004/113/CE du Conseil du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans l'accès ` des biens et services et la fourniture de biens et services, 373 2004; Directive 2006/54/CE du Parlement européen et du Conseil du 5 juillet 2006 relative à la mise en œuvre du principe de l'égalité des chances et de l'égalité de traitement entre hommes et femmes en matière d'emploi et de travail (refonte), 204 2006.

⁷¹⁷ Ces critères sont listés à l'article L. 1132-1 du code du travail et à l'article 225-1 du code pénal.

⁷¹⁸ Pour une approche européenne, voir notamment : FREDMAN Sandra, « Intersectional discrimination in EU gender equality and non-discrimination law », Commission européenne, 2016.

⁷¹⁹ « L'approche intersectionnelle va au-delà d'une simple reconnaissance de la multiplicité des systèmes d'oppression opérant à partir de ces catégories et postule leur interaction dans la production et la reproduction des inégalités sociales » - BILGE Sirma, « Théorisations féministes de l'intersectionnalité », Diogene, n° 225, Presses Universitaires de France, 2009. P.1. Sur ce sujet, voir notamment Kimberlé Crenshaw, « Demarginalizing the Intersection of Race and Sex: À Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics » (1989) 1989 U Chi Legal F 139; Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, PMLR, 2018.

exigence, s'avère être une tâche colossale, mais indispensable, qui passe par la gestion des outils statistiques.

Les seuils statistiques représentent concrètement le niveau d'exigence de qualité du fonctionnement attendu des systèmes d'IA présents sur le marché européen. Ils sont la représentation précise et technique du seuil d'acceptabilité légale des systèmes d'IA. Par conséquent, il est plus que souhaitable d'exiger que ce seuil d'acceptabilité des risques soit similaire, voire identique, pour toutes les catégories de population des résidents européens. Une certification globale des biais, et des risques discriminatoires associés, reviendrait à ignorer ce phénomène de concurrence et de hiérarchisation intra-indicateur. La surveillance de l'ensemble des seuils statistiques représentatifs est indispensable, afin de s'assurer que ce ne sont pas systématiquement les mêmes groupes d'individus qui voient leur seuil statistique sacrifié, c'est-à-dire que la discrimination ne soit pas ou soit mal reconnue.

2) La hiérarchisation des seuils inter-exigences

Le besoin d'articulation entre les exigences imposées aux systèmes d'IA à haut risque. Le phénomène du dilemme des métriques présenté précédemment met aussi en lumière la relation complexe et parfois conflictuelle entre différents indicateurs. Cette tension entre indicateurs ne se limite pas à l'enjeu définitionnel d'une même exigence mais s'étend aussi à la difficulté de satisfaire plusieurs exigences simultanément, c'est-à-dire d'atteindre différents seuils probabilistes. Dans la même logique que les principes éthiques de l'IA, vont apparaître des tensions entre les exigences relatives au fonctionnement des systèmes d'IA à haut risque contenues dans le règlement européen sur l'IA.⁷²⁰ À nouveau, ces tensions soulèvent le besoin d'articulation entre ces exigences. Les deux catégories de métriques d'équité démontrent la difficulté à satisfaire à la fois l'absence de biais dans les résultats et l'exactitude globale du système⁷²¹. De façon similaire, l'opposition entre un système d'IA explicable et/ou transparent est régulièrement mentionnée comme une exigence conflictuelle avec celle d'exactitude. Par

⁷²⁰ Voir B– La présence de conflits entre les principes.

⁷²¹ Cela tient notamment aux modalités de correction des biais algorithmiques qui peuvent impliquer d'ajouter des contraintes d'équité à l'objectif initial (la variable-cible) donné à l'algorithme : SERVICE DE RECHERCHE DU PARLEMENT EUROPÉEN, « A governance framework for algorithmic accountability and transparency - Think Tank », Parlement Européen, 2019. Pour plus de détails, voir les travaux sur l'optimisation multi-objectifs, notamment CHIANDUSSI G., CODEGONE M., FERRERO S. *et al.*, « Comparison of multi-objective optimization methodologies for engineering applications », *Computers & Mathematics with Applications*, 63, 2012.

exemple, l'usage des systèmes d'IA d'identification biométrique seront soumis à une obligation spécifique de transparence par l'article 50 du règlement européen : « *Les déployeurs d'un système de reconnaissance des émotions ou d'un système de catégorisation biométrique informent les personnes physiques qui y sont exposées du fonctionnement du système [...] »*⁷²². Dans le cas de la reconnaissance faciale, les technologies les plus performantes sont extrêmement opaques (les réseaux de neurones à convolution par exemple). L'exigence d'exactitude entre donc en tension avec l'obligation spécifique de transparence.

La hiérarchisation du fait de l'importance accordée à l'exigence. Les interactions entre les indicateurs issus d'obligations différentes peuvent produire une hiérarchisation de deux manières. Dans un premier cas, les deux exigences en question se voient attribuer des seuils statistiques non équivalents. Face à la difficulté de satisfaire deux exigences en tension, le niveau de conformité de l'une est moins élevé que l'autre au regard de l'état de l'art. Le législateur européen n'affiche ni officiellement ni explicitement une hiérarchie entre les obligations contenues dans le règlement européen sur l'IA. De plus, la requête de la Commission européenne de normalisation à destination des organismes de normalisation européens⁷²³, à savoir le CEN et le CELENEC, ne laisse pas apparaître d'attention plus insistante sur certaines dispositions plutôt que d'autres. En effet, l'annexe II dudit document contient notamment des exigences spécifiques concernant certaines normes techniques de qualité. Mais il appert que la totalité des normes techniques commandées par la Commission européenne font l'objet de précisions spécifiques, aucun d'eux n'est laissé à la marge. Malgré cela, il est toujours possible que cette hiérarchisation explicite des exigences se retrouve dans les normes européennes harmonisées, à l'initiative des organisations de normalisation. En raison de contraintes techniques méconnues de la Commission européenne notamment, le CEN et le CENELEC pourraient mettre en avant certaines normes techniques plutôt que d'autres. Les seuils statistiques sont alors une formalisation d'une hiérarchie consciente entre les obligations par les organisations de normalisation.

La hiérarchisation du fait de la nature de l'exigence. Dans un second cas, l'élaboration de seuils statistiques, et même de métriques, correspond au niveau de conformité européen des systèmes d'IA et peut jouer un rôle plus actif sur la hiérarchisation entre les exigences. Le

⁷²² *Législation sur l'intelligence artificielle, supra note 1, article 50.3. La suite de la disposition prévoit néanmoins une exception pour les systèmes d'IA de catégorisation biométrique autorisés par loi à des « fins de prévention et de détection des infractions pénales et d'enquêtes en la matière. »*

⁷²³ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence, supra note 424.*

chapitre 2 du titre III du règlement européen sur l'IA contient plusieurs exigences qui ne sont pas toutes égales face au processus de quantification. Par leur nature, certaines dispositions sont plus facilement transcriposables en métriques et seuils statistiques que d'autres. Par exemple, il est plus difficile de quantifier la transparence d'un système d'IA ou son contrôle humain que son exactitude ou sa robustesse. La facilité à être quantifié peut amener à créer une hiérarchie de fait entre les obligations juridiques, ou bien modifier une hiérarchie déjà présente entre elles. Premièrement, l'établissement d'un seuil statistique laisse en effet moins de marge d'interprétation au fournisseur du système d'IA sur l'objectif à atteindre pour obtenir la certification nécessaire à la mise sur le marché de son produit. À l'inverse, une formulation plus littérale pourra donner lieu, d'une part, à des disparités de niveau de conformité entre les fournisseurs de systèmes d'IA et, d'autre part, à une interprétation à la baisse du niveau de conformité exigée, dès lors que cela permet au fournisseur une réduction de coûts économiques ou humains. Deuxièmement, il est important de réaliser que cette hiérarchie de fait entre les obligations en raison de leur nature est un phénomène déjà observé qui sera renforcé par la réutilisation des seuils statistiques en droit. Lors de l'étude des textes éthiques, il était déjà apparu que le secteur privé démontrait une préférence nette pour les processus et obligations quantifiables⁷²⁴. Le caractère aisément quantifiable d'une obligation s'explique par le fait qu'il s'agit d'une préoccupation considérée comme importante par l'industrie elle-même depuis longtemps⁷²⁵. L'appropriation du seuil statistique expose donc à un risque de renforcement d'enjeux techniques ou factuelles au détriment d'obligations juridiques nouvelles pourtant cruciales pour garantir la qualité d'un système d'IA.

CONCLUSION DE LA SECTION I :

Les preuves techniques se révèlent être une pierre angulaire de la stratégie de certification des systèmes d'IA à haut risque par l'Union européenne. Elles sont destinées à représenter en pratique la vision de la qualité européenne de fonctionnement desdits systèmes d'IA. Ce rôle est reconnu partiellement par le législateur européen qui leur accorde une importance formelle au sein du règlement sur l'IA. Cette importance formelle vient s'ajouter à l'importance considérable de fond de la détermination des preuves techniques, bien moins prise

⁷²⁴ Pour plus d'informations sur le recours aux principes éthiques en fonction des acteurs, voir A– L'appropriation des principes éthiques par les acteurs.

⁷²⁵ Si les acteurs du secteur privé souhaitent majoritairement envisager l'équité sous le prisme des biais algorithmiques, cela s'explique en partie parce que les biais sont un problème plus large que l'équité et pose des problèmes plus globaux de performance. Ainsi ces acteurs privilégient la réutilisation de méthodes connues, réduisant ainsi les coûts économiques de mise en conformité.

en compte par le texte européen. Tout d'abord, en l'absence d'exigences précises de qualité, les indicateurs s'apparentent à une définition scientifique de la conception de la qualité qui serait conforme aux exigences juridiques du marché européen. En effet, en décidant de la manière dont est évalué un aspect du fonctionnement des systèmes d'IA à haut risque, l'indicateur détermine non seulement la preuve mais également son objet. Ensuite, les seuils probabilistes symbolisent le niveau d'acceptabilité validé par le droit de l'UE, qu'il s'agisse de l'acceptation sur le marché européen des logiciels ou des risques encourus par les futures personnes affectées. En plus de cette fonction de catégorisation, le caractère quantitatif des seuils dévoile des risques considérables concernant la hiérarchisation des obligations du règlement européen et plus globalement des droits et libertés des individus. Derrière les termes d'indicateurs et de seuils probabilistes se cachent donc des enjeux juridiques centraux sur le fond (la protection de la sécurité, de la santé et des droits fondamentaux des individus soumis aux logiciels d'IA) et sur la forme du droit de l'IA (les besoins de précision et d'articulation des dispositions).

Malgré ces enjeux majeurs, l'Union européenne a fait le choix de laisser les éléments techniques de preuves être déterminés lors de la normalisation. Par conséquent, cette tâche, et les enjeux associés, est confiée à un nouveau groupe d'acteurs qui aura la responsabilité de sélectionner les preuves techniques les plus conformes au droit de l'Union.

Section II : L'influence des acteurs dans la définition des normes techniques

Annonce de plan. Bien que les éléments techniques de preuves soient mentionnés au sein du règlement sur l'IA, ce dernier ne les définit pas. Le choix des indicateurs et des seuils probabilistes, constitutifs des tests de fonctionnement du système d'IA, sont des éléments, à la fois extrêmement contextuels et hautement complexes. Dès lors, les normes techniques apparaissent comme un instrument plus adéquat pour déterminer les éléments techniques de preuve les plus pertinents pour tester la conformité réglementaire des systèmes d'IA. Au regard de la place centrale occupée par les preuves techniques au sein de la réglementation, les normes techniques deviennent alors un point d'intérêt majeur dans l'étude de la réglementation sur l'IA. Les acteurs capables d'exercer une influence sur le contenu des normes techniques se voient donc octroyés, *de facto*, un pouvoir conséquent.

La détermination du contenu des normes techniques sur l'IA, et donc des preuves techniques, peut se diviser en deux temps. Tout d'abord, le processus d'élaboration desdites normes est un instant privilégié pour penser *ab initio* leur contenu. Les acteurs participant à cette phase bénéficient donc d'un pouvoir de fait important sur la sélection des éléments techniques de preuve (**Sous-section 1**). Ensuite, les normes techniques élaborées font l'objet de contrôles destinés à vérifier leur conformité aux règles de droit. Les acteurs participant à cette seconde étape ont alors la possibilité de contester les choix des preuves techniques réalisés lors de l'élaboration de la norme technique (**Sous-section 2**).

Sous-section 1 : Les acteurs en charge de l'élaboration des normes techniques

Annnonce de plan. Que ce soit pour les systèmes d'IA ou pour tout autre produit et service⁷²⁶, la production de normes techniques est assurée par des organismes de normalisation. Les normes techniques étant des outils anciens⁷²⁷, leur recours dans le but d'uniformiser la conception d'un produit n'est en aucun cas l'apanage de l'Union européenne. Dès lors, le terme « organisme de normalisation » ne renvoie pas spécifiquement à des institutions de l'Union européenne ou à des sous-traitants exclusifs à cette dernière mais à une multitude d'acteurs très divers. Toutefois, notre recherche portant sur les exigences contenues dans le règlement européen sur l'IA, les organismes concernés seront distingués ici au travers de leur relation avec l'Union européenne, entre les normalisateurs européens (**§1**) et les normalisateurs extra-européens (**§2**).

§1. Les organismes de normalisation officiels de l'Union européenne

La composition des organismes de normalisation européens. Tout d'abord, le système de normalisation de l'Union européenne se décompose en deux niveaux. Le premier niveau se situe à l'échelle des États membres, il s'agit des organismes de normalisation nationaux. Le second niveau est européen et regroupe trois organisations européennes de normalisation: le

⁷²⁶ *supra* note 7.

⁷²⁷ Benoît Frydman & Arnaud Van Waeyenberge, *Gouverner par les normes techniques et les indicateurs. De Hume aux rankings / Frydman Benoît et van Waeyenbergen Arnaud (dir.)*, Bruylant, 2013.

CEN⁷²⁸, le CENELEC⁷²⁹ et l'ETSI⁷³⁰. Avec le règlement sur l'IA, la Commission européenne a officialisé l'encadrement des systèmes d'IA comme une préoccupation européenne. Ainsi, il n'est plus possible pour les normalisateurs nationaux de lancer des initiatives qui iraient à l'encontre des normes harmonisées européennes⁷³¹. Par conséquent, l'analyse se concentrera sur les organisations européennes. Le CEN, le CENELEC et l'ETSI sont des associations à but non-lucratif de droit belge ou français⁷³² et sont les seules organisations de normalisation reconnues par l'Union européenne⁷³³. Le CEN et le CENELEC sont composés de représentants des organismes de normalisation nationaux de tous les États membres⁷³⁴ ainsi que les États membres de l'AELE⁷³⁵, la Turquie et la Macédoine⁷³⁶. Depuis le règlement 1025/2012, l'Union européenne a encouragé la diversification des membres des normalisateurs européens dans l'objectif de faciliter « *la représentation appropriée et la participation effective de toutes les parties prenantes, notamment les PME, des associations de consommateurs et des parties prenantes environnementales et sociales, à leurs activités de normalisation* »⁷³⁷. Pour autant, à titre d'exemple⁷³⁸, parmi les 319 organismes partenaires européens du CEN, seule une petite minorité représentent lesdits intérêts, l'immense majorité étant des représentants d'associations de secteurs d'industrie⁷³⁹. L'ETSI, quant à lui, compte actuellement plus de 900 membres provenant de plus de 60 pays répartis sur cinq continents. Il assure ainsi une représentativité mondiale plus significative que ses deux homologues restés encore très « européens » dans leur

⁷²⁸ Comité européen de normalisation, chargé des normes techniques dans la plupart des secteurs d'activité.

⁷²⁹ Comité européen de normalisation électrotechnique, spécialisé dans les normes relatives à l'électrotechnique.

⁷³⁰ Institut européen des normes de télécommunications, spécialisé dans les normes relatives aux télécommunications.

⁷³¹ Pour plus de détails sur la Nouvelle approche, voir §2. La concentration sur les exigences essentielles : le recours à la nouvelle approche.

⁷³² Le CEN et le CENELEC sont des associations de droit belge tandis que l'ETSI relève du droit français : Arnaud Van Waeyenberge, « La normalisation technique en Europe. L'Empire (du droit) contre-attaque » (2018) t. XXXII:3 *Revue internationale de droit économique* 305-317, p.309

⁷³³ « Règlement (UE) n ° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n ° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012. Annexe I.

⁷³⁴ À titre d'exemple, l'AFNOR est l'organisme national français de normalisation.

⁷³⁵ Il s'agit de l'Islande, de la Norvège et de la Suisse.

⁷³⁶ WAEYENBERGE Arnaud Van, « La normalisation technique en Europe. L'Empire (du droit) contre-attaque », *Revue internationale de droit économique*, t. XXXII, 2018.

⁷³⁷ « Règlement (UE) n ° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n ° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012, *supra* note 79, article 5.1.

⁷³⁸ Le CENELEC est dans une situation similaire à celle du CEN : « The CEN and CENELEC Annual Report 2021 is out! », sur *CEN-CENELEC*.

⁷³⁹ Voir rapport 2021 du CEN : *Ibid*.

composition. Loin d'être anodine, cette différence de composition entre l'ETSI et le CENELEC est un facteur très important dans l'élaboration des normes européennes harmonisées. Dans le contexte de l'IA, bien que les trois normalisateurs européens démontrent des compétences sur le sujet, seuls le CEN et le CENELEC ont été mandatés pour l'élaboration des normes techniques relatives au règlement européen sur l'IA⁷⁴⁰. L'ETSI se trouve quant à lui réduit à un rôle de conseil sur une liste de sujets précis. Cette exclusion de l'ETSI par la Commission européenne s'intègre dans un mouvement plus global de mise à distance de ladite organisation de normalisation⁷⁴¹. Sa gouvernance est considérée comme trop influencée par le secteur privé, plus particulièrement par des acteurs non-européens comme la Chine et les États-Unis⁷⁴². Cette décision illustre parfaitement la conscience par l'Union européenne du risque d'appropriation du cadre légal sur l'IA au travers des normes techniques. Toutefois, l'adoption d'un texte législatif a permis au législateur européen de réduire les divergences d'influence non désirées liées à l'origine géographique ou au secteur privé. Au-delà de la présence d'enjeux politiques, sociaux et économiques dans l'élaboration des normes techniques⁷⁴³, cette décision renforce le constat selon lequel les normes européennes harmonisées ont un rôle essentiel aux yeux du législateur européen, notamment en raison de leur travail de précision et d'articulation avec les règles de droit.

Le processus de création des normes techniques. Au sein du CEN et du CENELEC, l'initiative d'un projet de standard technique peut se manifester de deux manières. Le plus souvent, le processus de développement est une initiative interne à la suite de la constatation d'un besoin dans le secteur concerné⁷⁴⁴, généralement par ses membres communs, à savoir les représentants des normalisateurs nationaux⁷⁴⁵. De manière moins fréquente, ces organisations

⁷⁴⁰ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, *supra* note 49, article 1.

⁷⁴¹ Luca Bertuzzi, « Commission leaves European normalisation body out of AI standard-setting », (7 décembre 2022), en ligne: [www.euractiv.com](https://www.euractiv.com/section/artificial-intelligence/news/commission-leaves-european-normalisation-body-out-of-ai-standard-setting/) <<https://www.euractiv.com/section/artificial-intelligence/news/commission-leaves-european-normalisation-body-out-of-ai-standard-setting/>>.

⁷⁴² *Ibid.* BERTUZZI Luca, « European Commission sets out a plan to regain clout in standard-setting », sur [Www.euractiv.com](https://www.euractiv.com), publié le 3 février 2022.

⁷⁴³ *supra* note 7.

⁷⁴⁴ Ces initiatives internes ont notamment mené à la constitution par le CEN-CENELEC d'un « *focus group* » sur l'IA en décembre 2018 et à l'élaboration d'une feuille de route sur le sujet en 2019 : « -ARTIFICIAL INTELLIGENCE | Joinup », en ligne: <<https://joinup.ec.europa.eu/collection/rolling-plan-ict-normalisation/artificial-intelligence>>; Martin Ebers, « Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act » dans, Rochester, NY, Social Science Research Network, 2021.

⁷⁴⁵ « European Normes techniques », en ligne: *CEN-CENELEC* <<https://www.cencenelec.eu/european-standardization/european-normes-techniques/>>.

de normalisation peuvent recevoir une requête, de la part de la Commission européenne⁷⁴⁶ ou d'autres parties prenantes⁷⁴⁷. Concernant les futures normes techniques relatives au règlement sur l'IA, leur élaboration a été commandée expressément par la Commission européenne par une requête du 12 mai 2022⁷⁴⁸.

Désormais mandaté par l'Union européenne, le CEN et le CENELEC peuvent lancer leurs procédures d'élaboration de normes techniques. Lorsqu'un projet est proposé, il est nécessaire qu'un nombre suffisant de membres souhaitent s'y impliquer pour que la procédure se poursuive et que le projet soit confié à un comité technique compétent⁷⁴⁹. Le développement des normes techniques est alors exclusivement entre les mains des membres de ce comité, à savoir les membres du CEN et du CENELEC qui sont intéressés par travailler sur le projet. Les partenaires des institutions européennes, des associations d'industriels ou de la société civile ou encore des membres d'organismes internationaux de normalisation peuvent assister au processus de réflexion dudit comité technique en qualité d'observateur. Après parfois plusieurs années de travail, un consensus est trouvé entre les membres du comité technique et une proposition de standard est présentée aux restes des membres du CEN et du CENELEC. Celle-ci fait ensuite l'objet d'une enquête publique ouverte à toute partie intéressée. À la suite de cette enquête, les retours sont évalués et font l'objet d'un vote final. Si le résultat du vote final est satisfaisant, le projet de norme technique est adopté et ladite norme devient un standard officiel.

Le manque de légitimité démocratique des organismes européens de normalisation. Au regard des enjeux entourant les preuves techniques de fonctionnement des systèmes d'IA imposées aux fournisseurs, la légitimité des deux organisations européennes de normalisation désignées par la Commission européenne représente un enjeu décisif sur lequel il est nécessaire

⁷⁴⁶ « [...] Dans les limites des compétences fixées dans les traités, la Commission peut demander à une ou plusieurs organisations européennes de normalisation d'élaborer une norme européenne ou une publication en matière de normalisation européenne dans un délai déterminé. » - « Règlement (UE) n° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012. Article 10.1.

⁷⁴⁷ Voir le site du CEN – CENELEC : « European Standards », sur *CEN-CENELEC*.

⁷⁴⁸ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence.*

⁷⁴⁹ S'il n'existe pas de comité technique compétent sur le sujet, il est créé pour l'occasion. Dans le cas de l'IA, un comité technique joint (JTC n°21) entre le CEN et le CENELEC a été créé en 2020 à la suite de la publication du livre blanc sur l'IA de la Commission européenne : « CEN and CENELEC launched a new Joint TC on Artificial Intelligence », sur *CEN-CENELEC*.

de s'attarder⁷⁵⁰. La légitimité, entendue au sens du droit européen, repose « *d'une part sur le caractère démocratique de la procédure de leur adoption ; d'autre part, sur leur conformité à l'état de droit, sous le contrôle des juridictions.* »⁷⁵¹. Cette légitimité est donc intimement liée au principe de représentation appropriée des parties prenantes⁷⁵², consacré et encouragé par le règlement 1025/2012⁷⁵³. La prédominance des intérêts privés industriels en comparaison à d'autres intérêts⁷⁵⁴ démontre déjà un problème dans l'équilibre des intérêts représentés. En réalité, les dispositions du règlement 1025/2012 sont davantage programmatiques qu'opérationnelles. Les grandes lignes sont indiquées mais les précisions concernant le contenu demeurent très faibles⁷⁵⁵. De plus, le manquement à ce principe de représentation appropriée n'est accompagné d'aucune sanction. Cette absence de contraintes aussi bien sur le fond que sur la forme s'inscrit dans l'esprit de la stratégie « nouvelle approche »⁷⁵⁶. Les organismes de normalisation européens bénéficient d'une marge de manœuvre significative sur le processus d'élaboration des normes techniques qui deviendront des normes européennes harmonisées⁷⁵⁷. En effet, le statut d'observateur, accordé aux partenaires, n'octroie pas de droit de vote dans le processus de création d'un standard et leur capacité à faire appel d'un projet de standard est limitée⁷⁵⁸. En plus des contraintes procédurales, les partenaires des organisations européennes de normalisation doivent aussi faire face à des obstacles plus pratiques. La participation au processus d'élaboration nécessite une certaine expérience en normalisation que ne possèdent pas une très grande partie des associations représentant les consommateurs ou tout autre catégorie de la société civile. Les réunions nécessitent des déplacements fréquents à travers le

⁷⁵⁰ L'Union européenne elle-même qualifie le principe de représentation appropriée des intérêts dans les organismes de normalisation européens comme un « élément crucial » du fait de l'intérêt public en jeu : *Avenir de la normalisation européenne Résolution du Parlement européen du 21 octobre 2010 sur l'avenir de la normalisation européenne (2010/2051(INI))*, (2010), par. 9

⁷⁵¹ *supra* note 7.

⁷⁵² Dagmar Schiek, *Private Rule-Making and European Governance - Issues of Legitimacy*, Rochester, NY, 2007, pp. 443-452

⁷⁵³ « Règlement (UE) n° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012.

⁷⁵⁴ Voir précédemment la proportion de représentants d'acteurs privés industriels par rapport à des associations de consommateurs ou de défense de l'environnement.

⁷⁵⁵ Les éléments en question sont présents uniquement dans les considérants dudit texte : Pierluigi Cuccuru, *Interest Representation in European Normalisation: The Case of CEN and CENELEC*, Rochester, NY, 2019.

⁷⁵⁶ Pour rappel, la nouvelle approche désigne une stratégie européenne de réglementation qui prend la forme de renvois par le texte juridique à des normes techniques. Le texte énonce alors uniquement les « exigences essentielles » de la réglementation tandis que les normes techniques précisent les « spécifications techniques » associées.

⁷⁵⁷ CUCCURU Pierluigi, *op. cit.*

⁷⁵⁸ Voir le règlement intérieur 2022 du CEN et du CENELEC – Partie 1 : « Reference documents », en ligne: BOSS CEN <<https://boss.cen.eu/reference-material/refdocs/pages/>>

monde, ce qui nécessite du temps et des moyens financiers considérables pour suivre les nombreuses années du processus⁷⁵⁹. Comme la composition des partenaires le laissait présumer, le fonctionnement du CEN et du CENELEC ne permet pas réellement une participation effective des organismes partenaires⁷⁶⁰. Par conséquent, il appert que le mandat confié par la Commission européenne au CEN et au CENELEC sera réalisé par deux institutions amplement dépourvues de légitimité démocratique. Toutefois, les organisations européennes de normalisation ne sont pas à l'origine de toutes les normes européennes harmonisées, il est fréquent qu'elles s'inspirent, voire réutilisent des normes techniques produites par d'autres organismes de normalisation⁷⁶¹. Les différentes décisions autour de la preuve technique peuvent donc être réalisées par d'autres acteurs auxquels il faut aussi accorder une place dans cette analyse.

§2. Les organismes de normalisation non officiels de l'Union européenne

Enjeux de la collaboration entre les organismes de normalisation européens et internationaux pour la normalisation de l'IA. Dans les considérants de sa requête de normalisation destinée au CEN et au CENELEC, la Commission européenne indique l'intérêt d'impliquer les organismes de normalisation internationaux lors de l'élaboration des normes techniques. Il existe plusieurs organismes de normalisation internationaux⁷⁶² mais la Commission mentionne explicitement l'organisation internationale de la normalisation (l'ISO)⁷⁶³ et la Commission internationale sur l'électrotechnique (IEC)⁷⁶⁴. La reconnaissance et la désignation de ces acteurs n'a rien de surprenant de la part de l'Union européenne, que ce soit dans le contexte de l'IA ou dans un contexte plus global. Premièrement, l'IEC et ISO

⁷⁵⁹ CUCCURU Pierluigi, *op. cit.*

⁷⁶⁰ EBERS Martin, « Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act », in *The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics*, Social Science Research Network, 2022.

⁷⁶¹ FRYDMAN Benoit et VAN WAEYENBERGE Arnaud, *Gouverner par les standards et les indicateurs. De Hume aux rankings / Frydman Benoît et van Waeyenbergen Arnaud (dir.)*, Bruylant, 2013.

⁷⁶² Outre l'ISO et l'IEC cités ci-après, on retrouve également l'Institute of Electrical and Electronics Engineers (IEEE), l'International Telecommunications Union (ITU) et l'Internet Engineering Task Force (IETF). Il existe également des organismes régionaux de normalisation représentant l'Afrique, l'Amérique latine, les pays de l'Asie du Sud-Est, etc. Pour des raisons de clarté et de pertinence, ils ne seront pas traités et nous ne nous concentrerons que sur les normalisateurs les plus influents dans le contexte de la normalisation européenne sur l'IA.

⁷⁶³ International Organization for Standardization.

⁷⁶⁴ International Electrotechnical Commission.

collaborent depuis 2017 au travers d'un comité technique joint⁷⁶⁵ aujourd'hui à l'initiative de la majorité des travaux de normalisation dans le monde. En effet, les normes ISO/IEC relative à la robustesse dans l'IA, à la terminologie de l'IA⁷⁶⁶, la gouvernance⁷⁶⁷, l'éthique⁷⁶⁸ ou encore l'apprentissage automatique⁷⁶⁹ ont été élaborées par ce comité. L'expérience accumulée au cours des discussions menées par les différents sous-groupes de travail est un avantage considérable dont souhaite profiter l'Union européenne pour l'encadrement des systèmes d'IA⁷⁷⁰. Deuxièmement, l'implication des organismes internationaux de normalisation permet de mieux appréhender l'articulation entre les normes techniques. Si le règlement européen sur l'IA vient se superposer aux autres cadres juridiques existants, la Commission européenne est consciente qu'il en va de même pour les normes techniques. L'implication de l'ISO et de l'IEC est nécessaire pour conserver la conformité de produits, déjà sujets à des normes techniques précis, qui pourraient intégrer des technologies d'IA. À ce titre, l'annexe I du règlement sur l'IA sert non seulement à délimiter le cadre d'application des systèmes d'IA à haut risque, et donc du règlement lui-même, mais il liste aussi un ensemble de directives « nouvelle approche ». Les nombreuses normes techniques associées, dont certaines sont produites ou coproduites par les organismes internationaux de normalisation, devront être prises en compte ou modifiées pour assurer leur compatibilité avec les futures normes techniques spécifiques à l'IA. En d'autres termes, l'ISO et l'IEC sont impliqués pour répondre à un besoin d'articulation et de convergence des normes techniques.

Les relations entre les organismes européens et internationaux de normalisation. Si les usages multisectoriels des systèmes d'IA rendent la collaboration de l'IEC et de ISO plus nécessaire que d'accoutumée, ce travail d'équipe est en réalité une habitude consacrée en droit européen. Ainsi, l'accord de Vienne est un accord de coopération signé entre l'ISO et le CEN accordant la liberté du contexte de négociation des normes techniques⁷⁷¹. Le CEN est alors libre de décider si les groupes de travail auront lieu dans le contexte de l'ISO ou du CEN. Une fois

⁷⁶⁵ Le ISO/IEC JTC1/SC 42 a été le premier comité technique à travailler sur la normalisation autour de l'IA : EBERS Martin, « Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act », in *The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics*, Social Science Research Network, 2022.

⁷⁶⁶ ISO/IEC DIS 22989.

⁷⁶⁷ ISO/IEC DIS 38507.

⁷⁶⁸ ISO/IEC AWI TR 24368.

⁷⁶⁹ ISO/IEC DIS 23053.

⁷⁷⁰ « *International normalisation can, on the one hand, help to consolidate a common vision of trustworthy AI [...]* », *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, supra note 49, considérant 7.

⁷⁷¹ Débuté en 1989, l'accord de Vienne est publié en juin 1991 et sera confirmé en 2001 aux termes des résolutions 35/2001 pour l'ISO et 2/2001 pour le CEN.

le standard élaboré, une procédure de vote parallèle est mise en place, ce qui garantit une adoption commune simultanée par les deux organismes de normalisation⁷⁷². Un accord de coopération similaire, l'accord de Francfort⁷⁷³, est signé entre l'IEC et le CENELEC. Ces accords sont rendus possibles car les organismes européens et internationaux de normalisation ont des fonctionnements très proches. D'une part, tous les membres du CEN et du CENELEC, à savoir les organismes de normalisation nationaux, sont également membres de l'ISO et de l'IEC. D'autre part, les procédures de vote sont similaires entre les deux catégories d'organisme. Par conséquent, la parallélisation des procédures n'entraîne pas de difficultés particulières.

La légitimité démocratique des organismes internationaux de normalisation : l'exemple de l'ISO. Le rôle et la coopération forte entre les organismes européens et internationaux de normalisation impose de considérer ces derniers comme des acteurs significatifs de l'élaboration des normes européennes harmonisées. Par conséquent, il est important d'étudier leur légitimité démocratique au même titre que le CEN et le CENELEC. Pour cela, il sera pris comme exemple la figure de proue de ces organismes internationaux : l'ISO. Comme énoncé précédemment, le développement des normes techniques au sein de l'ISO suit un chemin très similaire à celui du CEN et du CENELEC, que ce soit au niveau des membres ou de la procédure de vote. Par conséquent, c'est sans surprise que les défaillances démocratiques présentes dans les organismes de normalisation européens se retrouvent au sein de l'ISO. Ces lacunes sont même exacerbées au niveau international car l'ISO (ainsi que l'IEC) n'est pas soumise à la réglementation européenne. L'organisation n'est donc pas contrainte de respecter le principe de représentation appropriée prévu par le règlement (UE) 1025/2012⁷⁷⁴. Les représentants de la société civile (associations de consommateurs, parties prenantes environnementales, etc.) ne font même plus face à des obstacles pratiques de participation car ils sont dès le départ mis à l'écart par le fonctionnement de l'ISO⁷⁷⁵. Au regard de la définition de la légitimité fournie précédemment, les organismes internationaux de normalisation présentent donc une légitimité,

⁷⁷² *supra* note 7.

⁷⁷³ Anciennement connu sous le nom de l'Accord Dresden (1996), l'Accord de Francfort a été voté en 2016 : « IEC - CENELEC Agreement on Common planning of new work and parallel voting (Dresden Agreement) - 1996 | IEC ».

⁷⁷⁴ « Règlement (UE) n ° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n ° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012. Article 5.

⁷⁷⁵ Mariolina Eliantonio & Pierluigi Cuccuru, « It is not all about judicial review: Internal appeal proceedings in the European normalisation process » dans *Le futur du droit administratif/The Future of Administrative Law*, Lexis Nexis, 2019 475.

en tant qu'acteur de l'élaboration des futures normes européennes harmonisées, encore plus critiquable que le CENELEC et le CEN.

Les organismes de normalisation sectoriels sollicités pour les normes techniques « verticales ». L'ISO et l'IEC sont des organismes extrêmement puissants dans l'écosystème de la normalisation et leurs compétences sont très vastes. Il existe cependant des organisations de normalisation aux compétences plus spécifiques qui élaborent des normes techniques pour un domaine précis. Malgré ce champ d'expertise plus restreint, il reste des acteurs pertinents dans l'élaboration de ce que la Commission européenne appelle des « spécifications verticales »⁷⁷⁶ qui visent à traiter des « risques verticaux ». Le terme « vertical » s'oppose au terme « horizontal »⁷⁷⁷ et renvoie à des problématiques propres à certains systèmes d'IA du fait de leur contexte d'utilisation ou à certains secteurs d'activité⁷⁷⁸. Contrairement aux spécifications horizontales, la Commission européenne ne mentionne pas d'organismes de normalisation susceptibles d'aider le CEN et le CENELEC dans leurs tâches.

Les autres types de normes techniques. Il est indispensable de garder à l'esprit que les preuves techniques seront déterminées par toute norme technique relative aux technologies d'IA, et non uniquement au sein des normes européennes harmonisées. Ces dernières ne seront pas les seules normes destinées à la conception des systèmes d'IA à haut risque. De nombreuses autres normes techniques privées, facultatives, seront produites et détermineront, ou aideront à identifier, les preuves techniques adéquates pour attester de la conformité d'un système d'IA aux réglementations applicables.

Tout d'abord, il est clair que la grande majorité des fournisseurs de systèmes d'IA à haut risque privilégieront l'application des normes européennes harmonisées. En effet, au pouvoir incitatif juridique de la présomption de conformité au règlement sur l'IA⁷⁷⁹ s'ajoute des considérations

⁷⁷⁶ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, supra note 49, article 5(c) et Annexe II, section 1.

⁷⁷⁷ Le terme « horizontal » renvoie ici à des exigences pertinentes pour tous les systèmes d'IA à haut risque au sens du futur règlement européen : *Ibid*, note de bas de page 4.

⁷⁷⁸ *Ibid*.

⁷⁷⁹ « Règlement sur l'intelligence artificielle », 2024. Article 40.1 (version compromis tchèque)

financières⁷⁸⁰, déterminantes dans la réflexion des acteurs du secteur. La CJUE a reconnu elle-même à plusieurs reprises ce constat⁷⁸¹.

Malgré ces inconvénients, il demeure des alternatives et certaines organisations peuvent souhaiter suivre d'autres normes techniques que celles proposées par le CEN et le CENELEC. Dans cette situation, la déclaration de conformité nécessitera une procédure particulière, plus contraignante que si le fournisseur avait choisi de respecter les normes harmonisées⁷⁸². Malgré ces contraintes supplémentaires, certains fournisseurs, en raison des spécificités de leur secteur par exemple, pourront choisir cette voie alternative en suivant soit des normes techniques tierces à celles de l'UE ou en élaborant eux-mêmes des normes techniques. L'existence de ces alternatives impose de nuancer la position, qui peut sembler hégémonique, des organismes de normalisation dans les choix de preuves techniques associées aux exigences du règlement sur l'IA. En effet, les acteurs absents de la composition de l'ISO, IEC, CEN et CENELEC (notamment les associations de consommateurs et les autres représentants de la société civile) peuvent être présents au sein d'autres organismes de normalisation moins importants. Dès lors, ils posséderont un pouvoir d'influencer le contenu de ces normes techniques, et donc le choix des preuves techniques considérées comme pertinentes. Toutefois, l'influence desdites normes techniques sur le choix des preuves produites par les fournisseurs de systèmes d'IA sera moindre car ces normes techniques seront moins suivies que celles produites par les principaux organismes de normalisation. À la différence des normes harmonisées, les normes techniques produites de manière indépendante au système de normalisation européen ne sont soumises à aucune exigence ou procédure spécifique européenne. Par conséquent, il existe un risque encore plus important d'appropriation et de dénaturation de l'exigence juridique par le biais de la détermination des éléments techniques de preuves appliqués au droit. Souvent éclipsées par les normes harmonisées et leur présomption de conformité, les autres normes techniques sont aussi un espace où les enjeux de précision et d'articulation liés à la preuve technique doivent également être considérés. L'origine des normes techniques au sein desquelles les preuves techniques de conception des systèmes d'IA sont précisées n'influence en rien le rôle central que ces dernières occupent dans le droit européen de l'IA. Il est donc essentiel de rester attentif

⁷⁸⁰ La Commission européenne prévoit ainsi des coûts supplémentaires pour les fournisseurs de système d'IA qui ne suivraient pas les procédures existantes dans l'état de l'art en ce qui concerne les exigences de robustesse et de précision : COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021. PP. 66-68.

⁷⁸¹ Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P. Par. 75. Voir également Cour de justice de l'Union européenne, 12 juillet 2012, n° C-171/11. Par. 29.

⁷⁸² Voir *supra*, Chapitre 2 : Le contrôle des preuves de conception.

à chacun des instruments normatifs par lesquels les preuves techniques pourraient être déterminées et certifiées par le droit européen.

Sous-section 2 : Les acteurs en charge du contrôle des normes techniques

Annonce de plan. Pour devenir des normes européennes harmonisées, les normes techniques produites par les organismes de normalisation européen doivent être publiées au Journal officiel de l'Union européenne. Si les organismes européens de normalisation (ainsi que l'ISO et l'IEC) ont le pouvoir de proposer une rédaction des normes techniques portant sur les systèmes d'IA à haut risque, ces normes techniques sont ensuite soumises à différents contrôles avant et après leur publication au Journal officiel de l'UE. Il s'agit d'un contrôle sur la conformité des normes techniques par rapport au droit européen⁷⁸³. Au regard des lacunes en termes de légitimité démocratique et du risque d'appropriation des notions juridiques identifiées lors de la phase d'élaboration desdites normes, cette phase de contrôle se révèle extrêmement importante. Cette attention grandissante est confirmée par la jurisprudence européenne (§1) qui implique alors de s'intéresser à ce processus de contrôle des normes techniques européennes (§2) ainsi qu'aux conséquences qui en découlent (§3).

§1. L'importance juridique des contrôles

Annonce de plan. Face aux limites de légitimité et de contrôle démocratiques lors de l'élaboration des normes techniques, il semble naturel d'exiger un contrôle fort dans le but de compenser lesdites lacunes. Ce contrôle est assuré par deux acteurs à deux moments distincts : la Commission européenne préalablement à la publication des normes techniques au Journal officiel de l'UE (A) et les institutions judiciaires européennes, plus particulièrement la CJUE, postérieurement à la publication officielle (B).

⁷⁸³ *Ibid.*

A-Le contrôle de la Commission avant la publication.

L'obligation d'un contrôle préalable à la publication au Journal officiel. Étant à l'origine du mandat, contenant les exigences relatives aux normes techniques, à destination du CEN et du CENELEC, la Commission européenne est particulièrement bien placée pour contrôler le résultat de son mandat de normalisation. En l'espèce, elle mandate des prestataires, le CEN et le CENELEC, pour obtenir un produit final, à savoir des normes techniques servant de spécifications techniques aux exigences contenues dans son règlement européen. Par conséquent, il est logique de vérifier que la prestation a été réalisée conformément au cahier des charges. Pour autant, les normes européennes harmonisées ont longtemps été directement publiées au Journal officiel de l'Union européenne sans qu'aucun contrôle ne soit prévu ou exercé⁷⁸⁴. La Commission européenne manifestait alors une confiance aveugle envers le travail des organismes de normalisation européens⁷⁸⁵ et se limitait à de rares objections formelles sur les normes techniques produits.

Ce contrôle extrêmement superficiel prit fin avec l'entrée en vigueur du règlement européen 1025/2012. Ce-dernier impose à la Commission européenne de contrôler les normes techniques qu'elle a commandé aux organisations européennes de normalisation :

« [...] En coopération avec les organisations européennes de normalisation, la Commission évalue la conformité des documents élaborés par les organisations européennes de normalisation avec sa demande initiale. »⁷⁸⁶

La validation par la Commission européenne des normes techniques lors de cette phase de contrôle est une condition *sine qua non* de leur publication au Journal officiel de l'Union

⁷⁸⁴ « Unlike in the case of direct referencing, an ex- ante control of the technical work by the legislator does not take place. The European legislator has faith in the accountability of the European normes techniques system. » - *Methods of referencing normes techniques in legislation with an emphasis on European regulation (Enterprise Guides)*, par Gerhard Leibrock, Bruxelles, Commission européenne, 2002. P. 9

⁷⁸⁵ « The Commission does not engage in any technical verification of the normes techniques and publishes their references blindly » Harm Schepel et al, *The Constitution of Private Governance: Product Normes techniques in the Regulation of Integrating Markets*, Oxford ; Portland, Or, Hart Publishing, 2005. P. 135

⁷⁸⁶ Règlement (UE) n ° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE, supra note 79. Article 10.5

européenne⁷⁸⁷. En cas de refus, la norme technique produite reste alors un simple standard privé et non un texte de droit européen à part entière⁷⁸⁸.

B- Le contrôle judiciaire après la publication.

Les fondements du contrôle judiciaire des normes harmonisées. Une fois les normes techniques publiées, la Cour de justice de l'UE joue le rôle de contrôleur. Plusieurs arguments permettent d'appuyer l'importance de ce contrôle judiciaire sur les normes techniques. Tout d'abord, le contrôle judiciaire d'actes juridiques, comme les normes européennes harmonisées, est un principe fondateur au cœur de la création de l'Union européenne. La Cour de justice rappelait ainsi en 1986 :

*« La Communauté économique européenne est une communauté de droit en ce que ni ses États ni ses institutions n'échappent au contrôle de la conformité de leurs actes à la charte constitutionnelle de base qu'est le traité. [...] Le traité a établi un système complet de voies de recours destiné à confier à la Cour de justice le contrôle de la légalité des actes des institutions. »*⁷⁸⁹

Ensuite, le contrôle judiciaire *ex post* des normes européennes harmonisées est central pour les citoyens concernés au regard de leur droit à un recours effectif et à accéder à un tribunal impartial, consacré aux articles 6 et 13 de la Convention européenne des droits de l'Homme⁷⁹⁰ et à l'article 47 de la Charte des droits fondamentaux de l'Union européenne⁷⁹¹. Ce droit a par ailleurs été qualifié de principe général du droit de l'Union par la Cour de justice à plusieurs reprises⁷⁹².

Enfin, la jurisprudence actuelle de la Cour de justice de l'Union européenne sur la normalisation européenne met l'accent sur la place du contrôle judiciaire comme garantie. Cette position de la Cour est d'autant plus intéressante qu'il n'en a pas toujours été ainsi. Au

⁷⁸⁷ « Lorsqu'une norme harmonisée répond aux exigences qu'elle vise à couvrir et qui sont définies dans la législation correspondante d'harmonisation de l'Union, la Commission publie une référence à cette norme harmonisée sans retard au Journal officiel de l'Union européenne [...] » - Article 10.6 du règlement 1025/2012.

⁷⁸⁸ Cour de justice de l'Union européenne, 3ème chambre, 27 octobre 2016, n° C-613/14. Pour plus de détails, voir B - Le contrôle judiciaire par la CJUE.

⁷⁸⁹ Cour de justice de l'Union européenne, 23 avril 1986, n° 294/83. Par. 23

⁷⁹⁰ CONSEIL DE L'EUROPE, « Convention européenne des droits de l'Homme », 1950. Article 6 : Droit à un procès équitable et Article 13 : Droit à un recours effectif.

⁷⁹¹ « Charte des droits fondamentaux de l'Union européenne », OJ C, 2012.

⁷⁹² « le principe de protection juridictionnelle effective constitue un principe général du droit » dans Cour de justice de l'Union européenne, 16 juillet 2009, n° C-12/08. Par 47 ; Cour de justice, 22 décembre 2010, n° C-279/09. Par. 29

départ, l'arrêt *Meroni*⁷⁹³ conditionnait une délégation de pouvoir, tel que le recours à la normalisation, à trois conditions : l'existence d'une délégation expresse desdits pouvoirs, les pouvoirs d'exécution visés devaient être « *nettement délimités* »⁷⁹⁴ et l'exercice de ces pouvoirs devait faire l'objet d'un contrôle rigoureux similaire à celui qui serait applicable à l'autorité délégante. Au sein de ce régime jurisprudentiel, les imprécisions dans les exigences juridiques d'un texte se limitant aux « exigences essentielles » et le mandat de normalisation émis par la Commission européenne ne satisfont pas les conditions de délégation de pouvoirs⁷⁹⁵. Toutefois, depuis l'arrêt *ESMA*⁷⁹⁶ rendu en 2014, la Cour de justice a assoupli lesdites conditions, venant nuancer sans pour autant inverser les contradictions de la normalisation européenne avec l'arrêt *Meroni*. Désormais, une délégation de pouvoirs est possible dès lors qu'elle fait l'objet d'un contrôle judiciaire adéquat. Face à cette évolution de jurisprudence, plusieurs auteurs considèrent que les lacunes identifiées lors de l'élaboration des normes techniques européennes sont tolérées dès lors qu'elles sont compensées par un contrôle judiciaire *ex post* approprié⁷⁹⁷. En d'autres termes, le contrôle judiciaire des normes techniques européennes est perçu comme un mécanisme permettant de corriger les problèmes juridiques identifiés lors de leur élaboration.

§2. Les modalités d'exercice des contrôles

Annonce de plan. Bien qu'essentielles, les preuves techniques relatives aux systèmes d'IA à haut risque (indicateurs, seuils probabilistes, etc.) ne représentent qu'une faible, et hautement complexe, fraction du contenu des normes techniques qui seront produits par le CEN et le CENELEC. À l'image du règlement lui-même, ces normes contiendront également des spécifications organisationnelles et documentaires. L'existence de moyens de contrôles sur ces normes techniques n'implique pas qu'ils soient suffisamment précis pour s'intéresser aux préoccupations relatives aux preuves techniques. Il faut également s'assurer que ces contrôles,

⁷⁹³ Cour de justice de l'Union européenne, 13 Juin 1958, n° 9-56.

⁷⁹⁴ *Ibid.* P. 43

⁷⁹⁵ Arnaud van Waeyenberge & David Restrepo, « James Elliot construction: À “New(ish) approach” to judicial review of normalisation » (2017) 42 European Law Review 882-893. P890 : « *regulatory tasks for harmonised normes techniques bodies Or, private/public networks such as the CEN clearly exceed the simple framework for the delegation of powers in the sense of Meroni.* »

⁷⁹⁶ Cour de justice de l'Union européenne, 22 janvier 2014, n° C-270/12.

⁷⁹⁷ Mariolina Elia Antonio, « Judicial Control of the EU Harmonized Normes techniques: Entering a Black Hole? » (2017) 44 Legal Issues of Economic Integration; van Waeyenberge & Restrepo, « James Elliot construction », *supra* note 580.

en particulier le contrôle judiciaire, sont effectivement actionnables par les acteurs intéressés. En d'autres termes, les modalités des contrôles exercés sur les normes techniques européennes déterminent la capacité des acteurs en charge de ces contrôles à contribuer à la détermination des preuves techniques. En suivant une logique chronologique, il sera étudié tout d'abord les modalités, d'initiation et de réalisation, du contrôle de la Commission européenne **(A)** puis celles relatives aux contrôles exercés par les juridictions européennes **(B)**.

A-Le contrôle par la Commission européenne.

Un contrôle superficiel du contenu des normes produites par les organisations européennes de normalisation. Tout d'abord, il convient de mentionner que l'obligation d'évaluation de conformité est énoncée à deux reprises dans le règlement 1025/2012 à l'article 10(5) et (6)⁷⁹⁸. Les deux versions de cette même obligation varient légèrement dans leur formulation, ce qui a pu donner lieu à des analyses détaillées des conséquences de ces différences, en particulier concernant l'objet de l'évaluation⁷⁹⁹ et le moment de son exécution⁸⁰⁰. Bien qu'intéressantes, ces réflexions ne seront pas traitées en détail car notre attention se porte ici davantage sur la profondeur de l'évaluation de conformité menée par la Commission européenne sur les normes techniques produites, en l'espèce celles concernant les systèmes d'IA à haut risque.

Premièrement, la réalisation d'une évaluation détaillée et approfondie de la conformité des normes techniques s'oppose totalement à la logique même de la Nouvelle approche⁸⁰¹, en plus de surcharger de travail la Commission européenne⁸⁰². Deuxièmement, la Commission européenne admet elle-même les limites de son analyse de conformité :

⁷⁹⁸ Voir *supra*, §1. L'importance juridique des contrôles.

⁷⁹⁹ L'article 10.5 dispose d'une évaluation au regard de la « *demande initiale* » de la Commission européenne, à savoir la requête de normalisation, tandis l'article 10.6 renvoie aux exigences « *définies dans la législation correspondante d'harmonisation de l'Union [...]* ». En pratique, les deux objets sont très souvent identiques. Voir *Legal Opinion On the European System of Harmonised Normes techniques*, par Kathrin Dingemann & Matthias Kottmann, Berlin, German Federal Ministry for Economic Affairs and Energy, 2020. P. 27.

⁸⁰⁰ *Ibid.*

⁸⁰¹ EBERS Martin, « Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act », in *The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics*, Social Science Research Network, 2022. P. 10

⁸⁰² Pour rappel, le CEN et le CENELEC ont publié sur l'année 2021 un total de 1617 livrables dont 1409 normes européennes, soit presque 4 normes par jour à évaluer pour la Commission européenne. « The CEN and CENELEC Annual Report 2021 is out! », sur *CEN-CENELEC*.

« En règle générale, la Commission ne doit pas examiner l'adéquation technique du contenu d'une norme. Les OEN [organisations européennes de normalisation] sont responsables du contenu et sont censées fournir des données correctes. Leurs procédures doivent garantir que les titres traduits, tels qu'ils sont présentés, renvoient au numéro d'identification correct de la norme. »⁸⁰³

Cette position de la Commission européenne vis-à-vis des normes techniques européennes n'a pas évolué suite à l'entrée en vigueur du règlement 1025/2012 qui a mis à jour la stratégie de la Nouvelle approche. Dans son « Guide bleu » de 2016 relatif à la mise en œuvre de la réglementation de l'UE sur les produits, la Commission rappelle explicitement les rôles et les responsabilités dans l'évaluation de conformité desdites normes techniques :

« Elles [les organisations européennes de normalisation] assument l'entière responsabilité du contenu technique de ces normes. [...] la législation d'harmonisation de l'Union ne prévoit pas de procédure permettant aux pouvoirs publics de vérifier systématiquement ou d'approuver, au niveau de l'Union ou au niveau national, le contenu des normes harmonisées adoptées par les organisations européennes de normalisation. »⁸⁰⁴

« Au cours de cette vérification, une révision du contenu technique n'est pas nécessaire dans la mesure où, en règle générale, la Commission ne prend pas position à l'égard du contenu technique ni n'en assume la responsabilité. »⁸⁰⁵

Dans la très grande majorité des cas, l'évaluation se limite à une comparaison formelle entre le contenu des différentes normes techniques et les exigences contenues dans la législation d'harmonisation et dans la requête de normalisation de la Commission européenne. En outre, la Commission européenne fait appel à un prestataire privé externe pour réaliser cette évaluation de conformité. Pour les requêtes réalisées auprès du CEN et du CENELEC, ce prestataire est le cabinet de conseil *Ernest & Young*⁸⁰⁶. En tant qu'intermédiaire entre la Commission européenne et les organismes européens de normalisation, ces consultants ont aussi pour mission de conseiller les organismes pour la mise en conformité de leurs normes techniques avec la requête

⁸⁰³ *Guidelines for the publication of references of normes techniques in the Official Journal of the European Union*, Commission européenne, 2005. P. 3. Traduction réalisée par l'auteur.

⁸⁰⁴ « Le Guide bleu relatif à la mise en œuvre de la réglementation de l'Union européenne sur les produits 2016 (2016/C 272/01) », Commission européenne, 2016. P. 41.

⁸⁰⁵ *Ibid.* P. 45

⁸⁰⁶ « European Normes techniques for citation in the Official Journal », en ligne: BOSS CENELEC <<https://boss.cenelec.eu/homegrowndeliverables/en/pages/enforojeu/>>.

de normalisation⁸⁰⁷. Tel qu'inscrit sur le site du CENELEC, si le résultat de l'évaluation de conformité est positif, alors « *la publication des références au JOUE est normalement possible sans préavis particulier* »⁸⁰⁸. Toutefois, l'évaluation de conformité ne semble pas totalement sous-traitée car la Commission européenne doit elle-aussi évaluer et valide *in fine* la proposition de norme technique produite par les organismes de normalisation⁸⁰⁹. Cette intervention d'un tiers privé ne permet pas de confirmer ou d'infirmer notre hypothèse sur l'absence de contrôle des preuves techniques au sein des normes européennes harmonisées relatives aux systèmes d'IA à haut risque. Néanmoins, la mention de ce sous-traitant est importante car il s'agit d'un nouvel acteur qui aura sans aucun doute une influence non négligeable sur le contenu des normes techniques et donc potentiellement les choix des preuves techniques de conception des systèmes d'IA.

Enfin, le règlement européen sur l'IA offre une petite nuance sur le contenu du contrôle de la Commission européenne portant sur les normes techniques que produiront le CEN et le CENELEC. Le texte mentionne la possibilité pour la Commission européenne de produire des spécifications communes sous certaines conditions, notamment lorsque « *les normes harmonisées ne sont pas conformes à la demande* »⁸¹⁰ mais aussi quand « *les normes harmonisées pertinentes ne répondent pas suffisamment aux préoccupations en matière de droits fondamentaux* »⁸¹¹. Cette préoccupation envers la protection des droits fondamentaux semble ainsi s'ajouter au contrôle de conformité traditionnel de la Commission européenne. Ce contrôle pourrait donc être plus complet que de coutume. Le texte n'indique pas explicitement si cet examen spécifique aux droits fondamentaux sera conduit par la Commission européenne ou par un acteur spécialisé. De la même manière, le législateur européen n'indique pas si cette condition de prise en compte insuffisante des droits fondamentaux peut être satisfaite après la publication des normes au Journal officiel de l'UE.

⁸⁰⁷ « European Normes techniques for citation in the Official Journal », en ligne: BOSS CENELEC <<https://boss.cenelec.eu/homegrowndeliverables/en/pages/enforojeu/>>.

⁸⁰⁸ « *The publication of the references in the OJEU is normally possible without any specific notice.* » - « European Normes techniques for citation in the Official Journal », en ligne: BOSS CENELEC <<https://boss.cenelec.eu/homegrowndeliverables/en/pages/enforojeu/>>.

⁸⁰⁹ « *La CE et le contractant HAS ont convenu de valider bilatéralement chaque "évaluation" avant de la fournir au CENELEC, et ce dans un délai de 5 semaines (35 jours calendaires).* » ; « *Lorsque les OEN proposent les références des normes harmonisées (hEN) pour citation, la CE vérifie et évalue chaque hEN nouvelle et révisée/modifiée en tenant compte des rapports d'évaluation et d'autres preuves éventuellement disponibles. Elle prend ensuite la position finale sur la publication des références d'un hEN au JOUE.* » - « European Normes techniques for citation in the Official Journal », en ligne: BOSS CENELEC <<https://boss.cenelec.eu/homegrowndeliverables/en/pages/enforojeu/>>.

⁸¹⁰ « Règlement sur l'intelligence artificielle », 2024, Version finale du trilogue. Article 41.1a)iv).

⁸¹¹ *Ibid.* Version finale du trilogue. Article 41.1a)iii).

En dépit de cet ajout, il demeure difficile d’imaginer que cela approfondisse suffisamment le contrôle de conformité pour que la Commission européenne évalue effectivement les décisions relatives aux preuves techniques. Ce constat met bien en avant l’influence extrêmement faible, voire inexistante en pratique, de la Commission européenne lors de son évaluation de la conformité des normes techniques. Cette situation, qui ne semble pas être remise en question dans le dernier « guide bleu » de 2022⁸¹², s’applique *a fortiori* sur les choix relatifs aux preuves techniques de conception des systèmes d’IA à haut risque.

B - Le contrôle judiciaire par la CJUE

Annonce de plan. La CJUE peut être amenée à se prononcer sur la validité des normes européennes harmonisées de deux manières différentes : le recours en annulation énoncé à l’article 263 TFUE (1) et le renvoi préjudiciel de la part d’une juridiction nationale au sens de l’article 267 TFUE (2).

1) Le recours direct en annulation auprès de la CJUE

Un mécanisme de contrôle du contenu des normes européennes harmonisées. Dans un premier temps, il existe un recours direct qui offre à la CJUE la compétence de se prononcer sur des actes « *des organes ou organismes de l’Union destinés à produire des effets juridiques à l’égard des tiers* » dès lors que le recours est formé soit par « *un État membre, le Parlement européen, le Conseil ou la Commission* » soit par « *toute personne physique et morale* » qui est concernée « *directement et individuellement* » par l’acte ou les actes contestés. Ce recours direct offre aux acteurs mentionnés la possibilité d’initier, auprès de la CJUE, un contrôle des normes techniques, et par conséquent des choix techniques et juridiques qu’elles contiennent. Pour ces acteurs, il s’agit donc d’une opportunité de corriger des décisions de normalisation relatives aux preuves techniques des systèmes d’IA à haut risque qui ne sembleraient pas conformes aux exigences du règlement européen sur l’IA, ou plus généralement au regard du droit de l’Union européenne. Toutefois, le recours en annulation est subordonné à deux séries de conditions : la

⁸¹² « Le «Guide bleu» relatif à la mise en œuvre de la réglementation de l’UE sur les produits (2022/C 247/01) », Commission européenne, 2022.

première relative à l'acte contesté en lui-même et la seconde sur la qualité de la personne effectuant le recours.

La nécessité d'une production d'effets juridiques par les normes européennes harmonisées. En ce qui concerne les conditions liées à l'acte contesté, ce dernier doit tout d'abord disposer d'effets juridiques. En ce qui concerne les normes européennes harmonisées, la présomption de conformité acquise lorsqu'elles sont référencées aux Journal officiel de l'Union européenne peut constituer un effet juridique pour plusieurs auteurs⁸¹³. Cette interprétation est par ailleurs confortée par la décision *Global Garden Products Italy SpA v. European Commission*⁸¹⁴ dans laquelle le tribunal énonce que la publication au JO UE confère une valeur juridique à la norme harmonisée⁸¹⁵. Néanmoins il est important de comprendre que la condition des « effets juridiques » est satisfaite seulement pour les exigences et les systèmes contenues dans la requête de normalisation de la Commission européenne, à savoir les systèmes d'IA présentant un risque élevé et les exigences contenus dans le chapitre 2 titre III du règlement européens sur les systèmes d'IA. À *contrario*, les normes techniques qui seront élaborées pour les autres systèmes d'IA (les systèmes d'IA ne présentant pas un risque élevé, les systèmes d'IA à finalité générale, ...) ne pourront pas faire l'objet d'un recours en annulation au sens de l'article 263 du TFUE. Par conséquent, les potentielles normes techniques relatives aux algorithmes de modération ou de publicité en ligne, utilisés par les fournisseurs de plateformes en ligne, seront situées hors du champ d'application de l'article 263 du TFUE. Pour autant, le fonctionnement de ces algorithmes est surveillé et encadré par le droit de l'Union au travers du récent règlement sur les services numériques 2022/2065⁸¹⁶ (ci-après DSA). Ce-dernier consacre la notion de « *risques systémiques*⁸¹⁷ » pouvant être entraînés par plusieurs facteurs dont la

⁸¹³ Eliantonio, « Judicial Control of the EU Harmonized Normes techniques », *supra* note 582. Harm Schapel, « The New Approach to the New Approach: The Juridification of Harmonized Normes techniques in EU Law » (2013) 20:4 Maastricht Journal of European and Comparative Law 521-533.

⁸¹⁴ Tribunal de l'Union européenne (Deuxième chambre), 26 janvier 2017, n° T-474/15.

⁸¹⁵ « *c'est la publication par la Commission de la référence d'une norme harmonisée au Journal officiel qui confère à celle-ci une valeur juridique permettant aux fabricants de machines ou à leurs mandataires de bénéficier, pour les machines qu'ils commercialisent et qui y sont conformes, d'une présomption de conformité* » - Ibid. Par. 60.

⁸¹⁶ « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022.

⁸¹⁷ Le DSA désigne les risques systémiques suivants : « a) la diffusion de contenus illicites par l'intermédiaire de leurs services ; b) tout effet négatif réel ou prévisible pour l'exercice des droits fondamentaux [...] ; c) tout effet négatif réel ou prévisible sur le discours civique, les processus électoraux et la sécurité publique ; d) tout effet négatif réel ou prévisible lié aux violences sexistes et à la protection de la santé publique et des mineurs et les conséquences négatives graves sur le bien-être physique et mental des personnes. Ibid. Article 34.1. Ladite notion et ses conséquences seront développées plus amplement au cours du chapitre 5 de la présente thèse.

majorité sont directement liés à des systèmes algorithmiques, et donc généralement des systèmes d'IA :

- « a) la conception de leurs systèmes de recommandation et de tout autre système algorithmique pertinent;
- b) leurs systèmes de modération des contenus ;
- c) les conditions générales applicables et leur mise en application ;
- d) les systèmes de sélection et de présentation de la publicité ;
- e) les pratiques du fournisseur en matière de données. »⁸¹⁸

Il ne fait aucun doute que des normes techniques seront élaborées concernant le fonctionnement des différents systèmes d'IA qui seront utilisés aux fins énoncées ci-dessus. De la même manière que pour les exigences contenues dans le règlement européen sur l'IA, ces normes techniques, sous l'apparence de normes purement techniques, contiendront des choix politiques et juridiques au travers de la manière de mesurer les risques systémiques, de les atténuer⁸¹⁹ ainsi que le seuil d'acceptabilité recommandé à partir duquel le fournisseur de plateforme en ligne s'inscrit en conformité avec l'état de l'art.

La condition de paternité de l'acte appliquée aux normes européennes harmonisées.

Ensuite, l'acte contesté doit répondre au critère de paternité, c'est-à-dire qu'il doit faire partie des « *actes législatifs, des actes du Conseil, de la Commission et de la Banque centrale européenne, autres que les recommandations et les avis, et des actes du Parlement européen et du Conseil européen* »⁸²⁰ ou des « *actes des organes ou organismes de l'Union* »⁸²¹. La satisfaction des normes harmonisées à cette condition n'est pas certaine et dépend de l'interprétation qu'en fera la CJUE. D'un côté, les normes harmonisées ont pour auteurs les organisations européennes de normalisation (CEN, CENELEC et ETSI) et ne sont communiquées que sous forme de référence par la Commission européenne⁸²², ce qui peut les

⁸¹⁸ *Ibid.* Article 34.2

⁸¹⁹ « Les fournisseurs de très grandes plateformes en ligne et de très grands moteurs de recherche en ligne recensent, analysent et évaluent de manière diligente tout risque systémique au sein de l'Union découlant de la conception ou du fonctionnement de leurs services et de leurs systèmes connexes, y compris des systèmes algorithmiques, ou de l'utilisation faite de leurs services. » *Ibid.* Article 35

⁸²⁰ « Traité sur le fonctionnement de l'Union européenne (version consolidée) », 2012. Article 263, alinéa 1

⁸²¹ *Ibid.*

⁸²² La qualité d'auteur des organisations de normalisation européennes est renforcée par le fait qu'elles sont autrices de leurs normes européennes harmonisées élaborées au sens du droit de la propriété intellectuelle.

exclure du champ d'application de l'article 263 du TFUE selon certains auteurs⁸²³. De l'autre côté, de nombreux auteurs considèrent que l'acte de publication de la part de la Commission européenne est en soi un acte contre lequel un recours en annulation est possible⁸²⁴. Les divergences d'opinions et les incertitudes des auteurs se cristallisent principalement sur les conséquences de l'arrêt *James Elliott* qui qualifie les normes européennes harmonisées de « [...] mesure de mise en œuvre nécessaire et strictement encadrée des exigences essentielles définies par cette directive, réalisée à l'initiative et sous la direction ainsi que le contrôle de la Commission [...] »⁸²⁵.

Contrairement à l'opinion de l'avocat général⁸²⁶, la CJUE ne définit pas explicitement les normes européennes harmonisées comme des actes produits par des institutions de l'Union. Bien que l'arrêt *James Elliott Construction* semble poser les bases d'un contrôle judiciaire *ex post* plus efficace contre les normes techniques européennes, la position de la CJUE reste alors incertaine. Le récent arrêt *Global Garden Products* de 2017 énonce que « les décisions relatives à la publication des normes harmonisées sont des actes juridiques susceptibles de recours en annulation »⁸²⁷. En l'espèce, et bien que l'arrêt *James Elliott Construction* dispose d'une autorité plus importante, cette dernière décision nous semble démontrer que les normes harmonisées peuvent faire l'objet d'un recours en annulation au sens de l'article 263 du TFUE. Toutefois, en l'absence de confirmation explicite par la CJUE, il ne sera pas nié la possibilité d'une approche inverse de la question par cette dernière. Par conséquent, il n'est pas encore certain, mais seulement hautement probable, qu'un recours en annulation au sens de l'article 263 du TFUE envers l'acte de communication d'une norme harmonisée par la Commission européenne, comme celles à venir dans le cadre du règlement européen sur l'IA, soit considéré comme recevable.

Les conditions relatives au demandeur du recours en annulation. La seconde série de conditions pour admettre la recevabilité d'un recours en annulation porte sur l'intérêt à agir du demandeur. Dans le contexte des normes techniques, la partie demanderesse doit généralement

⁸²³ SCOTT Joanne, « In legal limbo », *COLA*, 48, 2011. Eliantonio, « Judicial Control of the EU Harmonized Normes techniques », *supra* note 582. P. 399.

⁸²⁴ SCHAPPEL Harm, « The New Approach to the New Approach », *Maastricht Journal of European and Comparative Law*, 20, SAGE Publications Ltd, 2013. VAN WAEYENBERGE Arnaud et RESTREPO David, « James Elliott construction », *European Law Review*, 42, 2017. P. 13. MAZEAU Laurène, « Responsabilité. Les enjeux de la normalisation technique dans le domaine de l'intelligence artificielle », *Cahiers Droit, Sciences & Technologies*, PUP, 2019. P. 230.

⁸²⁵ VAN WAEYENBERGE Arnaud et RESTREPO David, *op. cit.*

⁸²⁶ 28 janvier 2016, n° C-613/14. Par. 40 à 63.

⁸²⁷ Tribunal de l'Union européenne (Deuxième chambre), 26 janvier 2017, n° T-474/15. Par. 60. La décision renvoie également à l'arrêt *James Elliott Construction*.

démontrer que l'acte ou les actes contestés « *dont elle est le destinataire ou qui la concernent directement et individuellement, ainsi que contre les actes réglementaires qui la concernent directement et qui ne comportent pas de mesures d'exécution.* »⁸²⁸. Au préalable, il convient d'écarter les États membres, la Commission européenne, le Parlement européen et le Conseil qui sont des requérants privilégiés. En d'autres termes, ils n'ont pas besoin de satisfaire les exigences de *locus standi* pour intenter un recours en annulation⁸²⁹. En outre, il a été mis en lumière que l'utilisation par ces requérants privilégiés d'un recours en annulation est en pratique très peu probable⁸³⁰.

L'intérêt à agir à l'encontre des normes européennes harmonisées. En ce qui concerne les autres types de requérants⁸³¹, la qualification juridique des normes harmonisées pose à nouveau problème au sein des auteurs⁸³². Il sera considéré ici que la communication de la Commission européenne des références des normes européennes harmonisées sont des mesures d'exécutions au sens du dernier cas de figure de l'article 263 alinéa 5 du TFUE⁸³³. Par conséquent, et ce depuis le traité de Lisbonne, il n'est pas nécessaire pour le requérant de prouver qu'il est concerné individuellement par la norme harmonisée contestée. En revanche il doit toujours prouver que celle-ci le concerne directement. Les organisations nationales de normalisation, notamment celles qui ont été minoritaires lors des votes au CEN, CENELEC ou ISO, pourraient prétendre à satisfaire cette condition. Dans un raisonnement similaire, les parties prenantes, énoncées à l'article 10.5 du règlement 1025/2012, présentes lors de l'élaboration desdites normes mais dont la participation est mise à mal par de nombreux facteurs⁸³⁴ semblent avoir la qualité pour agir. Cependant, il convient de nuancer cette possibilité car la CJUE a tendance à interpréter très strictement les exigences de *locus standi* dans le cas des recours en annulation⁸³⁵.

⁸²⁸ « Traité sur le fonctionnement de l'Union européenne (version consolidée) », 2012. Article 263 alinéa 4

⁸²⁹ *Ibid.* Article 263 alinéa 2.

⁸³⁰ HOFMANN Herwig C.H., ROWE Gerard C. et TÜRK Alexander H., « Judicial Supervision », in *Administrative Law and Policy of the European Union*, Oxford University Press, 2011.

⁸³¹ Il existe également des requérants semi-privilégiés, à savoir la BCE, la Cour des comptes et le Comité des régions, qui ne seront pas traités ici car ils ne sont pas pertinents dans le contexte des normes techniques européennes appliquées aux systèmes d'IA.

⁸³² Pour une opinion plus nuancée de celle présentée ici, voir Eliantonio, « Judicial Control of the EU Harmonized Normes techniques », *supra* note 582. P. 400

⁸³³ Voir également VAN WAEYENBERGE Arnaud et RESTREPO David, « James Elliot construction », *European Law Review*, 42, 2017. P. 892 ;

⁸³⁴ Voir *supra*

⁸³⁵ Van Waeyenberge & Restrepo, « James Elliot construction », *supra* note 580.

2) Le renvoi préjudiciel par les juridictions nationales

L'insuffisance du renvoi préjudiciel à des fins de contrôle des normes européennes harmonisées. Le recours en annulation n'est pas la seule possibilité d'initiation d'un contrôle de la Cour de justice de l'UE sur les normes européennes harmonisées, plus précisément sur les questions autour des éléments techniques de preuve utilisés pour attester de la conformité d'un système d'IA au règlement européen, et les enjeux qui y sont associés⁸³⁶. L'article 267 paragraphe 1 du TFUE dispose que :

« La Cour de justice de l'Union européenne est compétente pour statuer, à titre préjudiciel :

a) sur l'interprétation des traités,

*b) sur la validité et l'interprétation des actes pris par les institutions, organes ou organismes de l'Union. »*⁸³⁷

Il s'agit de la procédure de renvoi préjudiciel, mise en avant par l'Union européenne comme moyen alternatif au recours en annulation. Le recours contre les normes européennes harmonisées s'inscrit dans le dernier alinéa de l'article, à savoir la capacité de contester « *la validité et l'interprétation des actes pris par les institutions, organes ou organismes de l'Union. »*

Préalablement à l'étude de la possibilité de recourir à cette procédure, il est indispensable de se questionner sur sa capacité à être une réelle alternative au recours en annulation, et donc à être un véritable outil de contrôle des choix de preuves techniques relatives à la conformité des systèmes d'IA au règlement sur l'IA. Cette hypothèse a ainsi été remise en cause par l'avocat général Jacobs, dans l'arrêt *Unión de Pequeños Agricultores v. Council of the European Union*⁸³⁸, qui défend le fait que considérer le renvoi préjudiciel comme un substitut adéquat au recours en annulation est contraire au principe de protection juridictionnelle effective car le renvoi préjudiciel « *ne constitue toutefois pas une voie ouverte de plein droit aux particuliers* »⁸³⁹. Deux raisons viennent appuyer cette position : premièrement, les juges nationaux sont parfaitement libres de refuser de poser des questions préjudicielles à la CJUE, à

⁸³⁶ Voir Section I : L'importance des preuves au sein des normes techniques.

⁸³⁷ « Traité sur le fonctionnement de l'Union européenne (version consolidée) », 2012.

⁸³⁸ Cour de justice de l'Union européenne, 25 juillet 2002, n° C-50/00 P.

⁸³⁹ Cour de justice de l'Union européenne, 21 mars 2002, n° C-50/00 P.. Par. 42

l'exception des juridictions de dernière instance⁸⁴⁰. Pour ces dernières, les délais supplémentaires entraînés par une telle procédure peuvent provoquer des retards incompatibles avec l'exigence d'une sécurité juridique et du principe de protection juridictionnelle effective⁸⁴¹. Par conséquent, il ne sera pas envisagé ici le renvoi préjudiciel comme une possibilité satisfaisante de compenser les limites identifiées dans la procédure du recours en annulation.

En ce qui concerne les conditions d'application du renvoi préjudiciel, il a été démontré précédemment que les conditions liées à l'acte contesté étaient satisfaites pour les normes européennes harmonisées. Il faut maintenant distinguer le recours à la procédure du renvoi préjudiciel, d'une part, pour contester la validité de l'acte et, d'autre part, pour contester l'interprétation de l'acte contesté. En effet, la décision *James Elliott Construction* de 2016 a clairement énoncé la compétence de la CJUE à statuer, à titre préjudiciel, sur les normes européennes harmonisées sur la base de leur interprétation. L'interprétation et la validité sont deux aspects très proches d'un même recours mais il serait imprudent d'étendre la jurisprudence rendue dans le premier cas au second⁸⁴². En attente d'une décision explicite des juridictions européennes, comme ce fut le cas pour le recours en annulation avec l'arrêt *Global Garden Products*, le choix sera fait ici de considérer que le renvoi préjudiciel n'est pas un mécanisme permettant de contester la validité des normes européennes harmonisées mais uniquement leur interprétation.

§3. Les conséquences des contrôles

Annonce de plan. En raison de leur temporalité et de leur nature, les contrôles effectués par la Commission européenne et par la Cour de justice de l'Union européenne ne produisent pas les mêmes effets en cas de non-conformité des normes techniques au règlement européen. Alors que le contrôle judiciaire se révèle limité dans ses effets **(A)**, le contrôle de la Commission européenne lui permet d'être plus proactive **(B)**.

⁸⁴⁰ « Traité sur le fonctionnement de l'Union européenne (version consolidée) », 2012. Article 267 (ex-article 234 TCE), alinéa 4.

⁸⁴¹ À la problématique des retards vient s'ajouter le risque que la formulation des questions préjudicielles par la juridiction nationale limite le nombre de mesures communautaires contre lesquelles le requérant s'est pourvu en justice, voir : Cour de justice de l'Union européenne, 21 mars 2002, n° C-50/00 P..

⁸⁴² Pour une opinion similaire, voir VAN WAEYENBERGE Arnaud et RESTREPO David, « James Elliot construction », *European Law Review*, 42, 2017. P. 891.

A - Les conséquences limitées du contrôle judiciaire.

La limitation à l'acte de publication de la Commission européenne. Dans le cas très optimiste où l'un de ces deux mécanismes de recours aboutisse, il est envisageable que la Cour donne raison au requérant et prononce la non-conformité de la norme harmonisée au règlement européen sur l'IA. Les répercussions d'une telle décision seraient cependant minimales. En effet, les normes techniques produites par les organisations européennes de normalisation ne sont pas des actes produits par les instances européennes au sens des articles 263 et 267 du TFUE. C'est par la communication de la Commission et leur publication au Journal officiel qu'elles acquièrent une valeur juridique⁸⁴³. Par conséquent, c'est cette communication qui subira les conséquences de la décision judiciaire. La ou les normes techniques référencées ne seront pas pour autant supprimées car elles ne sont pas le sujet du litige. Le contrôle de la Cour se limitera à contrôler que la Commission européenne n'a pas fait d'erreurs en publiant lesdites références, c'est-à-dire une vérification de la bonne réalisation du contrôle de conformité au sens de l'article 10.5 et 10.6 du règlement 1025/2012. Or, la technicité des problématiques liées aux preuves techniques des systèmes d'IA à haut risque dépasse largement le niveau d'analyse réalisé par la Commission européenne. Malgré les enjeux et conséquences centrales que représente la détermination de la preuve technique, il n'est pas imaginable que les choix des preuves techniques puissent constituer un fondement suffisant pour faire annuler la communication de la Commission ou corriger l'interprétation qui aurait pu en être faite. En raison de la nature juridique des normes harmonisées et de la technicité du sujet des preuves techniques, il n'est pas possible de considérer que le contrôle judiciaire des normes techniques soit un moyen crédible de validation ou de contestation des choix des indicateurs ou seuils probabilistes réalisé par les organisations européennes de normalisation⁸⁴⁴.

⁸⁴³ Tribunal de l'Union européenne (Deuxième chambre), 26 janvier 2017, n° T-474/15.

⁸⁴⁴ De manière générale, les défauts et limites du contrôle judiciaire laisse à penser que les normes harmonisées sont en pratique intouchables par la CJUE : « *Although harmonized normes techniques have significant legal and practical implications, they are currently in essence immune from judicial review.* » - EBERS Martin, « Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act », in *The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics*, Social Science Research Network, 2022.

B - Les conséquences d'un contrôle par la Commission européenne

L'élaboration de normes techniques complémentaires. Malgré le caractère très formel de l'évaluation de conformité menée par la Commission européennes sur les normes harmonisées, il existe une possibilité que ce contrôle se solde par un refus de publier les références desdites normes techniques dans le Journal officiel de l'Union européenne. Cette situation permet alors à la Commission d'élaborer et proposer ses propres normes techniques, appelés spécifications communes, dont le respect par les acteurs entraînera lui aussi une présomption de conformité aux exigences du règlement européen sur l'IA.

Pour pouvoir bénéficier de cette prérogative et donc d'une capacité de détermination du contenu d'une norme technique, et plus précisément en l'espèce, des modalités de preuve technique, la Commission doit indiquer la réalisation de plusieurs conditions énoncées à l'article 41 du règlement sur l'IA⁸⁴⁵ :

« la Commission, en vertu de l'article 10, paragraphe 1, du règlement (UE) no 1025/2012, a demandé à une ou plusieurs organisations européennes de normalisation d'élaborer une norme harmonisée pour les exigences énoncées à la section 2 du présent chapitre ou, le cas échéant, pour les obligations énoncées au chapitre V, sections 2 et 3, et :

la demande n'a été acceptée par aucune des organisations européennes de normalisation; ou

les normes harmonisées faisant l'objet de cette demande n'ont pas été présentées dans le délai fixé conformément à l'article 10, paragraphe 1, du règlement (UE) no 1025/2012; ou

les normes harmonisées pertinentes ne répondent pas suffisamment aux préoccupations en matière de droits fondamentaux ; ou

les normes harmonisées ne sont pas conformes à la demande ; et

*b) aucune référence à des normes harmonisées couvrant les exigences visées au chapitre II du présent titre n'a été publiée au Journal officiel de l'Union européenne, conformément au règlement (UE) n 1025/2012, et aucune référence de ce type ne doit être publiée dans un délai raisonnable. »*⁸⁴⁶

⁸⁴⁵ L'information selon laquelle lesdites conditions sont satisfaites est notamment communiquée au comité au sens de l'article 22 du « Règlement (UE) no 182/2011 du Parlement Européen et du Conseil du 16 février 2011 établissant les règles et principes généraux relatifs aux modalités de contrôle par les États membres de l'exercice des compétences d'exécution par la Commission », OJ L, 2011.

⁸⁴⁶ « Règlement sur l'intelligence artificielle », 2024. Article 41(1).

Tout d'abord, il faut qu'aucune norme harmonisée n'ait été publiée au Journal officiel sur les objectifs principaux du règlement sur l'IA⁸⁴⁷. Ensuite, la seconde et plus importante condition est le fait que la Commission européenne ait formulé une demande aux organismes de normalisation mais que cette demande rencontre l'un des quatre problèmes listés. En l'espèce, le CEN et le CENELEC ont déjà accepté la requête de normalisation formulé par la Commission européenne, ce qui exclut la première possibilité. Ensuite, il est peu probable que les organismes de normalisation européens produisent leurs normes techniques après l'échéance fixée par la Commission européenne⁸⁴⁸, à savoir le 31 janvier 2025⁸⁴⁹. Les deux sous-conditions envisageables sont donc la prise en compte insuffisante des droits fondamentaux dans les normes harmonisées et leur non-conformité à la requête de la Commission européenne. Dans les deux cas, le problème serait révélé à l'issu du contrôle de conformité réalisé par la Commission européenne. Par conséquent, le recours à des spécifications communes par la Commission européenne relève davantage de la conséquence à la suite d'un contrôle de conformité négatif d'une norme harmonisée que du mécanisme de détermination des preuves techniques⁸⁵⁰. Les spécifications communes remplissant un rôle identique aux normes européennes harmonisées, la Commission européenne bénéficierait le cas échéant d'une capacité de détermination des preuves techniques identiques à celle des organismes de normalisation mandatés actuellement. Cependant, au regard de la capacité de contrôle de la Commission européenne sur les normes techniques produits par le CEN et le CENELEC, la probabilité que la Commission puisse recourir aux spécifications communes est extrêmement faible, voire nulle⁸⁵¹.

CONCLUSION DE LA SECTION II :

Les acteurs du système de normalisation européenne et leur importance respective diffèrent significativement des parties prenantes présentes dans le parcours plus classique suivi

⁸⁴⁷ Pour rappel, ces objectifs sont la sécurité, la santé et le respect des droits et libertés fondamentaux des individus.

⁸⁴⁸ Les coûts d'accès aux normes techniques représentent la très grande partie du chiffre d'affaire annuel des organismes de normalisation. La satisfaction en temps et en heure de la requête de la Commission européenne représente donc des enjeux économiques extrêmement importants qui permettent d'avancer l'improbabilité d'un tel retard.

⁸⁴⁹ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*. Annexe I.

⁸⁵⁰ Ce constat justifie le traitement du mécanisme des spécifications communes dans le paragraphe consacré aux conséquences du contrôle plutôt que dans celui concernant l'élaboration de la norme technique.

⁸⁵¹ Ce constat est renforcé par le fait que la Commission européenne participe, sans droit de vote certes, à l'élaboration des normes techniques. Si un élément est susceptible d'entraîner un refus au moment du contrôle de conformité, les représentants de la Commission ont de grandes chances de le signaler aux membres votants au moment des réflexions au sein des comités techniques.

par les règlements et les directives européennes. La détermination des preuves techniques de conception des systèmes d'IA, particulièrement ceux présentant des risques élevés, est un pouvoir concentré principalement entre les mains des acteurs privés que sont les membres des organisations européennes de normalisation. Les garanties juridiques mises en place se révèlent incapables d'assurer une participation représentative effective des parties prenantes ainsi qu'un contrôle démocratique du processus d'élaboration de ces normes techniques. Malgré les enjeux juridiques importants, la complexité de la question des preuves techniques et de la nature juridique des normes techniques rend toute procédure de contrôle prévue, administrative comme judiciaire, au pire inopérante, au mieux très inadaptée. Par conséquent, il apparaît clairement que le pouvoir de détermination des preuves techniques ne s'exerce qu'au travers de l'élaboration des normes techniques qui les contiennent, et cela par un nombre restreint d'acteurs peu représentatifs des enjeux en présence. Enfin, ce constat est exacerbé lorsque pour les normes techniques autres que celles demandées par la Commission européenne. En ce qui les concerne, ces normes techniques ne font l'objet d'aucune garantie juridique, qu'elle soit effective ou non. Pour autant, les débiteurs des obligations de qualité imposées par la réglementation européenne peuvent y recourir comme alternative, néanmoins plus contraignantes, aux normes européennes harmonisées. Par conséquent, leur élaboration expose d'autant plus à un risque de dénaturation et d'appropriation des exigences essentielles énoncées dans les textes européens élaborés et votés de façon démocratique.

CONCLUSION DU CHAPITRE 1 :

En apparence éloignées des aspects juridique, les preuves techniques occupent en réalité une place essentielle dans la réglementation européenne sur l'IA. Biais algorithmiques, précision, robustesse, des preuves techniques seront sollicitées dans chacune des exigences substantielles du règlement européen. Au sein des normes techniques, elles se révèlent être des outils particulièrement adaptés pour répondre aux besoins de précision et d'articulation, des obligations de fonctionnement des systèmes d'IA à haut risque, que le législateur européen a délégué aux normes techniques. Celui-ci semble avoir pris relativement conscience de l'importance des preuves techniques dans sa stratégie de certification des systèmes d'IA. Le texte mentionne l'intérêt d'utiliser des méthodologies de mesure appropriées pour évaluer la conformité des systèmes aux exigences. Dans le but de lutter contre les biais algorithmiques discriminatoires, le règlement européen autorise en contrepartie, et sous conditions, le recours à des catégories particulières de données personnelles. Mais la place accordée par le législateur européen aux preuves techniques ne semble pas à la hauteur du rôle qu'elles occupent en réalité. Notre étude constate que le choix des métriques d'évaluation s'apparente autant à un travail de définition d'une notion juridique qu'à une tâche informatique. De la même manière, les seuils probabilistes représentent le niveau de risques pour les droits et libertés fondamentaux auxquels l'Union européenne est prête à exposer ses résidents. Au travers de ces seuils, les exigences du règlement européen seront alors quantifiées et leur articulation peut, sans une vigilance accrue, prendre la forme d'une hiérarchisation. Le choix des preuves techniques appropriées, ou du moins leurs modalités de sélection, sera réalisée au sein des normes techniques, en particulier les normes européennes harmonisées. Dès lors, la charge des enjeux et risques juridiques associés aux preuves techniques est confiée aux acteurs qui déterminent le contenu de ces normes, principalement au travers de leur phase d'élaboration. Les différents contrôles, administratif et judiciaires, apparaissent quant à eux peut fiables, voire inefficaces, pour déterminer les preuves techniques appropriées dans le cadre des exigences du règlement européen sur l'IA. Le choix de ces preuves techniques est donc laissé entre les mains des organismes de normalisation européens. Couplé à un contrôle juridique incertain, le manque de représentativité et de légitimité démocratique de ces acteurs laisse redouter un risque d'appropriation des notions et des exigences juridiques. Néanmoins, le contexte d'utilisation des systèmes d'IA à haut risque est un facteur important dans le choix des preuves techniques adéquates. De ce fait, il est possible que l'élaboration des normes techniques ne soit qu'une première étape dans le choix des preuves techniques pertinentes. Les enjeux juridiques qui leur

sont associés incombent alors aux acteurs de terrain que sont les fournisseurs de systèmes d'IA obligés de mettre en conformité leur produits et les acteurs chargés du contrôle de cette même conformité.

CHAPITRE 2 : LE CONTROLE DES PREUVES DE CONCEPTION

Introduction. La production de preuves juridiques n'a de sens que si elle permet d'attester, du moins de convaincre, de la réalité d'un fait, d'une situation ou encore d'un acte juridique auprès d'une personne. Ainsi, les preuves produites par le fournisseur du système d'IA ont pour finalité de démontrer que son produit respecte les différentes exigences applicables contenues le règlement européen sur l'IA. Au regard de notre périmètre de recherche, la conformité juridique désigne alors la satisfaction par les fournisseurs de systèmes d'IA à haut risque des exigences substantielles du règlement européen, à savoir les dispositions relatives à la qualité de fonctionnement des systèmes d'IA. Ensuite, les preuves de cette conformité renvoient aux procédés de preuves et aux éléments de preuves que les fournisseurs de ces logiciels apportent pour démontrer la conformité juridique de leurs produits, en particulier les preuves techniques utilisées. Ainsi, le contrôle de ces preuves désigne les différents mécanismes juridiques existants destinés à évaluer la pertinence de ces preuves au regard des exigences présentes dans les normes techniques et/ou dans le règlement européen sur l'IA. En d'autres termes, la phase de contrôle est l'étape au cours de laquelle les indicateurs et les seuils probabilistes choisis sont acceptés, ou refusés, en tant que preuves juridiques au regard du règlement sur l'IA.

Il est désormais clair que les choix des preuves techniques qui seront considérées comme pertinentes au regard des exigences substantielles de certification des systèmes d'IA à haut risque sont délégués aux organismes de normalisation européens, en particulier au CEN et au CENELEC. Cependant, ces choix sont extrêmement difficiles à réaliser. Parmi les premières normes techniques produites sur les systèmes d'IA, plusieurs d'entre elles ne proposent pas d'indicateurs ou de seuils statistiques à respecter ou même de conseils pour sélectionner ces éléments techniques de preuves, notamment en ce qui concerne le respect des exigences relatives aux biais algorithmiques discriminatoires⁸⁵². À ce jour, l'absence de ces éléments dans les normes techniques européennes ne semble pas être un élément rédhibitoire pour leur publication au Journal officiel de l'UE. Par exemple, le rapport du *Joint Research Centre* de la Commission européenne relatif aux normes provisoires IEEE 7003 sur l'IA recommande aux

⁸⁵² C'est le cas de la version provisoire de la norme technique IEEE P70003/D1 concernant des recommandations sur les biais algorithmiques. JOINT RESEARCH CENTRE, « AI Watch : Artificial Intelligence Standardisation Landscape Update », Commission européenne, 2023. Un autre exemple est celui des lignes directrices sur l'IA de Microsoft : MICROSOFT, « Microsoft Responsible AI Standard, v2 "General Requirements" », 2022.

organismes de normalisation européens de s'inspirer desdits travaux⁸⁵³, malgré l'absence constatée de métriques et de seuils statistiques appropriés⁸⁵⁴. De surcroît, la diversité des secteurs et contextes de déploiement des systèmes d'IA à haut risque accentue la difficulté d'élaborer des preuves techniques pertinentes pour chacune des situations. Dès lors, cette problématique ne pourra pas être réglée intégralement par les normes techniques. Dans de nombreux cas, les fournisseurs de ces systèmes d'IA devront donc sélectionner *in fine* les preuves qu'ils jugent pertinentes, réalisant ainsi un travail d'interprétation des obligations contenues dans le règlement européen⁸⁵⁵. L'obtention de la certification par un des acteurs en charge du contrôle constitue une validation de la pertinence du choix des preuves techniques réalisé par le fournisseur en tant que preuves juridiques. Cette validation, ou refus, des choix probatoires du fournisseur contribue alors aussi à préciser les éléments techniques de preuve effectivement pertinents pour établir la conformité du système d'IA au règlement européen.

Annonce de plan. Il convient donc d'étudier les acteurs impliqués dans le contrôle de la conformité des systèmes d'IA à haut risque et les différentes modalités d'exercice de ce contrôle afin de mettre en lumière la mesure dans laquelle ils peuvent contribuer aux choix des preuves pertinentes de conformité. Les systèmes d'IA à haut risque devant être certifiés préalablement à leur entrée sur le marché, l'obtention de ce certificat européen représente donc la première phase de contrôle des preuves de conception (**Section I**). En outre, les systèmes d'IA doivent fonctionner conformément au règlement européen tout au long de leur cycle de vie. Par conséquent, ils peuvent aussi faire l'objet d'un contrôle après leur commercialisation ou leur mise sur le marché européen (**Section II**).

Section I : Le contrôle ex ante de la conception du système d'IA

Annonce de plan. Pendant toute la durée du processus de conception, le fournisseur d'un système d'IA doit réaliser les obligations qui lui incombent au regard du règlement européen

⁸⁵³ « The IEEE P7003 Draft Standard for Algorithmic Bias Considerations should be considered a relevant source of technical specification for the operationalization of AI Act requirements in relation to bias » - JOINT RESEARCH CENTRE, *op. cit.* P. 4.

⁸⁵⁴ « However, further detail and concrete examples and best practices may be beneficial for AI providers e.g. [...] the selection of suitable metrics and thresholds, Or, the creation of balanced datasets. » - *Ibid.* P. 15.

⁸⁵⁵ Soit parce qu'il choisit de ne pas suivre les spécifications techniques des normes européennes harmonisées, soit parce que lesdites normes ne sélectionnent pas de preuves techniques pertinentes dans le contexte du fournisseur (elles sont alors lacunaires).

sur l'IA à haut risque⁸⁵⁶. La satisfaction des exigences est attestée à l'aide de preuves, qu'ils s'agissent de preuves de la réalisation de processus, de qualité de fonctionnement ou des preuves de documentation. La certification est donc le processus de contrôle *ex ante* des preuves produites par le fournisseur du système d'IA au regard des exigences des normes techniques européennes et/ou des exigences essentielles du règlement. L'octroi du certificat européen signifie que le travail de mise en place des exigences substantielles du règlement, dont le choix et l'obtention de preuves techniques pertinentes, a été réalisé correctement par le fournisseur. Pour réaliser leur mission, les acteurs en charge de la certification doivent interpréter les exigences du règlement européen afin de créer une procédure d'évaluation adéquate. En élaborant ce référentiel, les auteurs de la certification contribuent donc eux aussi aux choix des preuves techniques pertinentes. Au regard des enjeux de cette tâche, il convient d'identifier les acteurs désignés par le règlement pour réaliser la mission de certification des systèmes d'IA à haut risque (**Sous-section 1**) avant d'analyser les modalités de réalisation de ce contrôle *ex ante* de la conception (**Sous-section 2**).

Sous-section 1 : Les acteurs de la certification

Annonce de plan. Le mécanisme de certification des systèmes d'IA à haut risque est un processus complexe constitué de différentes procédures chacune impliquant des acteurs distincts. Par conséquent, il est nécessaire dans un premier temps de cartographier les acteurs désignés par le règlement européen sur l'IA pour la réalisation de ce contrôle *ex ante* (§1). Si les acteurs sont multiples, l'acteur en charge de la certification varie selon les situations. Il convient donc d'étudier la procédure de désignation de l'acteur qui réalisera le contrôle *ex ante* (§2), et qui obtient par conséquent la charge, et le pouvoir, de caractériser la pertinence juridique des preuves techniques soumises.

⁸⁵⁶ De façon évidente, la prise en compte et le respect des obligations du droit européen des systèmes d'IA par les fournisseurs variera selon les fournisseurs.

§1. Les acteurs impliqués dans le contrôle *ex ante*

Annonce de plan. Il est possible de diviser les acteurs impliqués dans le contrôle *ex ante* de la conformité des systèmes d'IA à haut risque en deux catégories, selon le mécanisme de certification à l'œuvre. Tout d'abord, le système d'IA à haut risque peut faire l'objet d'une auto-certification par le fournisseur lui-même **(A)**. Dans ce cas, ce dernier évalue les mesures qu'il a lui-même mises en place pour son système d'IA à haut risque, s'intégrant alors dans la tradition juridique de la *compliance* ou conformité en français. Toutefois, il est aussi possible que la certification soit réalisée par un tiers, transformant alors le processus en relation tripartite entre le législateur, le certificateur et le fournisseur du système d'IA **(B)**.

A- Le fournisseur certificateur

Les origines de l'auto-certification des systèmes d'IA. L'idée de confier au fournisseur du système d'IA la tâche d'évaluer la conformité de son propre système au regard des exigences européennes relève du processus d'auto-certification. Ce choix de l'auto-évaluation de la conformité des systèmes d'IA à haut risque n'est pas une option envisagée dans les travaux préparatoires qui ont inspiré le règlement européen sur l'IA⁸⁵⁷. Dans son livre blanc pour une IA de confiance et d'excellence, la Commission européenne :

« considère à ce stade qu'il serait nécessaire de réaliser une évaluation de la conformité objective et préalable pour vérifier et garantir le respect de certaines des exigences obligatoires susmentionnées pour les applications à haut risque [...] »⁸⁵⁸.

Cette nécessité d'objectivité la conduit à recommander plusieurs solutions dont la certification⁸⁵⁹. À ce titre, le livre blanc prend notamment comme références le règlement (UE)

⁸⁵⁷ Les lignes directrices en matière d'éthique pour une IA digne de confiance proposées par le HLEG sur l'IA ne discute ainsi jamais d'un recours à l'auto-évaluation de la conformité des systèmes d'IA à haut risque par leur propre fournisseur.

⁸⁵⁸ COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020. P.26.

⁸⁵⁹ « L'évaluation préalable de la conformité pourrait inclure des procédures d'essai, d'inspection ou de certification. » - Ibid. PP. 26-27.

2019/881 relatif à l'ENISA et à la certification des TIC⁸⁶⁰. Celui-ci prévoit la possibilité du recours à l'auto-évaluation de la conformité mais seulement pour les éléments présentant un risque faible et donc ne nécessitant qu'une protection de base⁸⁶¹. De même, dans sa partie consacrée à la gouvernance, le livre blanc européen ne mentionne aucunement le recours à l'auto-évaluation de conformité, suggérant au contraire un recours à des tiers⁸⁶². Pour autant, le règlement européen sur l'IA place le fournisseur comme l'un des deux acteurs du contrôle *ex ante* de la conformité des systèmes d'IA à haut risque. Bien que cette option ne soit jamais ni recommandée ni justifiée dans le cadre des systèmes d'IA⁸⁶³, il s'agit du fonctionnement habituel de la logique de redevabilité lorsque le législateur européen a recours aux normes européennes harmonisées⁸⁶⁴.

Le contrôle interne de conformité dans le règlement européen sur l'IA. L'auto-évaluation de la conformité d'un système d'IA à haut risque par son fournisseur est énoncée par le règlement européen sur l'IA au travers de la procédure de contrôle interne⁸⁶⁵. Les fournisseurs remplissent alors deux fonctions distinctes : ils sont, d'une part, en charge d'identifier, d'évaluer et de répondre aux risques élevés que leur système d'IA génère et, d'autre part, ils doivent évaluer si les mesures qu'ils ont eux-mêmes mis en place sont conformes aux exigences juridiques du règlement. Avec cette fonction de contrôleur *ex ante*, le fournisseur du système d'IA à haut risque se retrouve donc à la fois juge et partie⁸⁶⁶.

⁸⁶⁰ « Règlement (UE) 2019/881 du Parlement européen et du Conseil du 17 avril 2019 relatif à l'ENISA (Agence de l'Union européenne pour la cybersécurité) et à la certification de cybersécurité des technologies de l'information et des communications, et abrogeant le règlement (UE) no 526/2013 (règlement sur la cybersécurité) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2019. « Décision n° 768/2008/CE du Parlement européen et du Conseil du 9 juillet 2008 relative à un cadre commun pour la commercialisation des produits et abrogeant la décision 93/465/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008.

⁸⁶¹ « *L'autoévaluation de la conformité n'est autorisée que pour les produits TIC, services TIC et processus TIC qui présentent un risque faible schéma correspondant au niveau d'assurance dit "élémentaire".* » - « Règlement (UE) 2019/881 du Parlement européen et du Conseil du 17 avril 2019 relatif à l'ENISA (Agence de l'Union européenne pour la cybersécurité) et à la certification de cybersécurité des technologies de l'information et des communications, et abrogeant le règlement (UE) no 526/2013 (règlement sur la cybersécurité) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2019. Article 53.

⁸⁶² Voir ci-dessous la partie suivante « B. Le tiers certificateur »

⁸⁶³ L'analyse d'impact de la Commission européenne sur sa proposition de règlement reste étonnamment silencieuse sur les raisons du recours à l'auto-évaluation de conformité pour les systèmes d'IA à haut risque. COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021.

⁸⁶⁴ KOP Mauritz, « EU Artificial Intelligence Act », *Transatlantic Antitrust and IPR Developments*, 2021. PP. 5-6.

⁸⁶⁵ « Règlement sur l'intelligence artificielle », 2024. Article 43.

⁸⁶⁶ Une approche identique est proposée dans la proposition états-unienne de règlement sur les prises de décisions automatisées *Algorithmic Accountability Act (AAA) : Algorithmic Accountability Act*, 3 février 2022. L'analyse d'impact sur la protection des données prévue à l'article 35 du RGPD est aussi un exemple de mécanisme d'auto-évaluation.

Le contrôle interne comme mécanisme insuffisant de protection. Ce choix de contrôle de la conformité est contestable pour plusieurs raisons. Premièrement, le fait de confier ce contrôle *ex ante* de la conformité au fournisseur lui-même est discutable au regard de l'importance que revêt cette étape dans le processus de conformité. Ce contrôle du respect des exigences européennes subordonne la possibilité pour le fournisseur d'apposer le marquage CE sur son produit, condition *sine qua non* pour sa mise en service ou sa mise sur le marché. Pour les résidents européens, il s'agit du dernier mécanisme de protection avant de se voir exposer aux risques élevés que présentent ces systèmes d'IA. L'étape du contrôle *ex post*⁸⁶⁷, quelle que soit son efficacité, n'interviendra qu'après les nombreux préjudices que peut causer un système d'IA à haut risque non conforme au niveau de qualité européen. L'auto-évaluation de la conformité par le fournisseur réduit considérablement la protection juridique au regard des personnes affectées par les futures décisions du système que représente la condition préalable de certification. Cet inconvénient est inhérent au mécanisme d'auto-évaluation lui-même, qui est un processus régulièrement critiqué pour son effectivité limitée⁸⁶⁸. Dans le cas du règlement européen néanmoins, cette lacune est particulièrement problématique car le texte limite déjà ses obligations aux systèmes d'IA à haut risque. Pour reprendre le parallèle avec le récent règlement européen (UE) 2019/881 sur la cybersécurité, ce-dernier justifie son recours au mécanisme d'auto-évaluation de la conformité par les acteurs par le fait qu'il est limité exclusivement et strictement aux situations où les risques sont faibles⁸⁶⁹. La différence entre les deux textes sur le périmètre du recours à ce mécanisme entraîne des conséquences considérables pour les victimes de préjudices du fait de ces logiciels, en plus de complexifier l'articulation entre les textes⁸⁷⁰. La mise en place de ce mécanisme d'auto-évaluation pour les systèmes d'IA à haut risque est largement critiquée par la doctrine pour son niveau de protection clairement insuffisant au regard de la criticité des risques en présence⁸⁷¹. À ce titre, l'abandon de ce

⁸⁶⁷ Voir la section 2 du présent chapitre.

⁸⁶⁸ KOP Mauritz, « EU Artificial Intelligence Act », *Transatlantic Antitrust and IPR Developments*, 2021.

⁸⁶⁹ « Règlement (UE) 2019/881 du Parlement européen et du Conseil du 17 avril 2019 relatif à l'ENISA (Agence de l'Union européenne pour la cybersécurité) et à la certification de cybersécurité des technologies de l'information et des communications, et abrogeant le règlement (UE) no 526/2013 (règlement sur la cybersécurité) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2019.

⁸⁷⁰ Le respect du règlement (UE) 2019/881 offre une présomption de conformité aux exigences relatives à la cybersécurité des systèmes d'IA à haut risque (Article 42 du règlement sur l'IA) mais l'inverse semble très peu probable au regard des différences de protection. Pour une comparaison plus détaillée entre les deux textes, voir CASAROSA Federica, « Cybersecurity certification of Artificial Intelligence », *Int. Cybersecur. Law Rev.*, 3, 2022.

⁸⁷¹ Voir parmi d'autres : Miriam Kullmann & Aude Cefaliello, *The Draft Artificial Intelligence Act (AI Act): Offering False Security to Undermine Fundamental Workers' Rights*, Rochester, NY, 2021; PP. 10-11. Aida Ponce, *The AI Regulation: Entering an AI Regulatory Winter? Why An Ad Hoc Directive on Ai in Employment Is Required*, Rochester, NY, 2021.P.3

mécanisme en faveur d'un recours majoritaire, voire exclusif, à l'évaluation par un tiers est une recommandation unanime, partagée également par des institutions internes à l'Union européenne⁸⁷².

Le contrôle interne comme moyen d'influence du fournisseur sur l'identification des preuves techniques pertinentes. Au travers de ce mécanisme, le fournisseur du système d'IA à haut risque se voit confier l'entière responsabilité de la mise en application des normes juridiques, c'est-à-dire le travail d'adaptation du texte à la pratique. Comme il a été mentionné précédemment, il existe une probabilité significative que les normes européennes harmonisées ne définissent pas, ou seulement en partie, précisément les preuves techniques pertinentes nécessaires pour satisfaire les obligations des systèmes d'IA à haut risque. Par conséquent, c'est le fournisseur qui sera en charge de définir et de satisfaire les métriques et seuils probabilistes adéquats pour la finalité de son produit. La nécessité de mettre en application les exigences du règlement octroie de fait aux fournisseurs de systèmes d'IA une influence considérable sur les choix probatoires en question⁸⁷³. Ce constat n'est pas spécifique à l'auto-évaluation de conformité en droit de l'IA mais est la simple conséquence de l'écart inévitable entre le texte juridique et la pratique. Toutefois le recours à un contrôle interne renforce considérablement cette modalité d'influence des fournisseurs de systèmes d'IA à haut risque. Outre la détermination précise de la nature et de la forme des preuves techniques, les fournisseurs ont la capacité de valider les choix qu'ils ont réalisés lors de cette mise en pratique. Les enjeux juridiques sous-jacents aux preuves techniques, en particulier l'adéquation avec les autres réglementations européennes, sont donc délégués à ces derniers. Cette autonomie d'évaluation de la conformité vient s'ajouter à la place prépondérante que les acteurs privés occupent déjà dans les organismes en charge de l'élaboration des normes européennes harmonisées⁸⁷⁴.

La résolution des difficultés liées au choix adéquat des preuves techniques est une tâche complexe et coûteuse. Les fournisseurs de systèmes d'IA ont d'autres intérêts que la conformité parfaite de leur produit à la réglementation, la rentabilité du système d'IA développé étant l'un

⁸⁷² *Avis conjoint 05/2021 de l'EDPB et du CEPD sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle)*, 2021 PP. 12-13; *Opinion of the European Economic and Social Committee on Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (COM(2021) 206 final — 2021/106 (COD))*, (2021). P.11, recommandation 4.25.

⁸⁷³ Cette influence peut aussi être exercée indirectement par un intermédiaire privé qui viendrait assister et/ou conseiller le fournisseur de systèmes d'IA à haut risque dans son travail de mise en conformité. Pour plus de détails sur le rôle de l'intermédiaire dans l'application d'une réglementation, voir le paragraphe suivant.

⁸⁷⁴ Voir Section II : L'influence des acteurs dans la définition des normes techniques.

d'entre eux. Ainsi, l'élaboration de procédures de tests pertinentes pour un système d'IA, selon ses spécificités technologiques, son contexte d'utilisation et les exigences juridiques applicables, nécessite des ressources considérables de la part du fournisseur. Au regard des coûts économiques, techniques et humains entraînés par la mise en conformité des systèmes d'IA à haut risque⁸⁷⁵, les fournisseurs de ces systèmes d'IA ont intérêt à choisir des solutions qui limitent la quantité de ressources à investir. Ainsi, les enjeux économiques incitent à l'utilisation d'un indicateur déjà existant, en dépit des obstacles disciplinaires et/ou culturels qui pourraient exister⁸⁷⁶. Par exemple, l'indicateur de l'égalité du taux d'erreur, c'est-à-dire la comparaison du taux d'erreur du système d'IA entre deux groupes de population, est plus aisé à satisfaire que l'indicateur de la symétrie du taux d'erreur, qui désigne une égalité de faux positifs et de faux négatifs entre deux catégories de population, car il s'agit d'un indicateur moins détaillé. Cependant, elle est aussi moins pertinente au regard de la notion juridique de discrimination car elle ne permet pas de savoir dans quelle mesure le taux d'erreur mesuré entraîne des conséquences préjudiciables ou bénéfiques pour chaque catégorie d'individus⁸⁷⁷. De la même manière, les métriques de biais discriminatoires les plus populaires, et donc accessibles, sont directement inspirées de notions issues du droit étatsunien (le *disparate treatment* et le *disparate impact*) et non de la notion de discrimination au sens du droit européen. Cet exemple est exacerbé dans la situation où un fournisseur souhaite commercialiser son système d'IA à haut risque au sein de plusieurs marchés, dont le marché européen. À nouveau, il est peu probable que ce fournisseur prenne volontairement l'initiative de tester son produit différemment selon la définition de la discrimination applicable sur chacun des marchés si l'indicateur scientifique correspond à cette définition n'a pas été établi préalablement. Cette hypothèse est d'autant plus probable que le changement d'indicateurs ou de seuils probabilistes peut entraîner un échec lors des tests et donc occasionner le réentraînement partiel du logiciel afin d'améliorer ses performances. Enfin, atteindre un niveau de biais acceptable pour chacune des caractéristiques discriminatoires prohibées par la loi est extrêmement difficile. Dès lors, une possibilité serait de privilégier certaines caractéristiques, comme l'origine ethnique ou le sexe, à d'autres lors de la minimisation des biais dans les données car obtenir un score élevé pour chacune de ces caractéristiques se révèle trop complexe. Par conséquent, il existe une forte

⁸⁷⁵ COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021.

⁸⁷⁶ Pour la présentation des obstacles culturels et disciplinaires dans le choix des métriques, voir §1. La détermination de l'objet de la preuve : l'indicateur.

⁸⁷⁷ CASTETS-RENARD Céline et BESSE Philippe, « Responsabilité ex ante de l'AI Act : Entre certification et normalisation, à la recherche des droits fondamentaux au pays de la conformité », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général ; Perspectives comparées*, Bruylant, 2023.

probabilité que le choix de la métrique et du seuil statistique, pour chaque exigence, soit réalisé en fonction de la preuve technique la moins coûteuse à satisfaire selon la situation d'espèce, et non au regard de sa pertinence pour les exigences européennes.

Au travers des preuves, le fournisseur peut donc choisir la vision de la conformité technique la plus proche de ses besoins pratiques. Par conséquent, l'influence conférée par le mécanisme de l'auto-évaluation aux fournisseurs de systèmes d'IA expose au risque que les pratiques techniques modèlent l'objectif de qualité européenne, représenté par la conformité technique aux exigences essentielles, au lieu que les normes juridiques modèlent les pratiques techniques. Comme de nombreux auteurs et organisations, nous recommandons le recours systématique à un contrôle de conformité *ex ante* par un tiers indépendant⁸⁷⁸. Sans pour autant résoudre tous les problèmes, le recours à un tiers indépendant pour réaliser cette étape de contrôle *ex ante* permet de limiter la capacité du fournisseur de déterminer et valider les preuves techniques considérées comme pertinentes pour satisfaire les exigences du droit européen.

B- Le tiers certificateur

Le besoin d'un tiers certificateur. La certification des systèmes d'IA sur le marché européen a très souvent été envisagée avec la possibilité de recourir à un organisme tiers qui serait en charge de l'évaluation de la qualité desdits logiciels. Dans ses lignes directrices sur l'éthique de l'IA, le groupe d'experts de haut niveau sur l'IA mentionne déjà l'idée « [d'] *organisations qui soient en mesure d'attester auprès du grand public qu'un système d'IA est transparent, responsable et juste.* »⁸⁷⁹. Proposée à l'origine par l'IEEE⁸⁸⁰, cette logique de gouvernance est reprise brièvement dans le livre blanc de la Commission européenne qui propose de déléguer

⁸⁷⁸ Miriam Kullmann & Aude Cefaliello, *The Draft Artificial Intelligence Act (AI Act): Offering False Security to Undermine Fundamental Workers' Rights*, Rochester, NY, 2021; PP. 10-11. Aida Ponce, *The AI Regulation: Entering an AI Regulatory Winter? Why An Ad Hoc Directive on Ai in Employment Is Required*, Rochester, NY, 2021.P.3. *Avis conjoint 05/2021 de l'EDPB et du CEPD sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle)*, 2021 PP. 12-13; *Opinion of the European Economic and Social Committee on Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (COM(2021) 206 final — 2021/106 (COD))*, (2021). P.11, recommandation 4.25.

⁸⁷⁹ GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019.

⁸⁸⁰ Dans ses lignes directrices, le HLEG AI mentionne l'IEEE comme l'un des auteurs de cette idée au travers de leur initiative relative à une conception alignée sur le plan éthique : « The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems », en ligne: *IEEE Normes techniques Association* <<https://normes-techniques.ieee.org/industry-connections/ec/autonomous-systems/>>.

la mission de certification des systèmes d'IA à des organismes notifiés⁸⁸¹. Malgré la forme apparente de suggestion, la Commission motive son idée d'intégrer des intermédiaires dans la gouvernance européenne de l'IA. Selon elle, ces organismes offrent à la fois une assistance pour les autorités nationales compétentes et également une évaluation indépendante et objective des systèmes d'IA, résultant en une plus grande confiance de la part du public dans ces technologies⁸⁸². L'argumentation semble avoir convaincu, au moins partiellement, car le règlement européen sur l'IA contient au sein de son chapitre 4 des dispositions relatives à des organismes notifiés en charge de la certification des systèmes d'IA à haut risque.

Le tiers en charge de la certification. Le règlement européen sur l'IA définit l'organisme d'évaluation de la conformité comme « *un organisme en charge des activités d'évaluation de la conformité par un tiers, y compris la mise à l'essai, la certification et l'inspection ;* »⁸⁸³. Il est donc défini au travers de sa fonction de surveillance de la conformité des systèmes d'IA avant leur entrée sur le marché européen. En d'autres termes, il s'agit d'un tiers au fournisseur qui vérifie que ce-dernier satisfait correctement les exigences imposées par le règlement sur l'IA sur son système d'IA à haut risque⁸⁸⁴. *A contrario*, il n'a pas pour mission de vérifier que le fournisseur respecte ses autres obligations, à savoir la réalisation d'un système de gestion de la qualité⁸⁸⁵, son devoir d'information en cas de risque national⁸⁸⁶ ou encore sa bonne coopération avec les autorités nationales compétentes⁸⁸⁷. La mission énoncée de l'organisme certifié se limite donc à la seule évaluation de la conformité directe du système d'IA, en non de la conformité légale plus globale de son fournisseur et/ou de ses utilisateurs.

Le tiers notifié. Contrairement à ce que peut laisser penser sa définition, l'organisme notifié n'est pas simplement un tiers spécialisé dans l'évaluation de la conformité des systèmes d'IA à haut risque. Il se démarque des organismes de conseils et autres prestataires de services de mise en conformité réglementaire du fait qu'il a été accrédité comme tiers certificateur par une autorité compétente. Ces autorités compétentes sont les autorités notifiantes au sens de l'article 28 du règlement IA⁸⁸⁸. L'autorité notifiante peut être la même entité que l'autorité de contrôle,

⁸⁸¹ COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020. P.29

⁸⁸² *Ibid.*

⁸⁸³ « Règlement sur l'intelligence artificielle », 2024. Article 3(21).

⁸⁸⁴ *Ibid.* Article 3(20).

⁸⁸⁵ *Ibid.* Article 17.

⁸⁸⁶ *Ibid.* Article 20.

⁸⁸⁷ *Ibid.* Article 21.

⁸⁸⁸ « Chaque État membre désigne ou établit au moins une autorité notifiante chargée de mettre en place et d'accomplir les procédures nécessaires à l'évaluation, à la désignation et à la notification des organismes d'évaluation de la conformité et à leur contrôle. [...] » - *Ibid.* Article 28.1.

qui cumule alors la fonction d'accréditation à ses autres missions, ou il peut s'agir de l'organisme national d'accréditation de l'État membre⁸⁸⁹. Dans ce dernier cas, l'organisme national d'accréditation est ce que l'on appelle un « méta-intermédiaire »⁸⁹⁰ car il s'agit d'un intermédiaire en charge de vérifier que les autres intermédiaires sont dignes de confiance dans la réalisation de leur mission d'évaluation de la conformité. Les autorités notifiantes européennes doivent informer la Commission européenne et les États membres à chaque fois qu'elles délivrent une notification à un organisme d'évaluation de la conformité⁸⁹¹. Pour obtenir sa notification, l'organisme d'évaluation de la conformité doit satisfaire certaines conditions listées à l'article 33 du règlement européen sur l'IA. Parmi ces exigences⁸⁹², il doit notamment présenter des garanties de fiabilité au travers de « *sa structure organisationnelle, [sa] répartition des responsabilités, [ses] liens hiérarchiques et [son] fonctionnement [...]* »⁸⁹³. Pour être notifié, l'organisme d'évaluation de la conformité doit aussi démontrer qu'il offre, à la fois au niveau organisationnel et au niveau de ses collaborateurs, des garanties sur l'indépendance, l'impartialité et l'objectivité dans son travail de certification⁸⁹⁴. En cas de recours à un sous-traitant pour l'évaluation de la conformité d'un système d'IA, lesdites obligations lui sont aussi applicables. En revanche, la responsabilité des actions du sous-traitant demeure entièrement assumée par l'organisme notifié⁸⁹⁵. Contrairement aux certificats européens accordés aux systèmes d'IA à haut risque, la notification d'un organisme d'évaluation de la conformité (OEC) ne possède pas de date d'expiration. Aucun contrôle régulier n'est prévu pour s'assurer que l'OEC est toujours digne de confiance dans l'exercice de son activité. Néanmoins, l'absence de prescription ne signifie pas pour autant que celle-ci est irrévocable. L'autorité notifiante peut ainsi réaliser des enquêtes qui, si elles révèlent un manquement de l'OEC aux

⁸⁸⁹ *Ibid.* Article 28.2. Conformément à l'article 4 du règlement (CE) 765/2008, il ne peut y avoir qu'un seul organisme national d'accréditation par État membre. En France, il s'agit du Comité français d'accréditation (Cofrac), une association à but non lucratif en charge de cette mission de service public depuis 1994.

⁸⁹⁰ ABBOTT Kenneth et LEVI-FAUR David, « Introducing Regulatory Intermediaries », *The ANNALS of the American Academy of Political and Social Science*, 670, 2017. P.7. Cet acteur de la gouvernance n'est pas une nouveauté du règlement européen. À titre d'exemple, aux États-Unis, l'agence fédérale des produits alimentaires et médicaux (FDA) désigne des organismes qualifiés pour accréditer les organismes en charge de l'évaluation des médicaments et des aliments avant leur entrée sur le marché états-uniens.

⁸⁹¹ « Règlement sur l'intelligence artificielle », 2024. Article 30.2.

⁸⁹² Le choix des exigences présentées ici se concentre sur les exigences principales mentionnées par la Commission européenne dans son étude d'impact : « *they [notified bodies] are compliant with a set of requirements, notably on independence, competence and absence of conflicts of interests.* » - COMMISSION EUROPÉENNE, « Impact Assessment on Artificial Intelligence Act », 2021. P.13.

⁸⁹³ « Règlement sur l'intelligence artificielle », 2024. Article 31.3.

⁸⁹⁴ *Ibid.* Article 31.6

⁸⁹⁵ *Ibid.* Article 33.

exigences ayant conditionné sa notification, peuvent ensuite aboutir à restreindre la portée de la notification⁸⁹⁶, la suspendre, voire même la retirer.

L'existence de cette prérogative d'enquête accordée à l'autorité notifiante est essentielle pour assurer une certaine efficacité du système de certification européen. Un risque très fréquent dans une gouvernance à trois acteurs (régulateur, intermédiaire et cible de la réglementation)⁸⁹⁷ comme celle du règlement sur l'IA est la perte d'indépendance des intermédiaires qui, dans un contexte concurrentiel, assouplissent les exigences de certification afin de s'assurer plus de clients, les fournisseurs de systèmes d'IA en l'espèce, et donc plus de revenus⁸⁹⁸. Dans le contexte des systèmes d'IA à haut risque, la notification représente pour les OEC l'accès à un marché et les clés de ce marché sont entre les mains de l'autorité notifiante. Par conséquent, le processus d'accréditation mis en place réduit significativement les risques de déviance liés au recours à un intermédiaire dans la gouvernance⁸⁹⁹. Cependant, l'absence de contrôle régulier par les autorités compétentes sur les organismes notifiés relativise l'efficacité du processus d'accréditation dans la lutte contre ces risques⁹⁰⁰.

La capacité de l'organisme notifié à déterminer les preuves techniques pertinentes. Bien que les fournisseurs eux-mêmes aient recours à des prestataires tiers pour les aider à certifier leurs systèmes d'IA à haut risque, le choix de désigner officiellement des intermédiaires et de leur accorder un statut spécifique confère à ces derniers une place et un pouvoir particulier dans la gouvernance du règlement européen sur l'IA. Dans une relation de gouvernance tripartite comme celle-ci, les intermédiaires jouent traditionnellement trois rôles⁹⁰¹.

⁸⁹⁶ *Ibid.* Article 36. Les demandes de notification des OEC comprennent dès le départ un champ d'activité. Ce périmètre d'exercice varie selon les activités d'évaluation de la conformité, les modules d'évaluation de cette conformité et les technologies d'IA pour lesquels l'OEC justifie de sa compétence. C'est ce périmètre d'activité qui peut être restreint par l'autorité notifiante à la suite d'une enquête ayant révélé des manquements.

⁸⁹⁷ ABBOTT Kenneth W., LEVI-FAUR David et SNIDAL Duncan, « Theorizing Regulatory Intermediaries », *The ANNALS of the American Academy of Political and Social Science*, 670, SAGE Publications Inc, 2017.

⁸⁹⁸ GALLAND Jean-Pierre, « Big Third-Party Certifiers and the Construction of Transnational Regulation », *The ANNALS of the American Academy of Political and Social Science*, 670, SAGE Publications Inc, 2017.

⁸⁹⁹ ABBOTT Kenneth et LEVI-FAUR David, « Enriching the RIT framework », *The ANNALS of the American Academy of Political and Social Science*, 670, 2017. P. 11. VAN DER HEIJDEN Jeroen, « Brighter and Darker Sides of Intermediation », *The ANNALS of the American Academy of Political and Social Science*, 670, 2017. Un des exemples de déviance est

⁹⁰⁰ De plus, les conséquences du retrait de la notification sur les systèmes d'IA qui ont déjà été certifiés par ledit organisme notifié ne sont pas envisagées dans les dispositions du règlement européen. Le texte ne mentionne que le transfert des dossiers des systèmes d'IA en cours d'évaluation de conformité (au moment du retrait de la notification) à un autre organisme notifié afin de ne pas pénaliser les fournisseurs concernés.

⁹⁰¹ De manière générale, les intermédiaires occupent une grande diversité de rôle selon les étapes dans lesquelles ils interviennent. Dans le cas des organismes notifiés au sens de la réglementation européenne, cette étape d'intervention se situe après l'adoption du texte législatif. Dans ce contexte, nous résumons leur contribution à trois rôles. Voir ABBOTT Kenneth et LEVI-FAUR David, « Introducing Regulatory Intermediaries », *The ANNALS of the American Academy of Political and Social Science*, 670, 2017. PP. 7-8.

Premièrement, les intermédiaires interviennent significativement dans la mise en application des règles⁹⁰². Que celles-ci soient décrites dans des textes réglementaires ou plus en détails dans des normes techniques, il existe toujours un décalage entre l'énoncé de la règle et son application concrète⁹⁰³. Dans le cas du droit de l'IA, la diversité d'utilisations des systèmes d'IA, la variété de technologies incluses dans la définition du système d'IA ainsi que l'évolution technologique sont à l'origine de cet écart entre le texte et la pratique⁹⁰⁴. Les intermédiaires aident les cibles de la réglementation, en l'espèce les fournisseurs de systèmes d'IA à haut risque, à franchir cet écart. Les exigences d'objectivité et d'absence de conflits d'intérêts empêchent les organismes notifiés de fournir des activités de conseil accessoires à leur activité de contrôle de la conformité⁹⁰⁵. Cependant, la réalisation du contrôle de conformité impliquera nécessairement la production de conseils à destination du fournisseur de systèmes d'IA qui voit son système audité. En effet, si le système d'IA à haut risque s'avère ne pas satisfaire les exigences de certification, il est certain que l'organisme notifié lui fournira des recommandations afin de s'améliorer d'ici la prochaine évaluation, recommandations fondées sur la mise en place des exigences juridique réalisées par l'organisme notifié en question. Ce rôle de mise en application des règles est en l'espèce très lié au second rôle des intermédiaires, à savoir le contrôle de la conformité⁹⁰⁶. En élaborant une méthodologie d'évaluation adaptée aux systèmes d'IA, l'organisme certifié impose à la fois sa transcription des règles à celle réalisée par le fournisseur mais il détermine aussi les preuves techniques acceptables pour obtenir la certification européenne. Si les preuves techniques discutées dans le précédent chapitre ne sont pas déterminées, partiellement ou totalement, par les textes juridiques, l'organisme notifié pourrait avoir la charge de sélectionner les éléments techniques de preuve pouvant être utilisés comme preuves juridiques.

⁹⁰² *Ibid.*

⁹⁰³ Ce constat est également valable si la méthodologie de contrôle de conformité est spécifiée dans les normes techniques. Un standard spécifique à la réalisation du contrôle de conformité a ainsi été commandée par la Commission européenne aux normalisateurs européens (CEN et CENELEC) : *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, *supra* note 424. Annexe I, point 10 « *European standard(s) and/or European normalisation deliverable(s) on conformity assessment for AI systems* »

⁹⁰⁴ Voir A- Le fournisseur certificateur.

⁹⁰⁵ « *Les organismes notifiés sont également indépendants de tout autre opérateur ayant un intérêt économique dans le système d'IA à haut risque qui fait l'objet de l'évaluation, ainsi que de tout concurrent du fournisseur.* » « Règlement sur l'intelligence artificielle », 2024. Article 31.4.

⁹⁰⁶ ABBOTT Kenneth et LEVI-FAUR David, « Introducing Regulatory Intermediaries », *The ANNALS of the American Academy of Political and Social Science*, 670, 2017. P.7.

Enfin, les intermédiaires contribuent à établir une relation et un dialogue de confiance entre les autorités et les entités réglementées⁹⁰⁷. Dans le cas du règlement sur l'IA, la position de confiance leur est attribuée par le texte lui-même au travers du processus de notification mais celle-ci peut s'exprimer au travers d'autres moyens⁹⁰⁸. Au-delà de cette confiance, la nature même des intermédiaires leur octroie un rôle de rapporteurs auprès des autorités sur la réalité du terrain. Du fait de leur proximité avec les fournisseurs de systèmes d'IA, les organismes notifiés seront en effet d'excellents interlocuteurs pour les autorités afin d'identifier les éléments fonctionnels et dysfonctionnels de la réglementation. Ces retours pourront ainsi alimenter des arguments en faveur ou en défaveur de modifications des règles en place⁹⁰⁹, en particulier les éléments pratiques hors de la portée et/ou de l'expertise des autorités. Par conséquent, ces tiers certificateurs sont idéalement placés pour rapporter l'existence de potentiels problèmes comme l'usage fréquent d'une métrique inadéquate ou encore l'existence d'un seuil statistique trop difficile à atteindre pour les fournisseurs dans un secteur particulier. L'ensemble de ces fonctions occupées par les organismes notifiés au sein du règlement sur l'IA leur permet de contribuer à sélectionner les preuves techniques pertinentes et non pertinentes.

Toutefois, l'autorité notifiante joue aussi un rôle dans l'élaboration des procédures de certification, et donc dans la sélection des preuves techniques pertinentes. En effet, l'autorité notifiante produit des documents à destination des organismes notifiés dans le but de les guider dans la réalisation de leurs missions. Ces derniers ne seront donc pas seuls, et donc totalement libres, dans la démarche de validation ou refus des preuves techniques offertes par le fournisseur d'un système d'IA souhaitant obtenir le certificat européen pour son produit. Par exemple, le comité français d'accréditation (COFRAC), qui occupe le rôle d'autorité notifiante dans différents secteurs, élabore notamment des formulaires d'auto-évaluation de la qualité des procédures de certification réalisées par les organismes notifiés⁹¹⁰. Les organismes notifiés doivent décrire leur méthode de certification et justifier qu'elle est appropriée au regard des exigences normatives associées (issues de normes juridiques ou techniques). Il existe alors des normes techniques, sur la manière de réaliser les tests de certification pour ces organismes

⁹⁰⁷ *Ibid.* PP.7-8.

⁹⁰⁸ LOCONTO Allison Marie, « Models of Assurance », *The Annals of the American Academy of Political and Social Science*, 670, [Sage Publications, Inc., American Academy of Political and Social Science], 2017.

⁹⁰⁹ Au travers d'actes délégués, la Commission européenne détient un pouvoir d'évolution significatif du règlement européen sur l'IA.

⁹¹⁰ Voir par exemple, sur le sujet de la métrologie : *Traçabilité des résultats de mesure – Politique du COFRAC et modalités d'évaluation* (GEN REF 10 – Révision 00) – COFRAC, 02/10/2019.

notifiés, produites par les autorités notifiantes⁹¹¹. Il convient donc de nuancer la capacité des organismes notifiés à mettre en avant les preuves techniques qu'ils considèrent comme pertinentes au travers de l'élaboration et de l'application de procédures d'évaluation des systèmes d'IA à haut risque. La documentation produite par les autorités notifiantes contribuent à préciser les éléments contenus dans les normes techniques, notamment européennes. Par conséquent, elles contribueront elles aussi, dans une certaine mesure, à déterminer les indicateurs et les seuils probabilistes les plus pertinents pour prouver la conformité du fonctionnement des systèmes d'IA à haut risque aux exigences du règlement européen sur l'IA.

Les intérêts des tiers certificateurs. Les capacités de détermination des preuves techniques pertinentes juridiquement sont la conséquence logique de la fonction confiée aux organismes notifiés et ne dépend pas de l'identité du tiers certificateur en question. En revanche, l'identité du tiers certificateur impacte ses intérêts personnels, qui peuvent influencer son interprétation des exigences dont il est chargé de vérifier le respect. Comme nous le verrons plus en détails dans le paragraphe suivant, l'identité du tiers certificateur varie selon le contexte d'utilisation auquel est destiné le système d'IA à haut risque. Pour la présente réflexion, nous nous concentrerons autour des situations où le tiers certificateur est une personne privée et non une autorité publique. Bien qu'ils soient désignés et évalués par une autorité publique (l'autorité notifiante), il est essentiel de garder à l'esprit que ces organismes notifiés poursuivent leurs propres intérêts privés⁹¹². Par exemple, un intermédiaire peut influencer la réglementation dans une direction encore plus complexe et standardisée afin de s'assurer une pérennité, une mainmise et/ou un accroissement de ses activités de certification dont la complexité laisserait encore plus à l'écart de nouveaux concurrents ou conduirait à un contrôle plus précis des

⁹¹¹ L'*European accreditation*, association hollandaise à but non-lucratif, est l'infrastructure européenne en charge de l'accréditation au sens du droit européen : « Règlement (CE) n° 765/2008 du Parlement européen et du Conseil du 9 juillet 2008 fixant les prescriptions relatives à l'accréditation et à la surveillance du marché pour la commercialisation des produits et abrogeant le règlement (CEE) n° 339/93 du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008. L'association produit des normes non contraignantes venant préciser des détails techniques relatifs aux exigences imposées sur les organismes notifiés. Voir, à titre d'exemple : EA-4/02 M: 2022 Evaluation of the Uncertainty of Measurement in calibration (04/04/2022).

⁹¹² ABBOTT Kenneth et LEVI-FAUR David, « Introducing Regulatory Intermediaries », *The ANNALS of the American Academy of Political and Social Science*, 670, 2017. Ces intérêts peuvent être institutionnels (rémunération, influence parmi les acteurs de la gouvernance par exemple) et/ou substantiels (modification du contenu de la réglementation pour l'aligner en accord avec leurs propres valeurs).

autorités de régulation⁹¹³. Ces intérêts divers peuvent se trouver aussi bien du côté de l'autorité de régulation ou que du côté des acteurs réglementés. En tant que contrôleurs de la conformité *ex ante* de certains systèmes d'IA à haut risque, les organismes notifiés ont la capacité de faire valoir leur vision de la mise en application des exigences contenues dans le règlement européen sur l'IA et/ou dans les normes techniques associées, en particulier par le biais de la validation ou le refus des preuves techniques issues des tests réalisés par le fournisseur sur son logiciel. À l'image des auteurs de textes éthiques, les organismes notifiés ont des intérêts propres qui peuvent les conduire à interpréter les exigences réglementaires de façon à servir lesdits intérêts.

§2. L'auteur choisi pour le contrôle *ex ante*

Annonce du plan. Il est désormais clair que le contrôleur *ex ante* d'un système d'IA à haut risque, qu'il soit le fournisseur lui-même ou un tiers, détient un pouvoir significatif sur la sélection des preuves techniques pertinentes juridiquement. Toutefois, l'intensité de leur influence dépend en grande partie de la fréquence à laquelle ces acteurs sont amenés à pouvoir l'exercer. En d'autres termes, plus un acteur est amené à contrôler *a priori* un système d'IA à haut risque, plus il contribue à déterminer, en moyenne, les preuves techniques pertinentes. Plus généralement, ce sont les conditions dans lesquelles un acteur est désigné pour certifier lesdits systèmes qui déterminent l'importance de son pouvoir sur ces choix probatoires. Ces conditions de réalisation du contrôle *ex ante* sont déterminées au travers des règles de désignation du certificateur. Le règlement européen sur l'IA présente non pas une mais de nombreuses procédures de choix de l'acteur de l'évaluation de la conformité *ex ante*. Ces procédures peuvent être regroupées en deux catégories. La première contient les procédures d'évaluation de conformité issues des réglementations sectorielles auxquelles le règlement sur l'IA renvoie **(A)** tandis que la seconde est la procédure de désignation élaborée spécifiquement pour les systèmes d'IA commercialisés ou mis en service de façon autonome **(B)**.

⁹¹³ L'arrêt *James Elliott* de 2016 (voir Sous-section 2 : Les acteurs en charge du contrôle des normes techniques) a soulevé de nombreuses critiques quant à la compétence des juges de la CJUE à se prononcer sur les normes techniques. Parmi les auteurs de ces critiques se trouvaient les organismes européens de normalisations (CEN et CENELEC en particulier), qui sont des intermédiaires dans la phase d'élaboration de la loi. Ils proposaient en revanche de mettre en place un service d'interprétation de leurs normes techniques dans le but de garantir leur interprétation correcte : « CEN and CENELEC position on the consequences of the judgment of the European Court of Justice on *James Elliott Construction Limited v Irish Asphalt Limited* », 17 mai 2017, p. 6, disponible à l'adresse suivante : <https://opil.ouplaw.com/view/10.1093/law-oxio/e246.013.1/law-oxio-e246-document-1.pdf> . La complexité des normes techniques est donc un argument permettant à leur auteurs de plaider pour une élargissement de leur place en tant qu'acteurs dans la réglementation européenne.

A- La désignation de l'acteur par la réglementation sectorielle européenne

Le système d'IA comme composant d'un produit réglementé. Le système d'IA n'est pas voué à toujours être commercialisé ou mis en service comme un produit à part entière. L'annexe II du règlement européen liste les produits au sein desquels l'utilisation d'un système d'IA comme composant représente un risque élevé. L'approche transversale du règlement européen sur l'IA vient superposer ses obligations spécifiques aux systèmes d'IA à celles déjà existantes dans la conformité de ces produits. Conscient du besoin d'articulation entre les textes, le législateur européen fait le choix de ne pas modifier les procédures de certification et/ou de mise en conformité applicables. Par conséquent, lorsque le système d'IA est un composant d'un produit déjà soumis à un mécanisme de certification propre, l'auteur du contrôle de conformité du produit global est également en charge d'évaluer la conformité du système d'IA qui le compose⁹¹⁴. En d'autres termes, la procédure d'évaluation de la conformité du composant qu'est le système d'IA vient s'aligner à celle du produit dans lequel il est intégré. Dans le cas où la réglementation du produit exige le recours à un tiers pour le contrôle *ex ante* de conformité, ce-dernier est automatiquement habilité à évaluer la conformité aux exigences relatives aux systèmes d'IA dès lors que sa notification garantit sa compétence, son intégrité professionnelle et son indépendance⁹¹⁵.

Les conséquences complexes sur l'influence des acteurs sur les preuves techniques. L'analyse des actes législatifs listés à l'annexe II section A du règlement européen sur l'IA révèle sans surprise que chacun des produits concernés par les textes font l'objet d'une évaluation de conformité préalablement à leur mise en service ou leur commercialisation. Dans chacun des douze textes étudiés, les deux évaluateurs potentiels, le fabricant et le tiers notifié, sont présents dans les procédures d'évaluation *ex ante* mises en place. Les critères de choix d'une procédure, à savoir le contrôle interne, plutôt qu'une autre, le contrôle externe par un tiers, varient selon les textes. Il peut s'agir de l'application des normes européennes

⁹¹⁴ « Règlement sur l'intelligence artificielle », 2024. Article 43.3 alinéa 1.

⁹¹⁵ *Ibid.* Article 43.3 alinéa 2

harmonisées⁹¹⁶, de la gravité des risques présentés par certaines sous-catégories de produits⁹¹⁷ ou encore de la nature ou l'importance de l'élément dans le produit final⁹¹⁸. Sans une étude très approfondie de chacun de ces textes⁹¹⁹, il n'est pas possible d'estimer la fréquence du recours à une auto-évaluation ou évaluation externe pour l'obtention du certificat européen⁹²⁰. Par conséquent, il n'est pas non plus possible de savoir si l'alignement du règlement européen des systèmes d'IA sur les procédures d'évaluation de conformité en question accroît le recours à un tiers notifié ou à la procédure de contrôle interne. Dès lors, nous ne pouvons tirer aucune conclusion des conséquences que cette règle procédurale spécifique entraîne sur l'influence de chaque acteur sur la détermination des preuves techniques pertinentes.

B- La désignation de l'acteur par le règlement européen sur l'IA

L'application des normes européennes harmonisées. Lorsque le fournisseur a appliqué les normes techniques européennes, qu'ils s'agissent des normes européennes harmonisées ou des spécifications techniques de la Commission européenne, le fournisseur d'un système d'IA à haut risque a toujours l'opportunité de réaliser lui-même son contrôle de conformité *ex ante* en interne. Dans la quasi-totalité des situations⁹²¹, le choix du recours à un contrôle interne est imposé aux fournisseurs dès lors qu'ils ont suivi les exigences des normes techniques européennes⁹²². L'unique exception à ce principe est l'utilisation d'un système d'IA pour *« l'identification biométrique à distance "en temps réel" et "a posteriori" des personnes*

⁹¹⁶ La directive 2006/42/CE du 17 mai 2006 relative aux machines et la directive 2009/48/CE relative à la sécurité des jouets permettent le contrôle interne dès lors que le fabricant a respecté les normes européennes harmonisées pertinentes. Il s'agit également du choix réalisé par le législateur européen dans le cas des systèmes d'IA à haut risque. Voir paragraphe suivant.

⁹¹⁷ « Règlement (UE) 2016/425 du Parlement européen et du Conseil du 9 mars 2016 relatif aux équipements de protection individuelle et abrogeant la directive 89/686/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 19.

⁹¹⁸ Voir notamment : « Directive 2014/68/UE du Parlement européen et du Conseil du 15 mai 2014 relative à l'harmonisation des législations des États membres concernant la mise à disposition sur le marché des équipements sous pression (refonte) Texte présentant de l'intérêt pour l'EEE », OJ L, 2014. « Directive 2013/53/UE du Parlement européen et du Conseil du 20 novembre 2013 relative aux bateaux de plaisance et aux véhicules nautiques à moteur et abrogeant la directive 94/25/CE Texte présentant de l'intérêt pour l'EEE », OJ L, 2013.

⁹¹⁹ Cette étude approfondie de chacun de ces textes se situe hors du cadre de recherche de la présente thèse.

⁹²⁰ À cette difficulté s'ajoute les situations intermédiaires où les deux acteurs interviennent sur un même produit mais sur différents éléments et/ou sur différentes exigences.

⁹²¹ Sur les vingt-et-une utilisations d'un système d'IA qualifié à haut risque par le législateur européen, vingt d'entre elles se voient imposer un contrôle interne de conformité si les fournisseurs ont suivi les normes européennes harmonisées.

⁹²² « Pour les systèmes d'IA à haut risque visés à l'annexe III, points 2 à 8, les fournisseurs suivent la procédure d'évaluation de la conformité fondée sur le contrôle interne visée à l'annexe VI, qui ne prévoit pas d'intervention d'un organisme notifié. » - « Règlement sur l'intelligence artificielle », 2024. Article 43.2.

physiques. »⁹²³. Dans ce cas précis, le fournisseur qui a suivi les normes techniques européennes a le choix entre réaliser le contrôle de conformité *ex ante* lui-même ou recourir à un organisme notifié⁹²⁴. Au regard des intérêts, notamment économiques, des fournisseurs de systèmes d'IA, cette liberté de choix de l'acteur réalisant le contrôle *ex ante* aboutira très probablement à l'utilisation du contrôle interne.

Le tiers notifié ne sera donc très vraisemblablement jamais impliqué dans la vérification du respect des normes européennes techniques par le fournisseur. La présomption de conformité et les coûts réduits entraînés par le respect des normes techniques européennes sont deux facteurs incitatifs qui conduiront la plupart des fournisseurs de systèmes d'IA à les utiliser comme des textes de référence. Par conséquent, une fois les normes techniques européennes adoptées, on peut supposer que la quasi-totalité des marquages européens sur les systèmes d'IA à haut risque seront réalisés à la suite d'un contrôle interne de conformité⁹²⁵.

La fréquence et l'intensité de la contribution du fournisseur évaluateur à la sélection des preuves juridiques. Ce choix procédural dans le règlement européen entraîne deux conséquences sur l'influence que détient le fournisseur du système d'IA à haut risque sur le choix des preuves techniques pertinentes en tant que preuves juridiques. Au préalable, il faut signaler que le fournisseur peut bénéficier d'une présomption de conformité aux exigences essentielles de conception du règlement. Sans intervention d'un tiers indépendant, organisme notifié ou autorité de contrôle, la charge de la preuve de conformité réelle du système d'IA à haut risque au droit européen repose sur la phase de contrôle *ex post*. Ce sera aux acteurs de cette étape de contrôler que le système d'IA à haut risque respectait, et respecte toujours, effectivement les exigences prévues par le texte européen.

Désormais il apparaît que le fournisseur sera probablement un acteur du contrôle *ex ante* de conformité bien plus fréquent que le tiers notifié. Au regard de cette hypothèse, cela signifie

⁹²³ « Règlement sur l'intelligence artificielle », 2024. Annexe III.1.

⁹²⁴ « Pour les systèmes d'IA à haut risque énumérés à l'annexe III, point 1, lorsque, pour démontrer la conformité d'un système d'IA à haut risque avec les exigences énoncées au chapitre 2 du présent titre, le fournisseur a appliqué les normes harmonisées visées à l'article 40 ou, le cas échéant, les spécifications communes visées à l'article 41, il suit l'une des procédures suivantes: (a) la procédure d'évaluation de la conformité fondée sur le contrôle interne visée à l'annexe VI; (b) la procédure d'évaluation de la conformité fondée sur l'évaluation du système de gestion de la qualité et l'évaluation de la documentation technique, avec l'intervention d'un organisme notifié, visée à l'annexe VII. » - Article 43.1.

⁹²⁵ CASTETS-RENARD Céline et BESSE Philippe, « Responsabilité ex ante de l'AI Act : Entre certification et normalisation, à la recherche des droits fondamentaux au pays de la conformité », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général; Perspectives comparées*, Bruylant, 2023. Ce postulat sur lequel se fonde la suite de notre analyse devra néanmoins être vérifié et confirmé (ou infirmé) lors de la mise en œuvre du règlement.

que les fournisseurs des systèmes d'IA seront de manière globale, des contributeurs importants aux choix des preuves pertinentes. Sur le plan individuel en revanche, l'influence des fournisseurs de systèmes d'IA est relativement limitée. Leur contribution dans le choix des preuves adaptées sera réduite, de façon plus ou moins importante, par les précisions contenues dans les normes techniques. Plus les organismes de normalisation seront précis concernant le choix des preuves visant à garantir le respect des exigences de fonctionnement de ces logiciels d'IA, moins la liberté d'interprétation des fournisseurs sera importante. Néanmoins, le fournisseur doit attester que son système d'IA à haut risque satisfait les exigences substantielles du règlement sur l'IA. Par conséquent, si les organismes de normalisation européens ne règlent pas, ou pas suffisamment, le problème du choix des éléments techniques de preuves pertinents juridiquement, l'influence des fournisseurs de systèmes d'IA à haut risque sur cette question deviendra d'autant plus importante.

La non-application des normes européennes harmonisées. Malgré la prédominance considérable du recours au contrôle interne, il existe toujours la possibilité que les fournisseurs de systèmes d'IA à haut risque n'appliquent pas les normes techniques européennes. Cette non-application peut être volontaire ou contrainte. Dans le premier cas, le fournisseur de systèmes d'IA à haut risque décide de son propre chef de ne pas utiliser les normes européennes harmonisées, ou le cas échéant les spécifications communes de la Commission, comme textes de référence pour la mise en conformité de son système. Ce refus peut être total, auquel cas le fournisseur ignore la totalité du standard européen, ou partiel, auquel cas il ne choisit d'appliquer que certaines des instructions contenues dans le standard européen. Pour rappel, la CJUE considère que le choix de ne pas recourir aux normes européennes harmonisées est tellement contraignant pour les opérateurs économiques que ces normes en deviennent *de facto* « *manifestement obligatoires* »⁹²⁶.

Dans le second cas, le fournisseur du système d'IA n'applique pas le standard européen parce que ce-dernier n'existe pas ou plus. L'échéance accordée par la Commission européenne aux organismes de normalisation européens (CEN et CENELEC) pour la production des normes européennes harmonisées est en 2025. Dans le scénario où le règlement européen sur l'IA entrerait en vigueur avant cette date⁹²⁷, les premiers fournisseurs de systèmes d'IA à haut risque seraient dans l'impossibilité de suivre lesdites normes pour attester de leur conformité aux

⁹²⁶ Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P.

⁹²⁷ De façon plus générale, il s'agit du cas où le texte entrerait en vigueur avant la publication des normes techniques du CEN et du CENELEC dans le Journal officiel de l'Union européenne.

exigences essentielles du règlement. Également, dans le cas, certes très improbable⁹²⁸, où une norme européenne harmonisée viendrait à être annulée, les fournisseurs de systèmes d'IA concernés par ladite norme devraient suivre d'autres normes techniques (ou leur propre standard) s'ils souhaitent pouvoir certifier leur système avant l'actualisation du standard européen.

La fréquence et l'intensité de l'influence des organismes notifiés. Compte tenu des conditions de désignation de l'organisme notifié comme évaluateur de la conformité d'un système d'IA à haut risque, il occupera probablement moins ce rôle que le fournisseur lui-même. Cette très faible fréquence de recours à un tiers évaluateur est une volonté assumée de la part de la Commission européenne qui considère que les organismes notifiés spécifiquement pour les systèmes d'IA à haut risque ont une expérience moins étendue que les autres professionnels de la certification déjà en place⁹²⁹. Cependant, les organismes notifiés bénéficient d'une marge d'interprétation du texte bien plus importante que les fournisseurs en tant que contrôleurs *ex ante* de la conformité. La procédure de désignation limite leur intervention aux situations où il n'y a pas de normes techniques européennes. Dès lors, les organismes notifiés ont la charge de vérifier que les normes techniques suivies par le fournisseur de systèmes d'IA, ou plus généralement les mesures qu'il a mises en place, sont conformes avec les exigences essentielles contenues dans le règlement européen. En d'autres termes, les organismes notifiés doivent réaliser le même travail d'interprétation, voire de façon encore plus précise, que celui réalisé par les organisations de normalisation afin de déterminer si la mise en place des obligations par le fournisseur du système d'IA à haut risque est conforme aux exigences de protection de la santé, de la sécurité et des droits fondamentaux prévues par le règlement. Dans cette tâche, ils seront cependant assistés et guidés par les normes produites par les autorités notifiantes nationales et européenne. Néanmoins, les organismes notifiés conservent une capacité significative de contribution aux choix des preuves techniques pertinentes.

⁹²⁸ Pour plus de détails sur le contrôle des normes européennes harmonisées, voir Sous-section 2 : Les acteurs en charge du contrôle des normes techniques.

⁹²⁹ « Étant donné l'expérience plus étendue des organismes professionnels de certification avant mise sur le marché dans le domaine de la sécurité des produits et de la nature différente des risques encourus, il convient de limiter, au moins dans une phase initiale d'application du présent règlement, le champ d'application des évaluations de la conformité réalisées par un tiers aux systèmes d'IA à haut risque autres que ceux liés à des produits. » « Règlement sur l'intelligence artificielle », 2024. Considérant 63.

Sous-section 2 : Le processus de contrôle ex ante de conformité

Annonce de plan. En tant qu'évaluateur de la conformité *ex ante* des systèmes d'IA à haut risque, le fournisseur du système d'IA et l'organisme notifié ont tous les deux la capacité de choisir les preuves considérées comme adéquates pour attester du respect des exigences contenues dans le règlement européen sur l'IA. L'analyse réalisée précédemment sur la nature de ces acteurs, de leur fonction et des conditions de leur intervention dans la conformité des systèmes d'IA à haut risque ne révèle cependant que l'aspect formel de leur influence. Pour révéler l'aspect substantiel de leur contribution à la détermination des preuves techniques pertinentes (§2), il faut s'intéresser au périmètre du contrôle *ex ante* de conformité (§1).

§1. Le périmètre des procédures de contrôle *ex ante*

Limites de l'analyse et détermination du périmètre des procédures de contrôle *ex ante*. La présente étude est réalisée avant l'élaboration et la publication des normes européennes harmonisées. En particulier, l'une des normes techniques commandées par la Commission européenne au CEN et CENELEC porte sur le système de gestion de qualité, ainsi que le système de surveillance post-commercialisation⁹³⁰. L'analyse de ces éléments constituant le périmètre du contrôle *ex ante* de la conception des systèmes d'IA est donc réalisée sans les précisions qui seront contenues dans les normes européennes harmonisées ou, le cas échéant, dans les spécifications communes produites par la Commission européenne.

Les dispositions relatives au contrôle interne de conformité réalisé par le fournisseur du système d'IA à haut risque ne sont pas les mêmes que pour le contrôle externe réalisé par l'organisme notifié⁹³¹. Malgré cette différence de textes, les périmètres de réalisation des procédures sont semblables en se concentrant sur les trois mêmes éléments produits par le fournisseur : le système de gestion de qualité, la documentation technique et le système de surveillance après commercialisation.

⁹³⁰ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, supra note 424. Annexe I.10.

⁹³¹ La procédure de contrôle interne réalisé par le fournisseur du système d'IA à haut risque est décrite à l'annexe VI du règlement européen sur l'IA tandis que la procédure du contrôle réalisé par l'organisme notifié est expliquée à l'annexe VII dudit règlement.

Le système de gestion de qualité. L'article 17 du règlement européen sur l'IA ne fournit pas de définition explicite d'un système de gestion de qualité mais le décrit comme un ensemble de politiques, procédures et instructions écrites documentées de façon ordonnée et méthodique⁹³². Le fournisseur d'un système d'IA à haut risque se voit donc imposer la mise en place de ce système de gestion de la qualité dont le contenu est exposé, de façon non exhaustive⁹³³, sous forme de liste par le législateur européen. Sans entrer dans le détail de chaque élément de la liste inscrite à l'article 17.1.a) à m), le document qu'est le système de gestion de qualité contient toutes les preuves documentaires rattachées aux obligations processuelles des fournisseurs de systèmes d'IA. Il contient notamment le système de gestion de risques (au travers duquel le fournisseur d'un système d'IA à haut risque détermine le niveau de risques résiduels qu'il considère comme acceptable⁹³⁴), les normes techniques appliquées par le fournisseur ainsi que les procédures de collaboration avec les autorités nationales. L'obligation de mise en place d'un système de gestion de qualité impose donc au fournisseur du système d'IA à haut risque de mettre en forme les procédures qu'il applique tout au long de la conception de son ou de ses logiciels afin de faciliter le contrôle de leur existence lors du contrôle *ex ante* mais également *ex post* de conformité.

La documentation technique. De la même manière, la documentation technique est présentée au travers d'une liste non exhaustive d'éléments énoncée à l'annexe IV du règlement sur l'IA. Elle a pour fonction explicite de :

*« démontrer que le système d'IA à haut risque satisfait aux exigences énoncées dans le présent chapitre et à fournir aux autorités nationales compétentes et aux organismes notifiés toutes les informations nécessaires pour évaluer la conformité du système d'IA avec ces exigences »*⁹³⁵

Contrairement au système de gestion de la qualité, la documentation technique est unique à chaque système d'IA à haut risque. De ce fait, malgré certaines redondances avec le système de gestion de la qualité⁹³⁶, la documentation technique contient des informations plus individualisées et détaillées sur le système d'IA auquel elle est rattachée. L'on retrouve ainsi non seulement des indications précises sur les procédures de tests mais également les résultats

⁹³² « Ce système est documenté de manière méthodique et ordonnée sous la forme de politiques, de procédures et d'instructions écrites [...] » - « Règlement sur l'intelligence artificielle », 2024. Article 17.1.

⁹³³ Outre le caractère non exhaustif de la liste fourni par le texte, l'étendue de l'obligation de mise en place desdits éléments varie proportionnellement « à la taille de l'organisation du fournisseur » - *Ibid.* Article 17.2.

⁹³⁴ Voir Section I : L'importance des preuves au sein des normes techniques.

⁹³⁵ « Règlement sur l'intelligence artificielle », 2024. Article 11.1 paragraphe 2.

⁹³⁶ La documentation technique doit notamment contenir « une description détaillée du système de gestion des risques » ainsi que les normes techniques appliqués par le fournisseur.

à ces tests. La documentation technique doit notamment contenir « *le niveau global d'exactitude prévu par rapport à la destination du système* »⁹³⁷ ainsi que « *le degré d'exactitude pour des personnes ou des groupes de personnes spécifiques* »⁹³⁸. Ce sont des informations capitales pour identifier un potentiel fonctionnement biaisé du logiciel. En outre, le fournisseur doit indiquer les métriques utilisées pour évaluer le respect des exigences substantielles imposées aux systèmes d'IA à haut risque⁹³⁹.

Le système de surveillance après la commercialisation du système d'IA. Le système de gestion de la qualité et la documentation technique sont des éléments de garantie de la conformité du système d'IA à haut risque aux exigences du règlement européen. Le système de surveillance est un mécanisme complémentaire à ces derniers. Il a pour fonction de vérifier que la conformité de départ du système d'IA est maintenue tout au long de son utilisation⁹⁴⁰. Bien que faisant partie de la documentation technique, le système de surveillance se voit accorder une attention toute particulière dans les procédures de contrôle *ex ante*. Il repose sur un plan de surveillance après commercialisation. Ce document n'est pas décrit ni dans le corps du règlement ni dans les annexes. Pour autant, l'élaboration de contenu précis n'est pas déléguée aux organismes de normalisation européens au travers des normes européennes harmonisées⁹⁴¹. Elle ne sera pas non plus confiée aux autorités compétentes car la Commission européenne a explicitement conservé la prérogative de son élaboration⁹⁴². Contrairement aux deux documents précédemment traités, ce système de surveillance se concentre exclusivement sur les performances de fonctionnement du système d'IA⁹⁴³. De façon très intéressante, si le contrôle *ex ante* de la conformité a été réalisé par un organisme notifié, la surveillance du fonctionnement du système d'IA pendant son utilisation n'est pas laissée à la seule charge du fournisseur. L'organisme notifié doit ainsi vérifier que le niveau de conformité du système d'IA qu'il a approuvé avant la commercialisation est maintenu pendant son utilisation. En revanche,

⁹³⁷ « Règlement sur l'intelligence artificielle », 2024. Annexe IV.3.

⁹³⁸ *Ibid.*

⁹³⁹ « *les procédures de validation et de test utilisées, y compris les informations sur les données de validation et de test utilisées et leurs principales caractéristiques; les paramètres utilisés pour mesurer l'exactitude, la robustesse, la cybersécurité et le respect des autres exigences pertinentes énoncées au titre III, chapitre 2* »

⁹⁴⁰ « Règlement sur l'intelligence artificielle », 2024. Article 61.2.

⁹⁴¹ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, supra note 424.

⁹⁴² « *La Commission adopte un acte d'exécution fixant des dispositions détaillées établissant un modèle pour le plan de surveillance après commercialisation et la liste des éléments à inclure dans le plan.* » - « Règlement sur l'intelligence artificielle », 2024. Article 61.3. Une exception est cependant faite lorsque le système d'IA est utilisé comme composant d'un produit faisant lui aussi l'objet d'un système de surveillance. Comme pour la désignation de l'acteur du contrôle de conformité, c'est le système de surveillance du produit global qui prime, tout en étant étoffé des contraintes spécifiques au système d'IA qui le compose.

⁹⁴³ *Ibid.* Article 61.2.

si le contrôle *ex ante* de conformité a été réalisé par le fournisseur lui-même, il vérifie seul la continuité de cette conformité. Par conséquent, l'acteur qui a qualifié la qualité d'un système d'IA comme conforme aux exigences juridiques est aussi celui chargé de la surveillance du maintien de cette qualité. Si les performances du système d'IA varient au cours de son utilisation, il est en charge d'évaluer à nouveau si les éléments de preuve techniques, à savoir les résultats des tests de conformité, sont conformes à son interprétation des exigences européennes. Dès lors, l'influence du contrôleur *ex ante* de la conformité sur le système probatoire dépasse la seule phase de conception car son interprétation de la conformité juridique est maintenue, dans une certaine mesure⁹⁴⁴, pendant tout le cycle de vie du système d'IA à haut risque.

§2. L'influence substantielle du contrôle *ex ante* sur le système probatoire

Annonce de plan. La procédure exacte de contrôle *ex ante* de la conformité est actuellement en cours d'élaboration par les organismes européens de normalisation⁹⁴⁵. Toutefois, les éléments de preuve soumis au contrôle *ex ante* permettent d'estimer le pouvoir substantiel conféré aux acteurs de ce contrôle sur la détermination des éléments techniques de preuves pertinents en droit de l'IA. À première vue, les acteurs de l'évaluation *ex ante* des systèmes d'IA à haut risque semblent être limités à un contrôle des preuves documentaires **(A)**. Cependant, les prérogatives accordées à ces acteurs par le règlement sur l'IA leur rendent aussi possible l'accès et le contrôle effectif de la réalité des éléments techniques de preuves de conception produites par le fournisseur **(B)**.

⁹⁴⁴ L'existence d'un contrôle *ex post* de la conformité du système d'IA à haut risque avec des acteurs différents vient nuancer ce constat.

⁹⁴⁵ *Draft normalisation request to the European Normalisation Organisations in support of safe and trustworthy artificial intelligence*, *supra* note 424. Annexe I.10. L'élaboration de ce standard particulier renforce d'autant plus le pouvoir de ces organismes sur le système probatoire des systèmes d'IA car il leur permet d'étendre leur influence à la phase de contrôle de conformité *ex ante*.

A- Le contrôle des preuves documentaires

Le contrôle du système de gestion de qualité est exclusivement un contrôle documentaire. En effet, en ce qui concerne le contrôle interne de conformité, l'annexe VI.2. du règlement énonce que : « *Le fournisseur vérifie que le système de gestion de la qualité établi est conforme aux exigences de l'article 17.* ». De la même manière, l'annexe VII relative au contrôle par un organisme notifié est rédigée presque identiquement sur ce point : « *Le système de gestion de la qualité est évalué par l'organisme notifié, qui détermine s'il satisfait aux exigences visées à l'article 17.* ». Il s'agit des deux seules dispositions concernant le déroulé du contrôle du système de gestion de qualité. Le législateur européen n'apporte ainsi aucune précision ou indication supplémentaire quant au déroulé du contrôle de cet élément de preuve. Aucune indication permettant un contrôle de la réalité des éléments de preuves techniques annoncés par le fournisseur n'est mentionnée aussi bien dans la procédure de contrôle externe que pour la procédure interne. Or, le système de gestion de qualité est une liste de « *politiques, de procédures et d'instructions écrites*⁹⁴⁶ ». Par conséquent, le rôle du contrôle *ex ante* est limitée à une simple comparaison des éléments listés dans le document par le fournisseur avec la liste d'éléments énoncés par le législateur européen. Néanmoins, cette limitation forte du contrôle du système de gestion de la qualité peut s'expliquer par deux aspects. D'une part, le contrôle *ex ante* se limite à évaluer la conformité d'un produit qui est un système d'IA à haut risque. Or, le système de gestion de qualité est un document qui concerne potentiellement une multitude, voire la totalité, des systèmes d'IA à haut risque produits et à produire par le fournisseur de systèmes d'IA. Par conséquent, le système de gestion de la qualité est un document qui dépasse en partie le périmètre d'action du contrôleur dont la mission est de contrôler la conformité du système d'IA à haut risque. D'autre part, le système de gestion de qualité contient des éléments redondants avec les autres documents soumis à l'évaluation du contrôleur *ex ante*, notamment la documentation du système d'IA à haut risque. L'évaluation de l'effectivité des procédures, politiques et stratégies contenues dans le système de gestion de la qualité peut donc être faite grâce au contrôle, plus poussé, de la documentation technique. Le système de gestion de la qualité sert alors également d'outil de comparaison pour observer si les éléments qu'il contient sont respectés dans leur déclinaison réalisée pour un système d'IA à haut risque précis. En conclusion, l'analyse de la procédure d'évaluation du système de gestion de la qualité fait

⁹⁴⁶ Article 17.1 du règlement européen sur l'IA. Pour plus de précisions sur le contenu de ce document, voir le paragraphe précédent.

apparaître que l'influence de l'acteur du contrôle *ex ante* est limitée aux systèmes d'IA et ne s'étend pas aux exigences plus générales imposées aux fournisseurs.

B- Le contrôle des éléments techniques de preuve

L'accès aux éléments techniques de preuve. Contrairement au système de gestion de la qualité, la procédure de contrôle de la documentation technique et du système de surveillance après commercialisation apparaît plus approfondie, en particulier lorsque le contrôle est réalisé par un organisme notifié. En effet, le législateur européen octroie au tiers certificateur deux moyens de contrôle des éléments techniques de preuves qui fondent la conformité du système d'IA à haut risque. Premièrement, le tiers évaluateur détient un accès à toutes les données de conception :

« À cette fin, l'organisme notifié se voit accorder un accès complet aux jeux de données d'entraînement et de test utilisés par le fournisseur, y compris par l'intermédiaire d'interfaces de programmation ou d'autres moyens et outils appropriés permettant un accès à distance. »⁹⁴⁷

Dès lors, ce-dernier est capable de vérifier lui-même le respect par le fournisseur du système d'IA des exigences relatives aux données mentionnées à l'article 10 du règlement européen, qu'il s'agisse des données en elles-mêmes⁹⁴⁸ ou de leur traitement⁹⁴⁹. De façon surprenante, la documentation technique ne mentionne pas la problématique des biais, que ce soit explicitement ou au travers des qualités de représentativité ou encore d'exactitude. En dépit de cette absence, cet accès complet aux données de conception permet à l'organisme d'évaluer la véracité des informations contenues dans la documentation technique. Dans une logique similaire, l'accès aux données de conception peut s'accompagner d'un accès au code source du système d'IA à haut risque⁹⁵⁰.

⁹⁴⁷ « Règlement sur l'intelligence artificielle », 2024. Annexe VII.4.3.

⁹⁴⁸ La documentation technique doit obligatoirement indiquer certaines informations sur les données telles que « la manière dont les données ont été obtenues et sélectionnées » mais aussi des « informations sur la provenance de ces jeux de données, leur portée et leurs principales caractéristiques ; » - Ibid. Annexe IV.2(d)

⁹⁴⁹ Il peut s'agir aussi bien des « méthodes et techniques d'entraînement [...] utilisés ; [...] les procédures d'étiquetage [...], les méthodes de nettoyage des données ». - Ibid.

⁹⁵⁰ Contrairement à l'accès aux données, l'accès au code source doit être nécessaire. Cette condition doit alors être justifiée dans une demande écrite par l'organisme notifié. Le fournisseur du système d'IA évalué ne semble en revanche pas en mesure de refuser et/ou contester cette demande de façon officielle. Ce type de recours semble en effet prévu seulement pour les décisions finales liées au contrôle *ex ante* de conformité.

L'examen effectif des preuves techniques. Deuxièmement, l'organisme notifié peut exiger la production d'autres preuves techniques en lien avec les exigences essentielles du règlement européen :

« Lors de l'examen de la documentation technique, l'organisme notifié peut exiger que le fournisseur apporte des preuves supplémentaires ou effectue des tests supplémentaires afin de permettre une évaluation correcte de la conformité du système d'IA avec les exigences énoncées au titre III, chapitre 2. Chaque fois que l'organisme notifié n'est pas satisfait des tests effectués par le fournisseur, l'organisme notifié effectue directement des tests adéquats, le cas échéant. »⁹⁵¹

Par cette prérogative, l'organisme notifié se voit octroyer deux modalités de contrôle des preuves techniques. Tout d'abord, la demande de production de nouveaux tests et/ou preuves techniques élargit la portée du contrôle de conformité. Il ne s'agit pas seulement d'examiner l'adéquation des preuves de conception soumises par le fournisseur avec le droit européen mais également d'évaluer concrètement la conformité du système d'IA. En d'autres termes, l'organisme notifié n'est pas limité par les choix probatoires réalisés par le fournisseur de systèmes d'IA lors de la mise en application des exigences européennes. Par exemple, il pourrait demander la réalisation de nouveaux tests sur la base de métriques différentes. Le tiers contrôleur de conformité a le pouvoir de faire valoir son interprétation et ses choix relatifs aux preuves techniques de conception pertinentes. De plus, s'il juge les résultats des tests supplémentaires, dont il a demandé la réalisation, insuffisants, il peut refuser l'octroi du certificat européen en dépit des bons résultats obtenus par le logiciel sur les premiers tests conçus par le fournisseur. Sa fonction de contrôleur de la conformité lui accorde ainsi la capacité d'imposer sa vision des preuves techniques pertinentes sur celle du fournisseur de systèmes d'IA.

Enfin, la possibilité de réaliser lui-même les tests, et donc d'obtenir lui-même les preuves techniques demandées, offre à l'organisme notifié un second moyen de contrôle relatif à la qualité des preuves techniques. Il peut ainsi vérifier l'exactitude des preuves présentées par le fournisseur. Les preuves documentaires ne permettent qu'un contrôle indirect des preuves techniques. Le contrôleur doit faire confiance au fournisseur du logiciel d'IA à haut risque quant à la véracité des preuves contenues dans la documentation technique ou encore le système de surveillance. En exécutant les tests de qualité, le tiers notifié peut réaliser un contrôle direct du

⁹⁵¹ « Règlement sur l'intelligence artificielle », 2024. Annexe VII.4.4.

respect des exigences relatives au fonctionnement du système d'IA. Il génère lui-même les preuves techniques qu'il considère comme pertinentes pour sa mission de certification. Il évite ainsi certains problèmes comme la manipulation mensongère des résultats de tests ou bien un résultat dont la fiabilité serait trop faible pour être considérée comme scientifiquement acceptable⁹⁵². Grâce à cette disposition du règlement européen, l'organisme notifié, malgré son statut de tiers extérieur à la conception du système d'IA à haut risque, peut réaliser un contrôle effectif des éléments techniques de preuve lors de sa mission de certification. Par conséquent, il peut contribuer de façon significative à la sélection des éléments techniques de preuves qu'il considère comme pertinents en tant que preuve juridique, et donc à l'interprétation opérationnelle des exigences du règlement européens sur l'IA.

CONCLUSION DE LA SECTION I :

Le contrôle *ex ante* de la conformité d'un système d'IA à haut risque met en avant deux nouveaux acteurs dans le processus de détermination des preuves techniques pertinentes en droit de l'IA. Tout d'abord, le fournisseur du système d'IA a la possibilité, au travers du mécanisme d'auto-certification, de valider la pertinence des éléments techniques de preuve, qu'il a choisis et obtenus lors des tests, comme preuves juridiques de la conformité de son système d'IA au règlement européen. Au regard des règles de désignation de l'acteur du contrôle *ex ante*, le législateur européen a souhaité confier la très grande majorité des évaluations de conformité aux fournisseurs eux-mêmes. Ces conditions favorables de désignation octroient de fait une influence formelle déterminante et incontestable aux fournisseurs de systèmes d'IA à haut risque sur les choix des preuves juridiques pertinentes. Ce pouvoir formel de contrôle vient renforcer l'importante influence substantielle dont ils bénéficient lors de la mise en application des exigences au cours de la conception du système. De l'autre côté, le règlement européen confie un pouvoir de contrôle substantiel, équivalent à celui des fournisseurs, aux organismes notifiés. Grâce à des prérogatives pratiques pertinentes, ces tiers notifiés ont la capacité d'imposer leurs choix probatoires, et donc leur interprétation des exigences juridiques, sur ceux du fournisseur. Toutefois, la rareté des situations dans lesquelles ils sont désignés comme contrôleur réduit formellement leur influence globale sur la sélection des éléments techniques de preuves pertinents juridiquement.

⁹⁵² Plusieurs facteurs peuvent permettre d'évaluer la qualité d'un résultat algorithmique (marge d'erreur par exemple). Ces éléments seront discutés plus en détails dans la partie 2 de la présente thèse.

Au regard du rôle joué par chacun des deux acteurs, il apparaît clairement un déséquilibre au bénéfice des fournisseurs de systèmes d'IA à haut risque. Nous rappelons ainsi l'importance capitale de généraliser le recours à un contrôle *ex ante* de conformité exercé par un tiers indépendant, comme cela est déjà le cas dans plusieurs réglementations⁹⁵³. La réalisation d'une telle tâche par le fournisseur lui-même doit ainsi être réservée aux systèmes d'IA présentant des risques non critiques. Dans cette hypothèse, les choix cruciaux relatifs aux indicateurs et aux seuils probabilistes, pouvant être utilisés comme preuves juridiques, seraient décidés par un acteur dont l'indépendance est garantie. Avec l'aide des autorités notifiantes, les tiers certificateurs vérifieraient ainsi les éléments techniques de preuve produits, à savoir la traduction quantitative des exigences, ne dénature pas de façon déraisonnable les principes juridiques originels. Cette proposition doit toutefois s'accompagner d'un contrôle régulier des organismes notifiés afin de vérifier que ces derniers démontrent toujours les garanties nécessaires à leur notification. Cette proposition est également bienvenue dans l'état actuel du règlement sur l'IA, et ce malgré l'influence formelle faible des organismes notifiés.

Section II : Le contrôle ex post de la conception du système d'IA

Annonce de plan. La stratégie de certification des systèmes d'IA à haut risque mise en place par le législateur européen met en avant l'importance du contrôle *ex ante* de la conception préalablement à la mise en service ou la mise sur le marché desdits logiciels. Pour autant, le contrôle juridique ne s'arrête pas avec l'apposition du marquage CE sur le système d'IA à haut risque par le fournisseur. Comme le système de surveillance post commercialisation le laissait deviner, la conformité du système d'IA aux exigences du règlement européen est vérifiée tout au long de son cycle de vie. Toutefois, le contrôle *ex post* de la conception ne se résume pas à une extension dans le temps de la procédure de contrôle *ex ante*. Il s'agit d'un mécanisme à part entière dans la volonté de garantir une conformité juridique européenne aux systèmes d'IA à haut risque. De ce fait, il implique de nouveaux acteurs et de nouvelles procédures qui, à l'image du contrôle *ex ante*, représentent autant d'intérêts et de modalités de contribution aux choix des éléments techniques de preuve pertinents. La phase de contrôle *ex post* se divise en deux

⁹⁵³ Voir notamment le règlement européen sur la cybersécurité : « Règlement (UE) 2019/881 du Parlement européen et du Conseil du 17 avril 2019 relatif à l'ENISA (Agence de l'Union européenne pour la cybersécurité) et à la certification de cybersécurité des technologies de l'information et des communications, et abrogeant le règlement (UE) no 526/2013 (règlement sur la cybersécurité) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2019.

procédures distinctes avec, dans un premier temps, le contrôle administratif de la conception (**Sous-section 1**) et, dans un second temps, un contrôle juridictionnel (**Sous-section 2**).

Sous-section 1 : Le contrôle administratif de la conception

Annonce du plan. Le contrôle administratif de la conformité est la première étape du contrôle *ex post* des systèmes d'IA à haut risque prévue par l'Union européenne. Il s'agit non seulement du contrôle de la conformité des systèmes d'IA mais également du contrôle de l'interprétation des exigences européennes par les acteurs du contrôle *ex ante*. Le contrôle administratif peut être le premier lieu de contradictions entre deux visions de ce que doit être la conformité des systèmes d'IA, dont les éléments techniques de preuve sont l'essence. L'exercice de ce contre-pouvoir est réalisé par des acteurs divers mais différents de ceux du contrôle *ex ante* (**§1**). Cependant, pour qu'ils puissent exercer une influence sur le système probatoire à la hauteur de celle des précédents acteurs, ils doivent bénéficier de prérogatives légales adéquates afin d'accéder, évaluer et proposer leurs choix de preuves (**§2**).

§1. Les acteurs du contrôle administratif

Annonce de plan. Au regard de l'envergure de la mission que représente la surveillance du marché européen des systèmes d'IA à haut risque, il n'est pas envisageable qu'un acteur unique puisse, seul, satisfaire cette tâche. Par conséquent, il est préférable que le contrôle administratif *ex post* de la conformité implique, de gré ou de force, la participation de plusieurs protagonistes à l'effort. Le système de surveillance du marché des systèmes d'IA à haut risque s'inscrit dans la logique du règlement (UE) 2019/1020 sur la surveillance du marché et la conformité des produits⁹⁵⁴. Dans ce contexte, il apparaît deux types d'acteurs avec, d'un côté, ceux à l'initiative du contrôle administratif du système d'IA à haut risque (**A**) et, de l'autre côté, ceux auteurs dudit contrôle (**B**).

⁹⁵⁴ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019.

A- Les initiateurs du contrôle

Annonce du plan. Contrairement au contrôle *ex ante* systématique, la mise en œuvre d'une évaluation *ex post* de conformité doit être initiée par un acteur au préalable. Les modalités d'initiation renseignent donc beaucoup sur l'ampleur de la réalisation de ce contrôle administratif *ex post*. Par conséquent, l'étude des règles d'initiation contribue à estimer formellement la capacité des autorités administratives à contrôler les éléments techniques de preuve choisis par les fournisseurs d'IA à haut risque et valider par les acteurs du contrôle *ex ante*. Deux acteurs se voient chargés d'initier un contrôle administratif *ex post* d'un système d'IA à haut risque. D'une part, on retrouve les acteurs économiques de la chaîne de production du système d'IA qui contribuent à la surveillance réactive du marché⁹⁵⁵ (1). D'autre part, les autorités administratives elles-mêmes peuvent réaliser une surveillance proactive du marché⁹⁵⁶ et ainsi initier un contrôle administratif *ex post* (2).

1. Les opérateurs économiques

Un devoir d'information de l'existence d'un risque. Le règlement européen impose aux opérateurs économiques un devoir de signalement de l'existence d'un risque aux autorités compétentes⁹⁵⁷. Il s'agit de l'application au contexte des systèmes d'IA à haut risque de l'obligation pour les opérateurs économiques d'informer les autorités de surveillance de marché si un produit présente un risque⁹⁵⁸. Dans notre étude, les opérateurs économiques du système d'IA sont le fournisseur, le déployeur, l'importateur et le distributeur. Au-delà de la prise de mesures correctives⁹⁵⁹, le fournisseur de systèmes d'IA à haut risque qui prend connaissance d'un tel risque doit en informer immédiatement les autorités de surveillance du marché de

⁹⁵⁵ La surveillance réactive du marché s'opère « *en réponse à un événement extérieur* » tandis que la surveillance proactive du marché désigne « *une activité planifiée de surveillance du marché* » - ADMINISTRATIVE COOPERATION GROUPS, « Bonnes pratiques pour la surveillance du marché », Administrative Cooperation Groups, 2017.

⁹⁵⁶ *Ibid.*

⁹⁵⁷ Le terme de « risque » renvoie ici à la situation où un système d'IA à haut risque est susceptible de nuire à la santé, à sécurité et à la protection des droits fondamentaux « *dans une mesure qui va au-delà de ce qui est considéré comme raisonnable et acceptable eu égard à l'objectif poursuivi ou dans les conditions d'utilisation normales ou raisonnablement prévisibles du produit concerné [...]* » - Article 65 de la proposition de règlement (UE) 2021/10106 et article 3.19) du règlement (UE) 2019/1020.

⁹⁵⁸ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 4.3.c)

⁹⁵⁹ « Règlement sur l'intelligence artificielle », 2024. Article 21.

chaque État membre dans lequel le système a été mis à disposition⁹⁶⁰. De façon similaire, les distributeurs qui considèrent ou suspectent un système d'IA à haut risque comme exposant à un risque allant au-delà du raisonnable et de l'acceptable⁹⁶¹ doivent en informer notamment les autorités nationales compétentes⁹⁶². La même obligation dans les mêmes conditions est imposée aux importateurs⁹⁶³ ainsi qu'aux déployeurs⁹⁶⁴. Ce devoir d'information imposé à la totalité des acteurs de la production d'un système d'IA fait donc endosser, de force, à ces derniers le rôle d'initiateur du contrôle administratif *ex post* des systèmes d'IA à haut risque.

L'obligation de notification des incidents graves et des dysfonctionnements. Outre ce devoir d'information aux autorités de surveillance du marché, le fournisseur est également soumis à une obligation de notification envers ces mêmes autorités. Il doit signaler à l'autorité de surveillance du marché tout incident grave⁹⁶⁵, à savoir :

« un incident ou dysfonctionnement d'un système d'IA entraînant directement ou indirectement :

(a) le décès d'une personne ou une atteinte grave à la santé d'une personne ;

(b) une perturbation grave et irréversible de la gestion ou du fonctionnement d'infrastructures critiques ;

(c) la violation des obligations au titre du droit de l'Union visant à protéger les droits fondamentaux ;

*(d) un dommage grave à des biens ou à l'environnement ; »*⁹⁶⁶

⁹⁶⁰ *Ibid.* Article 21 alinéa 2.

⁹⁶¹ *Supra* note 108.

⁹⁶² « Règlement sur l'intelligence artificielle », 2024. Article 27.4.

⁹⁶³ *Ibid.* Article 26.2.

⁹⁶⁴ *Ibid.* Article 29.4.

⁹⁶⁵ *Ibid.* Article 62.1.

⁹⁶⁶ *Ibid.* Article 3(44).

Dans une logique similaire à celle du RGPD⁹⁶⁷, le règlement européen sur l'IA ne fixe pas de seuil de criticité du risque comme condition de notification aux autorités⁹⁶⁸. Il convient néanmoins de se réjouir de cette absence de condition de criticité qui augmente considérablement l'information des autorités administratives et par conséquent leur capacité à exercer un contrôle *ex post* de la conformité des systèmes d'IA à haut risque⁹⁶⁹. Une fois cette notification reçue, l'autorité de surveillance du marché est dans l'obligation de transmettre cette information aux :

« [...] autorités et organismes publics nationaux qui supervisent ou font respecter les obligations au titre du droit de l'Union visant à protéger les droits fondamentaux en ce qui concerne l'utilisation des systèmes d'IA à haut risque visés à l'annexe III [...] »⁹⁷⁰

Cette dernière disposition renforce considérablement la probabilité que l'information communiquée par le fournisseur du système d'IA à haut risque aboutisse à un contrôle administratif *ex post* de la conformité dudit systèmes d'IA, que ce contrôle soit réalisé par une ou plusieurs autorités administratives.

2. L'autorité administrative

Une procédure d'initiation fondée sur les risques. Au travers de sa proposition de règlement, la Commission européenne offre une certaine liberté aux autorités de surveillance de marché pour initier un contrôle de conformité *ex post*. Ces dernières peuvent lancer une procédure de contrôle dès lors qu'elles ont des « *raisons suffisantes de considérer qu'un système d'IA*

⁹⁶⁷ Si le RGPD pose une condition de criticité pour la communication d'une violation de données à caractère personnel aux personnes concernées, la seule conséquence d'un risque pour les droits et libertés oblige le responsable de traitement à notifier l'autorité de contrôle de l'évènement. Voir « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Articles 33 et 34.

⁹⁶⁸ Ce constat peut potentiellement être nuancé par l'expression « *incident grave* » de l'article 62. Toutefois, nous n'interprétons pas cette expression comme pouvant avoir une conséquence limitative significative sur le rôle du fournisseur comme initiateur d'un contrôle administratif *ex post* de la conformité. D'une part, la condition ne porte alors pas sur la criticité de la conséquence mais sur celle de l'évènement. D'autre part, il semble peu probable qu'un incident survenant dans le contexte d'un système d'IA à haut risque qui aurait des conséquences sur les droits fondamentaux ne soit pas qualifié de « grave ».

⁹⁶⁹ À titre d'exemple, la proposition de loi canadienne sur l'intelligence artificielle et les données n'impose au responsable d'un système d'IA à incidence élevée seulement s'il « *entraîne, ou entraînera vraisemblablement, un préjudice important.* » - *Loi sur l'intelligence artificielle et les données*, 16 juin 2022. Article 12.

⁹⁷⁰ « Règlement sur l'intelligence artificielle », 2024. Articles 62.2 et 64.3.

présente un risque [pour la sécurité, la santé et les droits fondamentaux] »⁹⁷¹. Les justifications possibles pour initier un contrôle de conformité ne sont pas pour autant laissées à la discrétion desdites autorités⁹⁷². Elles doivent s'inscrire au sein d'une démarche par les risques. En d'autres termes, les autorités de surveillance du marché doivent prioriser les systèmes d'IA qui semblent présenter les risques les plus importants lors de leur utilisation. Les facteurs de détermination des risques sont listés exhaustivement par le règlement (UE) 2019/1020 :

« Lorsqu'elles décident des contrôles à effectuer, des types de produits à contrôler et de l'ampleur des contrôles, les autorités de surveillance du marché s'appuient sur une démarche fondée sur les risques tenant compte des facteurs suivants :

- a) les dangers et cas de non-conformité potentiels associés au produit et, si cette donnée est disponible, leur présence sur le marché ;*
- b) les activités et opérations qui sont sous le contrôle de l'opérateur économique ;*
- c) les antécédents de l'opérateur économique en matière de non-conformité ;*
- d) le cas échéant, le profil de risques établi par les autorités désignées en vertu de l'article 25, paragraphe 1 ;*
- e) les réclamations de consommateurs et autres informations provenant d'autres autorités, des opérateurs économiques, des médias et d'autres sources susceptibles d'indiquer une non-conformité. »*⁹⁷³

La capacité d'initiation des contrôles peut être influencée de façon globale soit au niveau européen, soit au niveau national. Au niveau européen tout d'abord, à la suite du vote et l'entrée en vigueur du texte européen, il est possible que les modalités d'initiation du contrôle administratif *ex post* de la conformité par les autorités de surveillance du marché soient spécifiées dans le contexte des systèmes d'IA à haut risque par la Commission européenne en fonction des tendances de non-conformité mises en avant pendant les premières années de contrôle⁹⁷⁴. De la même manière, il peut exister des variations au niveau national au travers des

⁹⁷¹ *Ibid.* Article 65.2.

⁹⁷² « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 11.3.

⁹⁷³ *Ibid.* Article 11.3.

⁹⁷⁴ *Ibid.* Article 11.4.

stratégies nationales qui définissent notamment les domaines prioritaires de surveillance ainsi que la fréquence des contrôles dans certains secteurs d'activité⁹⁷⁵.

B- Les auteurs du contrôle

Annonce du plan. Le rattachement du règlement européen sur l'IA à la logique plus globale du règlement 2019/1020 relatif à la surveillance de marché et à la conformité des produits invite logiquement les autorités de surveillance du marché à remplir ce rôle de contrôleur administratif *ex post* des systèmes d'IA à haut risque **(1)**. Toutefois, les risques sur les droits fondamentaux causés par ces logiciels sont une spécificité nouvelle pour ce type d'acteur, entraînant un besoin de coopération avec des organismes plus expérimentés dans le domaine **(2)**. L'implication de plusieurs acteurs dans la réalisation du contrôle administratif *ex post* de la conformité des systèmes d'IA à haut risque complexifie l'identification de l'acteur en capacité d'influencer la détermination des preuves techniques de la conformité de ces systèmes.

1. Le contrôle par les autorités de surveillance du marché

Définition de l'autorité de surveillance. L'autorité de surveillance du marché est définie par le règlement (UE) 2019/1020 comme « *une autorité désignée par un État membre en vertu de l'article 10 comme étant chargée d'assurer la surveillance du marché sur le territoire de ce dernier ;* »⁹⁷⁶. L'autorité de surveillance du marché poursuit trois objectifs :

« a) la surveillance efficace du marché des produits mis à disposition en ligne et hors ligne sur leur territoire en ce qui concerne les produits soumis à la législation d'harmonisation de l'Union ;

b) l'adoption par les opérateurs économiques de mesures correctives appropriées et proportionnées en ce qui concerne le respect de la législation d'harmonisation de l'Union et du présent règlement ;

c) l'adoption de mesures appropriées et proportionnées si l'opérateur ne prend pas de mesures correctives⁹⁷⁷. »

⁹⁷⁵ *Ibid.* Article 13.2.

⁹⁷⁶ *Ibid.* Article 3.4. Le règlement européen sur l'IA effectue un renvoi direct à cette définition à son article 3(26).

⁹⁷⁷ *Ibid.* Article 11.1.

Sa présence parmi les acteurs en charge du contrôle *ex post* du règlement sur l'IA découle directement de la stratégie européenne de considérer le système d'IA comme un produit. Dans le contexte du règlement européen, l'autorité de surveillance du marché des systèmes d'IA est davantage une fonction qu'une institution unique⁹⁷⁸. Ce statut de garantie supplémentaire nécessaire au contrôle *ex ante* est un besoin formulé dès les travaux préparatoires au texte européen⁹⁷⁹. Il démontre ainsi la place particulière de ces autorités le contrôle *ex post* des systèmes d'IA sur le marché européen. Enfin, l'autorité de surveillance du marché est soumise aux mêmes conditions d'indépendance, d'impartialité et d'objectivité dans la réalisation de ces missions⁹⁸⁰. Cette dernière caractéristique est centrale car elle est l'un des avantages intrinsèques au recours à une autorité administrative dans le contrôle de la conformité⁹⁸¹.

Identité de l'autorité de surveillance. Si le recours à une autorité de surveillance du marché est une option envisagée depuis longtemps par l'Union européenne⁹⁸², la question de l'identité de l'acteur qui endossera ce rôle est plus épineuse. Le principe est le suivant : l'autorité de contrôle nationale remplit également le rôle d'autorité de surveillance du marché⁹⁸³. Cependant les exceptions sont nombreuses avec, d'une part, des exceptions imposées par le texte européen et, d'autre part, une marge de manœuvre laissée aux États membres pour créer des exceptions sous couvert de justifications⁹⁸⁴. L'objectif n'est pas ici de lister l'ensemble des situations dans lesquelles l'autorité de surveillance du marché n'est pas également l'autorité de contrôle au sens du règlement européen, et ce pour deux raisons. Premièrement, les exceptions nationales ne sont pas encore connues, ou du moins les hypothèses énoncées ne peuvent être considérées comme certaines et ne nous semblent donc pas indispensables à commenter ici. Secondement, et le plus important, la liste exhaustive de ces situations exceptionnelles ne nous semble pas pertinentes pour étudier les acteurs capables de déterminer les preuves techniques pertinentes

⁹⁷⁸ De ce fait, la fonction de surveillance du marché et celle de notification des organismes d'évaluation de la conformité sont en principe réunies dans les mains de l'autorité de contrôle.

⁹⁷⁹ « *Toute évaluation préalable de la conformité devrait s'entendre sans fléchissement du contrôle de la conformité et de l'exécution ex post par les autorités nationales compétentes. Cela vaut pour les applications d'IA à haut risque, mais également pour d'autres applications d'IA soumises à des exigences légales [...]* » - COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020. P. 27.

⁹⁸⁰ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 11.2.

⁹⁸¹ SCHERER Matthew U., « Regulating Artificial Intelligence Systems », *Harvard Journal of Law & Technology*, 29, 2015. PP.383 à 387.

⁹⁸² COMMISSION EUROPÉENNE, *op. cit.* « *Ce contrôle de la conformité devrait s'inscrire dans le cadre d'un système continu de surveillance du marché.* ». P.28.

⁹⁸³ Ainsi que le rôle d'autorité notifiante. « Règlement sur l'intelligence artificielle », 2024. Article 59.2.

⁹⁸⁴ « *L'autorité de contrôle nationale agit en tant qu'autorité notifiante et autorité de surveillance du marché, sauf si un État membre a des raisons organisationnelles et administratives de désigner plus d'une autorité.* » - Ibid.

relatives la conformité des systèmes d'IA. Par conséquent, l'identité de l'autorité de surveillance ne sera abordée que lorsqu'elle peut avoir une influence sur ces choix probatoires.

De façon similaire à la désignation des organismes notifiés, lorsque des systèmes d'IA sont utilisés comme composants d'un produit plus global, l'autorité de surveillance du marché est celle désignée par le règlement encadrant ledit produit final. Par conséquent, chaque secteur mentionné dans l'annexe II du règlement sur l'IA est sous la surveillance d'une autorité dédiée⁹⁸⁵. Une situation exceptionnelle semblable⁹⁸⁶ s'applique pour les systèmes d'IA utilisés dans le cadre de services financiers, par des autorités répressives ou pour la gestion de la migration, de l'asile et des contrôles aux frontières⁹⁸⁷. Cependant, toutes ces autorités de surveillance du marché sont soumises à une obligation d'indépendance, d'impartialité et d'objectivité dans la réalisation de leurs missions⁹⁸⁸. L'identité de l'autorité en charge de la surveillance du marché devrait donc avoir un impact moindre sur la réalisation du contrôle, contrairement aux acteurs du contrôle *ex ante*, en ce qui concerne l'évaluation de la conformité des systèmes d'IA à haut risque⁹⁸⁹. Par conséquent, la validation, ou le refus, des choix de éléments techniques de preuve réalisés par le fournisseur d'IA, et validés lors du contrôle *ex ante*, ne sera pas significativement impacté par l'identité de l'autorité administrative en charge du contrôle *ex post*.

⁹⁸⁵ Pour illustrer, voici certains exemples d'autorité de surveillance de marché dans des secteurs de l'annexe II.A pour la France. Le contrôle *ex post* du règlement (UE) 2017/746 relatif aux dispositifs médicaux de diagnostic in vitro ainsi que le règlement (UE) 2017/745 relatif aux dispositifs médicaux est réalisé par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM). Pour la directive 2006/42/CE relative aux machines, la surveillance du marché est exercée par la Ministère du travail, de l'emploi et de l'insertion. Pour la liste de toutes les autorités de surveillance du marché par secteurs, voir COMMISSION EUROPÉENNE, « List of national market surveillance authorities by country », publié le 3 février 2023.

⁹⁸⁶ La situation n'est pas totalement identique car le règlement européen pour les systèmes d'IA ne désigne pas une seule autorité de surveillance de marché pour certains de ces secteurs mais énonce trois possibilités aux États membres. « Règlement sur l'intelligence artificielle », 2024. Article 63.3 à 63.5.

⁹⁸⁷ *Ibid. Ibid.*

⁹⁸⁸ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 11.2.

⁹⁸⁹ Pour les missions dépassant le contrôle de la certification des produits, le niveau d'indépendance, notamment vis-à-vis de l'État, présentée par les autorités de surveillance du marché peut être considéré comme insuffisant : VEALE Michael et ZUIDERVEEN BORGESIJUS Frederik, « Demystifying the Draft EU Artificial Intelligence Act », *Computer Law Review International*, 22, 2021. P. 111.

2. L'intervention des autorités de protection des droits fondamentaux

Un nouvel acteur du contrôle administratif ex post des produits. Les autorités de surveillance du marché ne sont pas les seules autorités administratives impliquées dans le contrôle *ex post* de la conformité des systèmes d'IA. Le règlement sur l'IA implique également une nouvelle catégorie d'acteurs étrangers à la surveillance du marché, à savoir :

« [Les] autorités ou organismes publics nationaux qui supervisent ou font respecter les obligations au titre du droit de l'Union visant à protéger les droits fondamentaux en ce qui concerne l'utilisation des systèmes d'IA à haut risque visés à l'annexe III [...] »⁹⁹⁰.

Ces nouveaux acteurs n'interviennent que pour les systèmes d'IA à haut risque présents dans l'annexe III. Encore inconnus aujourd'hui, ces acteurs seront listés par chaque États membres dans un délai de trois mois à compter de l'entrée en vigueur du règlement européen sur l'IA⁹⁹¹.

L'apport d'une expertise complémentaire sur les droits fondamentaux. L'apparition de ces acteurs dans le contrôle administratif *ex post* de la conformité est une nouveauté bienvenue dans la gouvernance de la conformité européenne des systèmes d'IA. Le recours à ces organismes se justifie par l'originalité du triple objectif poursuivi par le règlement européen, à savoir la protection de la santé, de la sécurité et des droits fondamentaux. En comparaison, les autres textes d'harmonisation du marché européen ne poursuivent qu'un objectif double : la santé et la sécurité⁹⁹². L'objectif de protection des droits fondamentaux est donc un objectif spécifique aux systèmes d'IA et par conséquent nouveau pour les autorités de surveillance du marché en place. Dès lors que cet objectif est mis à mal par la violation d'une obligation énoncée par le règlement, les autorités de surveillance doivent en informer ces nouveaux acteurs⁹⁹³ afin qu'ils

⁹⁹⁰ « Règlement sur l'intelligence artificielle », 2024. Article 64.3.

⁹⁹¹ *Ibid.* Article 64.4.

⁹⁹² À titre d'exemple : « Étant donné que l'objectif de la présente directive, à savoir de fixer les exigences essentielles de santé et de sécurité relatives à la conception et à la construction afin d'améliorer la sécurité des machines » - « Directive 2006/42/CE du Parlement européen et du Conseil du 17 mai 2006 relative aux machines et modifiant la directive 95/16/CE (refonte) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2006. Considérant 28 ; « établir un cadre réglementaire rigoureux, transparent, prévisible et durable pour les dispositifs médicaux, qui garantisse un niveau élevé de sécurité et de protection de la santé » - « Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017. Considérant 1.

⁹⁹³ « Dès la réception d'une notification relative à un incident grave visé à l'article 3, paragraphe 44, point c), l'autorité de surveillance du marché concernée informe les autorités ou organismes publics nationaux publiques nationales visées à l'article 64, paragraphe 3. » - « Règlement sur l'intelligence artificielle », 2024. Article 62.2.

collaborent dans la réalisation du contrôle *ex post*⁹⁹⁴. La limitation explicite de l'intervention de ces nouveaux acteurs aux risques pour les droits fondamentaux⁹⁹⁵ permet de présenter clairement leur rôle comme l'apport d'une légitimité et d'une expertise manquante aux autorités de surveillance du marché traditionnelles. Cette fonction leur confère donc une place importante dans la détermination des preuves techniques de certaines exigences du règlement européen, en particulier la détection des biais dans les données⁹⁹⁶ et des performances au regard des différentes catégories de population⁹⁹⁷.

§2. La réalisation du contrôle administratif

Annonce de plan. Les autorités de surveillance de marché et les organismes publics nationaux en charge de la protection des droits fondamentaux sont donc les deux acteurs du contrôle administratif *ex post* de la conformité des systèmes d'IA. À première vue, ce statut leur confère à tous les deux une influence théorique certaine sur les choix des preuves techniques pertinentes des systèmes d'IA. Afin d'évaluer leur influence exacte dans ces choix probatoires, il est indispensable de comprendre la place du contrôle *ex post* dans la détermination des preuves techniques pertinentes **(A)** avant d'examiner les prérogatives octroyées aux acteurs en charge de ce contrôle **(B)**.

A- La place du contrôle ex post dans les choix des preuves techniques pertinentes

Un contrôle de la conformité *ex post*. La mission première confiée aux autorités de surveillance du marché européen est de contrôler la conformité des systèmes d'IA à haut risque mis sur le marché. Contrairement au contrôle *ex ante*, le contrôle *ex post* n'est pas systématique

⁹⁹⁴ L'autorité de surveillance de marché et les organismes de protection des droits fondamentaux sont mentionnés à plusieurs reprises comme travaillant en collaboration lors de la survenance de violation des droits fondamentaux en lien avec les obligations imposées aux systèmes d'IA. Voir notamment les articles 62,64 et 65 du règlement européen sur l'IA. .

⁹⁹⁵ « Dès réception d'une notification relative à une violation des obligations au titre du droit de l'Union visant à protéger les droits fondamentaux [...] » ; « Dès réception d'une notification relative à une violation des obligations au titre du droit de l'Union visant à protéger les droits fondamentaux [...] » - Respectivement l'article 63 et l'article 65 du règlement sur l'IA. Il s'agit des conditions d'intervention des organismes et autorités publics nationaux chargés de la protection des droits fondamentaux selon le texte européen.

⁹⁹⁶ « Règlement sur l'intelligence artificielle », 2024. Article 10.

⁹⁹⁷ *Ibid.* Article 15.

et la fréquence de contrôle dépend des informations à disposition de l'autorité de surveillance du marché, de ses ressources pour réaliser lesdits contrôles et, dans certaines situations des objectifs posés par la Commission européenne. Le rôle d'évaluation de la conformité *ex post* permet aux acteurs, notamment les autorités de surveillance du marché, de confronter la conformité établie par le fournisseur de systèmes d'IA à leur interprétation des exigences réglementaires. Dans certains cas, la conformité contrôlée sera établie par un organisme notifié⁹⁹⁸. En intervenant après l'évaluation *ex ante*, les autorités de surveillance du marché, au travers de leur contrôle, ont donc le pouvoir d'imposer leur interprétation des exigences européennes sur celles de tous les acteurs qui les ont précédés. Par conséquent, la temporalité de leur évaluation de la conformité révèle une influence formelle décisive sur l'élaboration des preuves techniques.

Un contrôle au regard des objectifs du règlement. Ensuite, les autorités de surveillance du marché contrôlent la conformité des systèmes d'IA à haut risque aux spécifications techniques mais également au regard des exigences essentielles du règlement européen⁹⁹⁹. De ce fait, elles peuvent constater le respect par le fournisseur des normes européennes harmonisées et pour autant conclure qu'il n'est pas conforme au règlement sur l'IA car, du fait d'un contexte particulier par exemple, les spécifications techniques sont lacunaires. Les autorités de surveillance peuvent donc remettre en question la présomption de conformité octroyée par le respect des normes techniques européennes, ou des spécifications communes le cas échéant. Par conséquent, elles évaluent également, pour un contexte donné, l'adéquation entre les spécifications techniques et les exigences du règlement sur l'IA. De ce fait, si les normes techniques indiquent des seuils statistiques ou des métriques à respecter, l'autorité de surveillance peut conclure que ces preuves techniques ne sont pas pertinentes (soit parce qu'elles sont inadéquates au règlement, soit parce qu'elles ne sont pas suffisantes). Pour rappel, le règlement européen sur l'IA impose que l'évaluation des risques soit réalisée notamment sur la base de « *métriques et de seuils probabilistes préalablement définis* »¹⁰⁰⁰. Par conséquent, il est certain que le contrôle de la conformité de ces systèmes implique la vérification de la réalisation correcte de ces procédures de tests et des résultats obtenus car ils représentent des arguments décisifs pour déclarer la conformité du système d'IA aux exigences textuelles. Alors

⁹⁹⁸ Cour de justice de l'Union européenne, 12 juillet 2012, n° C-171/11.

⁹⁹⁹ « *En particulier, l'autorité de surveillance du marché indique si la non-conformité découle d'une ou plusieurs des causes suivantes : (a) le non-respect, par le système d'IA, des exigences énoncées au titre III, chapitre 2; (b) des lacunes dans les normes harmonisées ou dans les spécifications communes visées aux articles 40 et 41 qui confèrent une présomption de conformité.* » - « Règlement sur l'intelligence artificielle », 2024. Article 65.6.

¹⁰⁰⁰ *Ibid.* Article 9.7.

que la Commission européenne et les juridictions ont démontré des limites conséquentes à leur capacité de contrôle des normes techniques européennes¹⁰⁰¹, les autorités de surveillance du marché apparaissent comme la première possibilité de les contrôler de façon effective. Ces autorités de contrôle *ex post* peuvent donc contribuer de façon considérable à la détermination des éléments techniques de preuves pertinents pour attester du respect des obligations du règlement sur l'IA.

B- Les prérogatives octroyées aux acteurs du contrôle ex post

Annonce de plan. La capacité formelle de réaliser un contrôle ne signifie pas pour autant que les éléments techniques de preuve seront évalués dans le cadre du contrôle. Du fait de sa complexité et de son caractère hautement technique, la problématique des choix pertinents de ces preuves peut ne pas être abordée lors d'un contrôle de conformité. Pour les acteurs en charge du contrôle *ex post*, la capacité à contribuer substantiellement à ces choix probatoires cruciaux dépend des prérogatives qui leur sont octroyées pour la réalisation de leurs missions. Ainsi, le législateur leur a confié la capacité d'accéder aux éléments techniques de preuve **(1)** ainsi que le pouvoir de produire eux-mêmes les preuves techniques de conception **(2)**.

1. L'accès aux preuves du fournisseur.

L'accès aux preuves précisé pour les autorités de surveillance de marché. Les autorités de surveillance du marché des systèmes d'IA bénéficient de certaines prérogatives minimales garanties par leur statut. Parmi ces pouvoirs se trouve le pouvoir d'exiger l'accès aux éléments de preuves que sont les : « *documents, spécifications techniques, données ou informations pertinentes concernant la conformité du produit et ses caractéristiques techniques* »¹⁰⁰² ainsi que l'accès aux « *informations sur la chaîne d'approvisionnement, sur les détails du réseau de distribution, sur les quantités de produits sur le marché [...]* »¹⁰⁰³ tant que les éléments demandés sont pertinents aux fins de la mission de contrôle. Les autorités de surveillance du marché ont donc la possibilité d'accéder à un panel bien plus large de preuves que les preuves

¹⁰⁰¹ Voir Section II : L'influence des acteurs dans la définition des normes techniques.

¹⁰⁰² « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 14.4.a).

¹⁰⁰³ *Ibid.*

documentaires spécifiques aux systèmes d'IA et auxquelles les acteurs du contrôle *ex ante* sont limités¹⁰⁰⁴. Cet accroissement des preuves exploitables par les acteurs du contrôle est non seulement bienvenu mais aussi indispensable au regard du périmètre de la mission à couvrir¹⁰⁰⁵. L'interprétation des diverses exigences légales en preuves techniques quantifiables est la stratégie poursuivie par le législateur européen concernant les systèmes d'IA et cette tâche renferme des enjeux juridiques centraux¹⁰⁰⁶. Toutefois, le travail d'identification des limites desdites preuves techniques de conception par rapport aux risques produits par les systèmes d'IA nécessite l'accès à d'autres types d'informations¹⁰⁰⁷. Ensuite, au regard des particularités des technologies d'IA, le législateur européen précise également que les autorités de surveillance du marché peuvent demander l'accès à toutes les données de conception utilisées par le fournisseur et au code source si nécessaire¹⁰⁰⁸.

L'accès aux preuves garanti pour les entités nationales. Contrairement aux autorités de surveillance du marché, les autorités et organismes publics nationaux en charge de la protection des droits fondamentaux ne disposent pas d'un statut légal commun leur offrant des prérogatives garanties et unifiées. Encore inconnues, ces autorités devront être désignées par chaque État membre dans les trois mois suivants l'entrée en vigueur du texte¹⁰⁰⁹. Outre cette possible variation entre ces entités, leurs prérogatives actuelles pourraient ne pas s'avérer suffisantes pour réaliser leur mission spécifique face à l'opacité et la complexité des systèmes d'IA. Afin qu'elles puissent faire valoir leur expertise en tant que contrôleur *ex post* des systèmes d'IA, le règlement européen leur confère le droit d'accéder aux preuves documentaires établies ou conservées par les opérateurs économiques du marché des systèmes

¹⁰⁰⁴ À savoir le système de gestion de la qualité, la documentation technique et le système de surveillance après commercialisation. Voir la section 1 du présent chapitre pour plus d'informations.

¹⁰⁰⁵ L'accès à des informations relatives à la chaîne d'approvisionnement et au réseau de distribution est une opportunité de produire une évaluation plus globale de la légalité de la production d'un système d'IA. Voir MATUS Kira J.M. et VEALE Michael, « Certification systems for machine learning », *Regulation & Governance*, 16, 2022.

¹⁰⁰⁶ Voir Section I : L'importance des preuves au sein des normes techniques.

¹⁰⁰⁷ Une distinction est opérée par la doctrine entre les qualités « pratiques » (*experiential qualities*), observables au travers de l'utilisation de l'objet, et les qualités « acquises » (*credential qualities*), qui sont généralement plus difficiles à mesurer et/ou à attribuer. DARBY Michael R. et KARNI Edi, « Free Competition and the Optimal Amount of Fraud », *The Journal of Law and Economics*, 16, The University of Chicago Press, 1973. Maria Loureiro & Jill Mccluskey, « Reputation and Production Normes techniques » (2005) 30 *Journal of Agricultural and Resource Economics*.

¹⁰⁰⁸ « Règlement sur l'intelligence artificielle », 2024. Article 64.1 et 64.2.

¹⁰⁰⁹ Juliette Sénéchal, « L'AI Act dans sa version finale – provisoire –, une hydre à trois têtes » (2024) 11 (Dalloz actualités). « Règlement sur l'intelligence artificielle », 2024. Article 64.1.

d'IA¹⁰¹⁰. Ce nouveau pouvoir permet à ces entités d'exercer un contrôle documentaire sur la conformité des opérateurs économiques aux droits fondamentaux. Cette prérogative est d'autant plus importante qu'elle est à l'initiative seule desdites entités, sans nécessiter l'intervention de l'autorité de surveillance du marché¹⁰¹¹. Ce droit d'accès aux informations nécessaires au contrôle permet à ces deux autorités d'évaluer l'adéquation des éléments techniques de preuve documentés par rapport aux obligations dont elles ont la charge d'assurer le respect.

2. L'obtention proactive de preuves techniques.

La réalisation directe des tests techniques. En ce qui concerne la conformité du fonctionnement du système d'IA, les preuves documentaires fournies par les opérateurs économiques ne sont que des représentations de la réalité technique du système d'IA¹⁰¹². Malgré les sanctions encourues¹⁰¹³, la véracité des informations et la validité des preuves techniques communiquées ne dépendent en pratique que de la bonne foi lors de la collaboration de l'acteur¹⁰¹⁴. Pour faire face à cette limite et contrôler plus efficacement les risques juridiques envers les personnes affectées, les autorités de contrôle *ex post* se voient confier la compétence d'organiser des tests directement sur le système d'IA à haut risque examiné. En ce qui concerne les autorités de surveillance du marché, il s'agit simplement de l'application du règlement (UE) 2019/1020 au contexte des systèmes d'IA¹⁰¹⁵. La réalisation d'inspections et de contrôles physique devient simplement l'organisation de tests techniques sur les systèmes d'IA.

¹⁰¹⁰ « Les autorités ou organismes publics nationaux qui supervisent ou font respecter les obligations découlant du droit de l'Union en matière de protection des droits fondamentaux, y compris le droit à la non-discrimination, [...] sont habilités à de demander et d'accéder à toute documentation créée ou conservée en vertu du présent règlement » - « Règlement sur l'intelligence artificielle », 2024. Article 64.1.

¹⁰¹¹ L'autorité de surveillance du marché de l'État membre doit toutefois être informée de la demande d'accès documentaire réalisée. *Ibid.* Article 64.1.

¹⁰¹² Voir §1. Les obligations juridiques posées pour les systèmes d'IA à haut risque.

¹⁰¹³ Le règlement européen sur l'IA sanctionne la communication d'informations « *inexactes, incomplètes ou trompeuses* » aux acteurs du contrôle de la conformité, *ex ante* et *ex post*, par une amende administrative pouvant s'élever à 10.000.000 € ou 2% du chiffre d'affaires mondial annuel. Article 71.5 du règlement sur l'IA. Pour autant, cette sanction ne permet que de dissuader ces comportements.

¹⁰¹⁴ Dans un autre contexte, les limites du contrôle documentaire ont été mises en avant dans l'affaire des prothèses mammaires PIP. Cour de cassation, 10 octobre 2018, n° 15-26.093. « *il résulte de la directive que lors de l'examen de la demande, l'organisme notifié n'avait pour obligation que d'examiner le dossier technique qui lui était soumis ; qu'aucun élément ne pouvait laisser suspecter que le gel Nusil avait été remplacé par un gel non approuvé* ». Néanmoins, il convient de préciser dans cet exemple que le mécanisme de fraude du fournisseur étant particulièrement élaboré, une intervention sur place et une obtention de preuves techniques n'aurait probablement pas permis d'identifier ladite fraude selon la Cour de cassation.

¹⁰¹⁵ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 14.

La capacité de « *procéder à des inspections inopinées sur place et à des contrôles physiques des produits* »¹⁰¹⁶ poursuit un objectif de contrôle effectif du respect du règlement européen sur l'IA¹⁰¹⁷. Le législateur européen permet ainsi aux autorités de surveillance du marché de procéder à leurs propres tests. Par conséquent, ces autorités peuvent élaborer leurs propres procédures de contrôle, avec leurs propres métriques notamment, afin d'évaluer le système d'IA à haut risque. Ensuite, les autorités peuvent procéder à leur contrôle sur les nouveaux éléments de preuves qu'elles ont produites elles-mêmes. Premièrement, la possibilité de réaliser de nouveaux tests permet de garantir la véracité des preuves techniques communiquées par le fournisseur. Par exemple, si l'autorité en charge du contrôle réalise des procédures de tests identiques à celles indiquées dans la documentation mais que les résultats sont différents, alors les éléments techniques de preuve inscrits par le fournisseur sont au mieux, erronés, au pire, trompeurs. La différence entre les résultats de l'autorité et ceux de la documentation peuvent s'expliquer soit par une évolution du fonctionnement du système d'IA, soit par une divergence relative à la procédure de test réellement effectuée, soit par une manipulation des résultats obtenus. Deuxièmement, les nouveaux tests effectués par l'autorité de contrôle peuvent être utilisés comme tests complémentaires. L'autorité ne doute pas de la fiabilité des tests effectués par le fournisseur mais plutôt de leur pertinence au regard du règlement et du contexte d'utilisation du logiciel. Ainsi, l'autorité réalise d'autres procédures de tests afin de vérifier différemment la conformité du produit. La possibilité de réaliser des tests alternatifs confirme l'existence d'un choix dans les preuves techniques et l'importance de ces choix. Par cette capacité de réaliser des contrôles directement sur le produit, les autorités de surveillance du marché peuvent contrôler effectivement la véracité et la pertinence des preuves techniques produites par le fournisseur dans sa documentation.

La réalisation indirecte de tests techniques. En ce qui concerne les autorités et organismes publics nationaux en charge de la protection des droits fondamentaux, il convient de nuancer cette prérogative. Ces derniers n'ont en effet pas le droit de réaliser directement lesdits tests auprès des opérateurs concernés. Ils doivent en réalité formuler une demande à l'autorité de surveillance du marché qui organise ensuite, avec leur collaboration, les tests souhaités sur le système d'IA¹⁰¹⁸. En d'autres termes, ces organismes et autorités de protection des droits

¹⁰¹⁶ « Règlement sur l'intelligence artificielle », 2024. Article 63.3. « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 14.4.d)

¹⁰¹⁷ *Ibid.*

¹⁰¹⁸ « Règlement sur l'intelligence artificielle », 2024. Article 64.5.

fondamentaux ont le pouvoir d'initier des tests techniques sur les systèmes d'IA mais pas de les réaliser, cette prérogative étant réservée aux autorités de surveillance du marché. Néanmoins, en ce qui concerne notre analyse, cette divergence est négligeable. En effet, quelle que soit la nature de l'acteur du contrôle administratif *ex post*, ce-dernier a la capacité d'obtenir proactivement les éléments techniques de preuve qu'il souhaite pour réaliser son évaluation. Par conséquent, les autorités de contrôle *ex post* peuvent imposer leur choix de preuves, aussi bien concernant les choix juridiques associés que sur la qualité des preuves obtenues, aux interprétations des acteurs précédents dans la gouvernance des systèmes d'IA européens (élaborateurs de la norme et contrôleurs *ex ante* de la conformité).

Sous-section 2 : Le contrôle juridictionnel ex post

Annonce de plan. Le contrôle juridictionnel constitue la dernière étape de notre étude la détermination des éléments techniques de preuves juridiques pertinents dans le règlement européen sur l'IA. De ce fait, le mécanisme revêt une importance toute particulière car les choix réalisés sur les preuves techniques représenteront la vision définitive de la conformité. Au travers du droit de recours, le contrôle juridictionnel permet d'impliquer les individus, jusqu'à présent absent de notre étude. Ils peuvent alors faire-valoir leur vision de la conformité technique en contestant les décisions d'autres acteurs, et par conséquent s'opposer indirectement à la vision de la conformité desdits acteurs. Dans le cadre du droit de la conformité de l'IA, le contrôle juridictionnel se manifeste au travers de voies de recours offerts à certains acteurs (§1) qui peuvent alors solliciter l'intervention des juridictions dont les modalités de contrôle des preuves techniques sont encore incertaines (§2).

§1. Les activités faisant l'objet d'un recours juridictionnel

Annonce de plan. La place du contrôle juridictionnel dans la détermination du système probatoire de la conformité des systèmes d'IA est tout d'abord subordonnée à l'existence de voies de recours. Les conditions des recours juridictionnels existants représentent en effet la fréquence à laquelle les acteurs judiciaires pourront apposer leur influence sur les preuves techniques. Dans le contexte du règlement sur l'IA, le législateur européen propose une

possibilité de recours à l'encontre de deux types de décisions : celle prise par l'organisme notifié **(A)** et celle prise par l'autorité de surveillance du marché **(B)**. Les recours existants dépassent ceux inscrits dans le règlement européen sur l'IA. Toutefois, les deux situations sur lesquelles ce dernier s'attarde constituent deux étapes importantes pendant lesquelles le juge peut avoir une influence sur la sélection des éléments techniques de preuve pertinents, indépendamment de la voie de recours choisie.

A- La décision de l'organisme notifié

Les parties ayant un intérêt légitime. Le droit d'intenter un recours contre les décisions prises par les organismes notifiés est ouvert aux « *parties ayant un intérêt légitime dans ces décisions*¹⁰¹⁹ ». L'identification des acteurs qui pourraient présenter ledit intérêt légitime est une tâche complexe qui peut être élargie plus généralement à la présence d'un intérêt à agir. À ce propos, deux catégories de personnes sont envisageables.

L'ouverture d'un recours pour le fournisseur du produit. Premièrement, le fournisseur du système d'IA à haut risque appert comme le principal concerné par les décisions de l'organisme notifié. En effet, le fournisseur de systèmes d'IA entreprend d'importants investissements pour concevoir ledit systèmes d'IA et la décision de refus de délivrer le certificat de conformité par l'organisme notifié peut représenter pour lui un préjudice économique. De plus, l'organisme notifié et le fournisseur de systèmes d'IA sont liés par une relation contractuelle. Ce contrat qui lie les deux acteurs renforce encore plus l'existence d'un intérêt légitime pour le fournisseur du système d'IA sur lequel porte la décision contestée. Sans grande hésitation, ce-dernier est donc bénéficiaire de ce droit de recours.

La mise en place d'un recours juridictionnel pour l'individu affecté laissé à la discrétion des États membres. La question est cependant plus épineuse concernant l'existence d'un intérêt à agir pour les personnes affectées par l'utilisation du produit en question. Dans l'attente de l'entrée en vigueur du règlement européen sur l'IA, nous réaliserons notre étude à l'aide d'exemple d'autres marchés réglementés, notamment celui des dispositifs de santé. Il convient donc de rappeler que cet exercice est délicat mais il demeure indispensable pour formuler des hypothèses prospectives dans le cadre des systèmes d'IA. En outre, même au travers des autres législations d'harmonisation, l'identification de l'intérêt à agir des individus affectés, que ce

¹⁰¹⁹ *Ibid.*

soit au travers du recours spécifique ou de la responsabilité civile des organismes notifiés, est particulièrement complexe. Ainsi, dans le cadre des dispositifs médicaux, la CJUE indique que les organismes notifiés sont « *tenus à une obligation générale de diligence attachée à leur intervention dans le cadre de la procédure relative à la déclaration CE de conformité* »¹⁰²⁰ mais elle ne précise pas envers qui cette obligation de diligence est dûe ni si son manquement ouvre la possibilité d'un recours juridictionnel pour la personne affectée par ledit manquement¹⁰²¹. La CJUE indique dans son arrêt *Schmitt* de 2017 que l'établissement d'une responsabilité des organismes notifiés dans le cadre de l'évaluation de la conformité des dispositifs médicaux est à laissée à l'entière discrétion des États membres¹⁰²². Celle-ci n'est pas requise par l'Union européenne mais n'est pas interdite non plus¹⁰²³. Par conséquent, la capacité pour le juge de contrôler le travail de conformité de l'organisme notifié, notamment du choix des éléments techniques de preuve pertinents, est dépendante de la volonté de chaque État membre. En ce qui concerne la législation française, il semble que la responsabilité civile des organismes notifiés puisse être engagée par les patients¹⁰²⁴. Toutefois, il s'agit d'un exemple unique qui se base sur la directive 93/42/CEE, aujourd'hui abrogée par le règlement (UE) 2017/745 concernant les dispositifs médicaux.

B- La décision de l'autorité de surveillance du marché

La décision des autorités de surveillance du marché. La seconde voie de recours dans le cadre de la conformité des systèmes d'IA est ouverte tout d'abord à l'encontre de « *toute mesure, décision ou injonction prise par les autorités de surveillance du marché [...]* »¹⁰²⁵. Il peut s'agir aussi bien des mesures consistant à apposer des avertissements sur un produit que du retrait du marché de ce dernier¹⁰²⁶. L'inscription de ce mécanisme en tant que droit

¹⁰²⁰ Cour de justice de l'Union européenne, 16 février 2017, n° C-219/15. Par. 46.

¹⁰²¹ WALLERMAN Anna, « Pie in the Sky When You Die », *Common Market L. Rev.*, 55, 2018.

¹⁰²² Cour de justice de l'Union européenne, 16 février 2017, n° C-219/15.

¹⁰²³ À condition que la responsabilité des organismes notifiés ne chevauche pas le régime de responsabilité du fait des produits défectueux. WALLERMAN Anna, *op. cit.* Dans son opinion, l'avocat général Sharpston considère que l'exigence pour les organismes notifiés de contracter une assurance de responsabilité civile suggère que le législateur a souhaité que leur responsabilité puisse être engagée sur ce fondement : Cour de justice de l'Union européenne, 15 septembre 2016, n° C-219/15. Par. 35. Cette exigence est aussi présente dans l'article 33.8 du règlement européen sur l'IA, l'argument apparaît donc logiquement comme transposable au marché des systèmes d'IA.

¹⁰²⁴ Cour de cassation, 10 octobre 2018, n° 15-26.093.

¹⁰²⁵ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019.

¹⁰²⁶ *Ibid.* Article 16.3.

procédural au sein du règlement 2019/1020 relatif à la surveillance du marché révèle qu'il s'agit d'un droit général à toutes les réglementations européennes concernant des produits au sens de son annexe I. De ce fait, la présence de ce droit de recours contre la décision d'une autorité de surveillance du marché est observable dans d'autres marchés plus établis que celui des systèmes d'IA. À titre d'exemple, l'Agence nationale de sécurité du médicament et des produits de santé (ANSM ci-après) est l'autorité française de surveillance du marché¹⁰²⁷ des dispositifs médicaux¹⁰²⁸ ainsi que du marché des dispositifs médicaux de diagnostic *in vitro*¹⁰²⁹. Les décisions prises par le directeur général de l'ANSM, visant à autoriser ou refuser l'entrée sur le marché desdits produits, sont considérées comme des décisions administratives et peuvent donc faire l'objet d'un recours devant la juridiction administrative compétente¹⁰³⁰. De façon identique aux autorités de surveillance du marché, les décisions des organismes et autorités nationaux publics en charge de la protection des droits fondamentaux, mentionnés par le règlement européen sur l'IA, sont des décisions administratives et peuvent donc faire l'objet d'un recours juridictionnel.

L'intérêt de l'opérateur économique. Contrairement au recours contre les décisions des organismes notifiés, les décisions prises par les autorités de surveillance du marché sont contestables auprès d'une juridiction uniquement par « *l'opérateur économique concerné* »¹⁰³¹. Les titulaires potentiels d'un intérêt à agir sont donc explicitement indiqués par le législateur européen au sein du titre III chapitre 3 du règlement européen sur l'IA¹⁰³². Il s'agit des fournisseurs de systèmes d'IA, des fabricants de produits, des importateurs, des distributeurs et

¹⁰²⁷ COMMISSION EUROPÉENNE, « List of national market surveillance authorities by country », publié le 3 février 2023.

¹⁰²⁸ « RÈGLEMENT (UE) 2017/ 745 DU PARLEMENT EUROPÉEN ET DU CONSEIL - du 5 avril 2017 - relatif aux dispositifs médicaux, modifiant la directive 2001/ 83/ CE, le règlement (CE) no 178/ 2002 et le règlement (CE) no 1223/ 2009 et abrogeant les directives du Conseil 90/ 385/ CEE et 93/ 42/ CEE », 2017.

¹⁰²⁹ « Règlement (UE) 2017/746 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic *in vitro* et abrogeant la directive 98/79/CE et la décision 2010/227/UE de la Commission (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017.

¹⁰³⁰ CONSEIL D'ETAT, « Le juge administratif et le droit des médicaments - Dossier thématique », Conseil d'Etat, 2014.

¹⁰³¹ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 18.2.

¹⁰³² « Règlement sur l'intelligence artificielle », 2024. Article 63.1 : « (a) toute référence à un opérateur économique en vertu du règlement (UE) 2019/1020 s'entend comme incluant tous les opérateurs identifiés au titre III, chapitre 3, du présent règlement ; »

des déployeurs concernés par la décision ou concernés par le système d'IA ayant fait l'objet de la décision¹⁰³³.

§2. Le contrôle juridictionnel de la conformité des systèmes d'IA

Annonce de plan. Le contrôle juridictionnel étant confié aux législations nationales, il sera ici traité du cas du droit national français. L'objectif est alors d'étudier si les juridictions françaises sont en capacité de contrôler et d'influencer effectivement la pertinence des choix des éléments techniques de preuve (indicateurs et seuils probabilistes). Au préalable, il convient d'identifier les juridictions compétentes en droit français **(A)** pour ensuite étudier leur capacité à contrôler le choix des preuves techniques de conception **(B)**.

A- Les acteurs du contrôle juridictionnel

Les juridictions compétentes pour le contrôle des autorités de surveillance du marché.

L'identification de la juridiction compétente pour contrôler les activités des autorités de régulation en la matière dépend dans un premier temps de l'identité de ces autorités de régulation. En ce qui concerne la France, l'autorité pressentie pour devenir l'autorité nationale de surveillance du marché au sens de l'annexe III du règlement européen sur l'IA est la Commission nationale de l'informatique et des libertés (CNIL). En effet, le Conseil d'État dans une étude du 31 août 2022 relative à la gouvernance de l'IA préconise un renforcement des pouvoirs de la CNIL afin qu'elle puisse assumer ledit rôle¹⁰³⁴, cette dernière ayant par ailleurs déjà commencé à préparer sa nomination¹⁰³⁵. Si la CNIL est l'autorité nationale compétente au sens de la réglementation européenne sur l'IA, elle ne sera pas pour autant l'autorité de surveillance de la totalité des marchés visés par le règlement, en particulier ceux listés à

¹⁰³³ « opérateur économique : le fabricant, le mandataire, l'importateur, le distributeur, le prestataire de services d'exécution des commandes ou toute autre personne physique ou morale soumise à des obligations liées à la fabrication de produits, à leur mise à disposition sur le marché ou à leur mise en service conformément à la législation d'harmonisation applicable de l'Union; » - Article 3.13) du règlement 2019/1020 sur la surveillance du marché et la conformité des produits.

¹⁰³⁴ Le Conseil d'État, « S'engager dans l'intelligence artificielle pour un meilleur service public », (30 août 2022), en ligne : *Conseil d'État* <<https://www.conseil-État.fr/actualites/s-engager-dans-l-intelligence-artificielle-pour-un-meilleur-service-public>>.

¹⁰³⁵ « Création d'un service de l'intelligence artificielle à la CNIL et lancement des travaux sur les bases de données d'apprentissage | CNIL ».

l'annexe II de la proposition européenne, ni même la seule autorité en charge du contrôle administratif *ex post*¹⁰³⁶.

Bien que la CNIL soit une autorité administrative indépendante, le juge administratif n'est pas pour autant la seule juridiction compétente à pouvoir contrôler son action. En effet, le contentieux lié à l'action des autorités de régulation telles que la CNIL se partage entre le juge administratif et le juge judiciaire selon le type de contentieux en l'espèce¹⁰³⁷. Dans le cadre de la CNIL, plus spécifiquement des recours à l'encontre de ses décisions au titre de sa mission de contrôle et de régulation, la juridiction compétente en premier et dernier ressort est le Conseil d'État¹⁰³⁸. De plus, le Conseil d'État est également la juridiction compétente en cas de recours contre des actes de « droit souple » qui seraient pris par la CNIL¹⁰³⁹, notamment l'élaboration d'avis ou de recommandations sur le choix des preuves techniques de conception dans les systèmes d'IA à haut risque¹⁰⁴⁰. Le Conseil d'État a donc la compétence de contrôler les actions de la CNIL, en tant qu'autorité de surveillance du marché des systèmes d'IA, et ainsi potentiellement de se prononcer sur les choix relatifs aux preuves techniques réalisés par la CNIL. Par conséquent, le Conseil d'État peut contribuer à la détermination des preuves techniques pertinentes au regard du règlement européen sur l'IA mais aussi au regard des droits fondamentaux (le principe de non-discrimination par exemple).

B- Le contenu du contrôle juridictionnel

Annonce de plan. Grâce à son rôle d'autorité de surveillance des marchés, l'autorité administrative française chargée du contrôle *ex post* des systèmes d'IA, très probablement la CNIL, sera amenée à produire différents documents dans le but de contribuer à la bonne

¹⁰³⁶ Voir Sous-section 1 : Le contrôle administratif de la .

¹⁰³⁷ L'intérêt premier est celui d'une bonne administration de la justice. Conseil constitutionnel, 23 janvier 1987, n° 86-224. Par. 16. Le juge judiciaire est notamment compétent lorsque les décisions de l'autorité administrative indépendante « *tendent à régler des litiges entre opérateurs économiques privés* » - « Régimes législatifs de répartition des compétences – Régimes dérogeant en faveur de l'ordre juridictionnel judiciaire », 2015.

¹⁰³⁸ Code de justice administrative, article R311-1.4° « *Le conseil d'État est compétent pour connaître en premier et dernier ressort : [...] 4° Des recours dirigés contre les décisions prises par les organes des autorités suivantes, au titre de leur mission de contrôle ou de régulation : [...] - la Commission nationale de l'informatique et des libertés ;* »

¹⁰³⁹ BENEZECH Ludovic, « "L'affermissement de la justiciabilité des actes de droit souple devant le Conseil d'État", Droit administratif, LexisNexis, 2020, n° 2, p. 33 », *Droit administratif*, 2020.

¹⁰⁴⁰ Sur la compétence du Conseil d'État concernant les recours à l'encontre des actes de « droit souple » élaborés par les autorités de régulation de manière générale, voir : Conseil d'État, 21 mars 2016, n° 368082. Conseil d'État, 21 mars 2016, n° 390023. Pour une application aux actes de la CNIL, voir Conseil d'État, 16 octobre 2019, n° 433069.

application de la réglementation européenne. Sur le sujet des preuves techniques de conception ou sur tout autre sujet, l'autorité de contrôle peut exercer une influence considérable non seulement au travers de décisions individuelles prises à l'issue d'un contrôle de conformité mais aussi par le biais de l'élaboration d'actes réglementaires ou non-réglementaires **(1)**. Pour chacun de ces actes, le juge administratif dispose de pouvoirs de contrôle lui permettant potentiellement de contrôler les choix probatoires de l'autorité administrative, voire d'imposer les siens dans certaines situations **(2)**.

1. Les décisions susceptibles de faire l'objet d'un contrôle par le juge administratif

L'élaboration d'actes influençant la détermination du système probatoire des systèmes d'IA. Au-delà de son action de contrôle *ex post*, l'autorité nationale en charge de l'application de la réglementation sur l'IA a la capacité de produire des actes réglementaires ou des actes de « droit souple » afin de contribuer à la bonne application de la réglementation européenne sur les systèmes d'IA. Il peut s'agir de lignes directrices¹⁰⁴¹, de guide pratiques¹⁰⁴² ou bien d'outils¹⁰⁴³ mis à disposition des acteurs soumis à la réglementation, comme c'est déjà le cas de la CNIL dans son rôle d'autorité nationale de la protection des données personnelles¹⁰⁴⁴. Cette prérogative de production d'actes est cruciale pour préciser certaines notions floues dans des textes juridiques ou détailler la mise en application de certaines exigences. Ces actes sont le résultat d'une certaine interprétation desdites notions et/ou exigences juridiques, par l'autorité administrative. Dans le cadre du règlement sur l'IA, le choix des preuves techniques pertinentes pour attester de la conformité juridique des systèmes d'IA à haut risque est un élément absent du règlement européen¹⁰⁴⁵ et qui ne sera probablement pas traité exhaustivement par les normes européennes harmonisées¹⁰⁴⁶. La problématique des éléments techniques de preuve nécessitera des précisions, aussi bien sur l'interprétation des exigences que sur leur mise en application, et les actes produits par l'autorité de surveillance pourraient répondre à ce besoin de façon adéquate. Par conséquent, par le biais du contrôle des actes produit par l'autorité de

¹⁰⁴¹ « Lignes directrices et recommandations de la CNIL | CNIL ».

¹⁰⁴² « Le guide du délégué à la protection des données | CNIL ».

¹⁰⁴³ « Outil PIA ».

¹⁰⁴⁴ « Comme il a été indiqué à l'occasion de l'examen des pouvoirs des Autorités administratives indépendantes, celles qui sont dotées d'un pouvoir réglementaire sont la CNIL, l'AMF, l'ARCEP, le CSA et la CRE. » - *Les autorités administratives indépendantes : évaluation d'un objet juridique non identifié*, Conseil d'État.

¹⁰⁴⁵ Voir §2. La délégation des objectifs de qualité aux normes techniques.

¹⁰⁴⁶ Voir Chapitre 2 : Le contrôle des preuves de conception.

surveillance des marchés, le Conseil d'État peut évaluer les éléments techniques de preuves considérés comme pertinents par l'autorité administrative. Dès lors, le Conseil d'État peut contribuer indirectement aux choix des preuves juridiques pertinentes dans le cadre de la certification des systèmes d'IA à haut risque.

Les limites de l'influence des actes de la CNIL sur les éléments techniques de preuve.

L'étendue du pouvoir de contrôle du Conseil d'État est intimement liée à la capacité de l'autorité nationale de surveillance du marché relatif aux systèmes d'IA, dans notre hypothèse la CNIL, à produire des recommandations ou des actes sur les éléments techniques de preuve de conception. En d'autres termes, plus la CNIL peut se prononcer sur la question des preuves pertinentes, plus le Conseil d'État aura de probabilités d'être saisi sur ces questions par le biais d'un recours. Cependant, la CNIL possède des limites quant au contenu des recommandations et/ou actes qu'elle publie. Par exemple, en 2019, la CNIL publie des lignes directrices relatives à l'application de l'article 82 de la loi du 6 janvier 1978, portant notamment sur le caractère libre du consentement des internautes concernant les cookies et autres traceurs¹⁰⁴⁷. Dans ce document de droit souple, la CNIL posait une interdiction de la pratique des « cookies walls¹⁰⁴⁸ ». Cette interdiction se fondait exclusivement sur une interprétation de la notion de consentement libre prévue par le RGPD¹⁰⁴⁹. Par une décision du 19 juin 2020, le Conseil d'État a annulé ladite ligne directrice au motif qu'un acte de droit souple ne peut en aucun cas énoncer une interdiction générale et absolue¹⁰⁵⁰.

À la lumière de cet exemple, il est intéressant de se demander comment le Conseil d'État aborderait des lignes directrices de la CNIL qui interdiraient le recours à certaines métriques en raison de leur inadéquation avec les exigences juridiques imposées par le règlement européen. Par exemple, la CNIL pourrait considérer que la métrique de l'égalité d'opportunités¹⁰⁵¹, à savoir le fait que le logiciel d'IA présente un taux d'erreur identique pour deux catégories de personnes, n'est pas une information suffisamment claire et complète sur la performance du

¹⁰⁴⁷ Délibération n° 2019-093 du 4 juillet 2019 portant adoption de lignes directrices relatives à l'application de l'article 82 de la loi du 6 janvier 1978 modifiée aux opérations de lecture et écriture dans le terminal d'un utilisateur (notamment aux cookies et autres traceurs)

¹⁰⁴⁸ « pratique qui consiste à bloquer l'accès à un site web ou à une application mobile pour qui ne consent pas au dépôt ou à la lecture, sur son terminal, de traceurs de connexion » - Conseil d'État, 19 juin 2020, n° 434684. Par. 9.

¹⁰⁴⁹ *Ibid.* Par. 10.

¹⁰⁵⁰ *Ibid.*

¹⁰⁵¹ Cette métrique est plus connue sous sa dénomination anglaise *equality of odds*. Pour plus de détails, voir B-Le choix d'un indicateur adéquat à l'exigence légale.

logiciel pour l'utilisateur¹⁰⁵². Par analogie, une telle interdiction générale fondée sur l'interprétation d'une exigence du règlement européen sur l'IA serait probablement considérée comme un excès de pouvoir de la part de la CNIL. Celle-ci devrait se contenter de fournir les éléments contextuels permettant d'établir un seuil statistique suffisant ou un indicateur adapté. Il semble donc que la CNIL n'excède pas ses compétences si elle vient à aborder la question des éléments techniques de preuve pertinents pour la certification des systèmes d'IA à haut risque. Par conséquent, cela confirme que le Conseil d'État pourra se prononcer sur la question des éléments techniques de preuve dans le cadre des potentiels recours contre les actes de la CNIL (en tant qu'autorité de surveillance), et ne limitera donc pas son analyse à la seule question de la compétence de l'autorité.

Les décisions individuelles. Le second type d'activité pouvant faire l'objet d'un recours juridictionnel sont les décisions individuelles prises par la CNIL. Dans le cadre d'une sanction, le recours peut porter sur la décision de sanction prise par la CNIL mais également sur l'absence de sanction prise par la CNIL à la suite d'un contrôle par exemple. La décision prise par l'autorité nationale de régulation des systèmes d'IA à la suite d'un contrôle de conformité, quelle que soit cette décision, est susceptible d'un recours devant la juridiction administrative. De plus, les décisions d'agréments ou d'autorisation produite par des autorités administratives de régulation entrent dans la catégorie de décisions individuelles¹⁰⁵³. Le juge administratif apparaît donc parfaitement en capacité de réaliser un contrôle juridictionnel sur ces décisions.

2. Le contenu du contrôle exercé par le juge administratif

Les mesures d'instruction à disposition du juge. Tout d'abord, le juge administratif peut ordonner une enquête sur les faits, que ce soit à la demande des parties ou d'office, dès lors qu'une telle mesure est utile à l'instruction de l'affaire¹⁰⁵⁴. Les parties en possession des éléments visés par les mesures d'instruction du juge doivent les ajouter au dossier. En cas de refus, le juge lui-même peut ajouter les éléments utiles au dossier contradictoire¹⁰⁵⁵ sans quoi il

¹⁰⁵² « Les systèmes d'IA à haut risque sont accompagnés d'une notice d'utilisation dans un format numérique approprié ou autre, contenant des informations concises, complètes, exactes et claires, qui soient pertinentes, accessibles et compréhensibles pour les utilisateurs. » - « Règlement sur l'intelligence artificielle », 2024. Articles 13.2 et 13.3.

¹⁰⁵³ Voir notamment les décisions d'agrément prises par le CSA : Conseil d'État, 11 avril 2014, n° 348972.

¹⁰⁵⁴ Article R. 623-1 du Code de la justice administrative.

¹⁰⁵⁵ Conseil d'État, Assemblée, 06/11/2002, 194295, M Sun Myung Moon, .

ne peut pas en tenir compte dans sa décision¹⁰⁵⁶. Ces mesures d’instruction apparaissent indispensables dans le cadre de dossiers complexes comme la conformité technique des systèmes d’IA. Le juge peut ainsi recourir à des expertises¹⁰⁵⁷, à des enquêtes à la barre ou sur les lieux¹⁰⁵⁸ ou encore à l’avis technique¹⁰⁵⁹.

Contenu du contrôle. Dans le contexte du contrôle des décisions administratives de la CNIL par le juge administratif, les choix concernant les preuves techniques ne se trouveront pas dans le contrôle de la légalité externe mais vraisemblablement dans le contrôle de la légalité interne desdits actes. En d’autres termes, la problématique probatoire se trouve davantage dans le fond de la décision administrative que dans la forme. En tant qu’autorité administrative indépendante, la CNIL bénéficie d’un large pouvoir d’appréciation dans l’exercice de ses compétences¹⁰⁶⁰. Dès lors, le contrôle juridictionnel par le juge administratif est allégé¹⁰⁶¹. Dans le cas d’un refus de sanction par la CNIL après un contrôle d’un système d’IA à haut risque, le juge ne contrôle pas le détail des motifs justifiant la décision de l’autorité. Il ne procède qu’à un contrôle dit « d’erreur manifeste »¹⁰⁶², aussi appelé contrôle restreint¹⁰⁶³. En d’autres termes, seule une erreur grossière de la part de l’autorité administrative pourra conduire à l’annulation de sa décision¹⁰⁶⁴. Par conséquent, les éléments précis pris en compte par la CNIL lors du contrôle et qui ont conduit au refus de sanctionner se situent hors du spectre de contrôle du juge administratif. Dans cette situation, le juge n’aura donc pas la compétence de vérifier l’examen des éléments techniques de preuve de conception réalisé par la CNIL.

En ce qui concerne les décisions individuelles autres que les refus, le juge exerce le plus souvent un contrôle dit de « proportionnalité »¹⁰⁶⁵, aussi appelé le contrôle normal¹⁰⁶⁶. Le juge administratif regarde alors si la décision prise par l’administration est nécessaire, adaptée et proportionnée¹⁰⁶⁷. Il s’agit donc d’un contrôle plus précis qui offre un plus grand pouvoir au

¹⁰⁵⁶ Conseil d’État, 23 décembre 2016, n° 405791.

¹⁰⁵⁷ Articles R. 621-1 et suivants du code de justice administrative

¹⁰⁵⁸ Article R. 623-1 du code de justice administrative

¹⁰⁵⁹ Article R. 625-2 du code de justice administrative

¹⁰⁶⁰ Voir notamment Conseil d’État, 21 octobre 2022, n° 459254.

¹⁰⁶¹ SIMMONEAUX Antoine, « Le contentieux du Data Protection Officer devant le juge administratif », *AJDA*, 2022.

¹⁰⁶² Conseil d’État, 21 octobre 2022, n° 459254.

¹⁰⁶³ ROUQUETTE Rémi, « Petit traité du procès administratif. Chapitre 222 - Typologie des recours contentieux », *Praxis Dalloz*, 2021.

¹⁰⁶⁴ Dans le cas d’un refus de sanctionner, le recours contentieux prend la forme d’un recours en excès de pouvoir. En revanche, la décision de sanctionner peut faire l’objet d’un recours de plein contentieux.

¹⁰⁶⁵ CONSEIL D’ÉTAT, « Le juge administratif et les autorités de régulation économique - Dossier thématique », 2016.

¹⁰⁶⁶ ROUQUETTE Rémi, *op. cit.*

¹⁰⁶⁷ CONSEIL D’ÉTAT, *op. cit.*

juge administratif de se prononcer sur la question des preuves techniques. Au travers de ces contrôles, le juge est amené à vérifier l'exactitude matérielle et la qualification juridique des faits sur lesquels se fondent les motifs de la décision¹⁰⁶⁸. Face à une décision de non-conformité, il sera nécessairement amené à vérifier la présence ou l'absence des preuves techniques produites par la CNIL ou indiquées dans la documentation technique par le fournisseur du système d'IA. Par exemple, un fournisseur de système d'IA à haut risque a considéré qu'un niveau de biais entre les hommes et les femmes dans les données de conception, après la réalisation des tests et des mesures de minimisation, égal à 35% est un risque résiduel qu'il qualifie d'acceptable. Il a ensuite obtenu la certification de son produit à l'issue d'une procédure d'évaluation *ex ante* interne. Au cours de la réalisation d'un contrôle *ex post*, la CNIL considère que 35% de biais dans les données d'entraînement est en réalité un risque résiduel inacceptable, et ainsi sanctionne le fournisseur du logiciel. Si le fournisseur choisit de contester cette décision de sanction, le juge devra alors se prononcer sur l'acceptabilité d'un score de biais de 35% entre les hommes et les femmes. Ce faisant, le Conseil d'État contribuera à déterminer le seuil statistique adéquat d'acceptabilité pour les risques résiduels de discrimination algorithmique. Si le juge administratif considère qu'il s'agit d'un risque résiduel acceptable, alors les autres fournisseurs de systèmes d'IA, dans un contexte similaire, pourront ajouter cette information dans leurs procédures de test. Par ce type de décision, le juge administratif vient valider une interprétation quantitative, le seuil statistique en l'espèce, d'une notion qualitative floue, l'acceptabilité pour les risques résiduels, dans un contexte donné. Au travers de ce cas d'illustration apparaît clairement le rôle du juge dans l'élaboration d'un seuil statistique pertinent pour une exigence juridique donnée. Un raisonnement identique pourrait être mené avec le choix d'un indicateur d'équité plutôt qu'un autre (l'égalité des chances par rapport à la parité statistique par exemple). La qualification de l'indicateur comme pertinent confirmerait la légalité de cette interprétation de l'équité pour les autres fournisseurs commercialisant des produits dans un contexte similaire.

Conséquences du contrôle. Les conséquences du contrôle de l'action administrative par le juge administratif dépendent de la nature du recours en présence. Dans un premier temps, le recours en excès de pouvoir limite le juge dans ses compétences à l'annulation de l'acte contesté. En ce qui concerne l'étude des éléments techniques de preuve, le juge apparaît donc comme un acteur de contrôle du pouvoir d'influence de l'autorité administrative sur la sélection des preuves techniques considérées comme pertinentes. En tant qu'acteur, le juge ne dispose

¹⁰⁶⁸ CE, ass., 30 janv. 1959, Sieur Grange, req. no 33908, Lebon

pas d'une influence proactive sur le sujet. Par exemple, le juge administratif pourra annuler des lignes directrices déconseillant certains indicateurs et seuils statistiques aux fournisseurs de systèmes d'IA, ou des facteurs permettant d'effectuer ce choix, mais ne pourra en revanche pas remplacer l'interprétation de l'autorité par la sienne au travers de l'arrêt. Dans un second temps, le recours de plein contentieux offre au juge la possibilité de s'affranchir de cette position « passive » dans l'établissement du système probatoire. En effet, dans ce cas, le juge détient la capacité de réformer la décision qui lui est déférée¹⁰⁶⁹. Ce recours est notamment exercé face à des décisions individuelles produites par une autorité administrative. Ainsi, une sanction prononcée par la CNIL, le retrait d'un certificat européen de conformité par exemple, peut faire l'objet d'un recours de plein contentieux¹⁰⁷⁰. Le juge administratif ne peut toutefois pas aggraver la sanction administrative prononcée¹⁰⁷¹, transformant par exemple un prononcé de mesures correctives en un retrait d'un système d'IA à haut risque du marché. Si le litige repose totalement ou partiellement sur l'analyse du fonctionnement technique du système d'IA à haut risque, le juge peut donc parfaitement contribuer activement à la détermination des preuves techniques adéquates non seulement en annulant mais aussi en allégeant la décision prise par l'autorité de contrôle.

CONCLUSION DE LA SECTION II :

Le contrôle *ex post* des systèmes d'IA est la troisième et dernière étape de la détermination de la conformité légale, et des preuves techniques qui la constituent. Les acteurs de cette mission de contrôle *ex post* se voient alors confier le pouvoir de confronter leur vision des preuves techniques pertinentes à celles établies par les précédents protagonistes, lors de l'élaboration des normes techniques et lors du contrôle *ex ante* de conformité. Du fait de leur place finale dans la chaîne de détermination de la conformité et de leur capacité de sanction, les contrôleurs *ex post*, en particulier le juge administratif, peuvent imposer leurs conceptions des preuves techniques pertinentes et détiennent donc une influence substantielle forte. En outre, le contrôle *ex post* est une étape importante d'élargissement de la place des preuves techniques dans le droit. Premièrement, le contrôle *ex post* de la conformité fait entrer de nouveaux acteurs que sont les autorités de surveillance du marché et les juridictions, jusqu'alors quasiment absents du système de mise en œuvre et de contrôle des systèmes d'IA. Malgré la mise à l'écart

¹⁰⁶⁹ Cour de justice de l'Union européenne, 7 décembre 2023, n° C-634/21. DOUVILLE Thibault, « Du renforcement de l'effectivité du RGPD : contrôle juridictionnel plénier de la décision de rejet d'une réclamation prononcée par une autorité de contrôle compétente en matière de protection des données », *RTD Com.*, 2024

¹⁰⁷⁰ Conseil d'État, 16 février 2009, n° 274000.

¹⁰⁷¹ Conseil d'État, 6 avril 2016, n° 374224.

persistante et dommageable de la société civile comme acteur de cette procédure, ces nouveaux protagonistes permettent d'offrir une vision différente et bienvenue des éléments techniques de preuves pertinents des systèmes d'IA en tant que garantie du respect de la sécurité, de la santé et des droits fondamentaux des résidents européens. Deuxièmement, le contrôle *ex post* s'étend s'entend en-dehors des seules exigences du règlement européen sur l'IA à haut risque. Les preuves techniques de fonctionnement, centrales dans la détermination de la conformité des systèmes d'IA à haut risque, sont désormais des outils à disposition pour attester du respect de nombreux autres règlements. Cette extension est notamment symbolisée par l'implication et l'octroi de prérogatives aux organismes et autorités en charge du respect des droits fondamentaux en droit européen. Troisièmement, les autorités de surveillance et les juridictions sont toutes les deux des acteurs présentant des garanties d'indépendance considérables. Contrairement aux acteurs du contrôle *ex ante* et aux acteurs en charge de l'élaboration des normes techniques, les acteurs du contrôle *ex post* de la conformité représentent la seule étape de contrôle des preuves techniques où les acteurs les plus présents sont indépendants. Pour toutes ces raisons, le contrôle *ex post* de la conformité est une étape cruciale, mais malheureusement non systématique, de la détermination des preuves techniques pertinentes en droit européen de l'IA.

CONCLUSION DU CHAPITRE 2 :

L'étape de contrôle de la conformité des systèmes d'IA occupe une place déterminante dans la détermination des preuves techniques pertinentes au regard du règlement européen sur l'IA. C'est au cours du contrôle de conformité que sera évaluée la pertinence des éléments techniques de preuve proposés par le fournisseur du système d'IA. L'analyse de ce processus révèle que les acteurs intervenant lors du contrôle *ex ante* ont un pouvoir sur ces choix probatoires bien plus important que les acteurs du contrôle *ex post*. En effet, leur interprétation des exigences du règlement sur l'IA pour les traduire en éléments de preuves techniques est systématique car le contrôle *ex ante* est une étape obligatoire pour tous les systèmes d'IA à haut risque. À l'inverse, le contrôle *ex post* n'est que ponctuel. Les acteurs du contrôle *ex post* ne pourront donc pas faire valoir aussi fréquemment leur interprétation des exigences que ceux du contrôle *ex ante*. De plus, les acteurs du contrôle *ex ante* caractérisent les preuves techniques pertinentes avant la mise sur le marché du logiciel d'IA à haut risque. De ce fait, le système d'IA à haut risque sera déployé selon leur interprétation de la conformité juridique, et les décisions reçues par les personnes affectées retranscriront leur conception de la qualité de fonctionnement européenne. Au contraire, les contrôleurs *ex post* n'interviennent que de manière réactive, c'est-à-dire après la production de résultats potentiellement préjudiciables. Cependant, en tant que dernières étapes de vérification de la conformité, le contrôle *ex post* permet aux acteurs qui le réalisent de mettre en avant leur interprétation de la conformité des logiciels d'IA par rapport à celle retenue par les contrôleurs *ex ante*. Enfin, notre étude des acteurs du contrôle de conformité met en lumière que les acteurs ayant la plus grande influence sur la définition d'une preuve technique pertinente sont également ceux qui présentent le moins de garantie d'indépendance. En effet, le principal contrôleur *ex ante* est le fournisseur du système d'IA lui-même qui est par nature totalement partial car il évalue la pertinence de ces propres choix probatoires. Par conséquent, malgré la place centrale qu'occupent les éléments techniques de preuve dans le nouveau règlement sur l'IA, le choix des preuves est réalisé principalement par l'acteur offrant le moins de garantie d'indépendance de tout le processus de contrôle de la conformité des systèmes d'IA à haut risque.

CONCLUSION DU TITRE 2 :

Les éléments techniques de preuve cristallisent à la fois les enjeux centraux de fond (équité, transparence, fiabilité, etc.) et les enjeux de forme (précision et articulation des exigences) de l'encadrement de l'IA. De ce fait, ils occupent une place centrale dans le règlement européen sur l'IA. Ces choix probatoires contribuent, de façon importante, à définir la notion de conformité européenne imposées aux logiciels d'IA à haut risque et, par conséquent, le niveau de protection des droits et libertés fondamentaux garanti par le texte. La détermination des preuves techniques pertinentes juridiquement est un processus essentiel dans le règlement européen et y contribuer confère un pouvoir majeur sur le droit européen de l'IA. Ce pouvoir est partagé entre les différents acteurs qui participent à ce processus. En fonction des rôles et des prérogatives juridiques octroyés par le règlement sur l'IA, chaque acteur bénéficie d'un pouvoir plus ou moins grand sur le choix des preuves techniques pertinentes. Entre la phase d'élaboration et la phase de contrôle desdites preuves, il apparaît que la grande majorité de ce pouvoir de détermination se trouve entre les mains d'acteurs privés, dépourvus de légitimité démocratique et de garanties d'indépendance. À ce jour, les autres acteurs, comme les autorités de surveillance de marché et les juridictions, se révèlent être des contributeurs incertains et mineurs dans la réalisation de ces choix probatoires.

CONCLUSION DE LA PARTIE 1 :

Avec le règlement sur l'IA, le législateur européen met en place des obligations visant à garantir que les systèmes d'IA à haut risque fonctionnent de façon sûre et respectueuse des droits et libertés fondamentaux de l'Union européenne. Pour atteindre cet objectif de qualité, le règlement met en avant la réalisation de tests techniques sur les logiciels dans le but de démontrer que le fonctionnement est conforme aux exigences, et ainsi obtenir l'indispensable certificat européen. En recourant à une approche aussi technique que la certification, le législateur européen fait reposer sur les spécifications techniques la cohérence de son régime juridique. Parmi tous les éléments que le règlement européen sur l'IA délègue aux normes techniques, les éléments techniques qui seront utilisés comme preuves de conception apparaissent comme la clé de voûte du droit des systèmes d'IA. Le choix de ces preuves techniques qui permettront d'obtenir la certification réunit en effet l'ensemble des enjeux juridiques majeurs de l'encadrement de l'IA. La détermination des preuves techniques pertinentes devient alors un processus central dans la définition et la cohérence de la qualité européenne que garantit le règlement. Situées hors du texte, les preuves techniques se retrouvent hors de vue des acteurs du processus législatif. Du fait de leur technicité, le choix des preuves techniques est confié majoritairement à des acteurs privés et, dans une moindre mesure, à des acteurs publics indépendants tels que les autorités de surveillance du marché et les juridictions. La démonstration de la place centrale qu'occupent les preuves dans le règlement européen sur l'IA illustre ainsi la porosité de la frontière entre la technique et le juridique.

Il est désormais clair que les preuves juridiques liées au fonctionnement des systèmes d'IA ne sont pas que de simples détails techniques d'un règlement mais bien un élément central et indispensable de celui-ci. Toutefois, les conséquences des preuves techniques choisies pour certifier les systèmes d'IA à haut risque ne se limitent pas au règlement européen. Ces preuves continuent à avoir des conséquences tout au long du cycle de vie du système d'IA. En tant que garantie de fonctionnement, ces preuves techniques sont réutilisées lors du déploiement du système. Elles sont donc vouées à s'articuler avec d'autres régimes juridiques applicables aux technologies d'IA.

PARTIE 2 : LES CONSEQUENCES DES PREUVES DE CONCEPTION LORS DE L'UTILISATION DU SYSTEME D'IA

Introduction. Pour parvenir à établir un marché européen de l'IA sûr, l'articulation entre le règlement européen sur l'IA et les législations relatives aux secteurs d'utilisation de ces logiciels est une condition indispensable¹⁰⁷². En effet, la mise en place du nouveau texte européen sur l'IA, en particulier le processus de certification des systèmes d'IA à haut risque, n'a d'intérêt que s'il entraîne des conséquences lors du déploiement. Par conséquent, les preuves de conception, à savoir les éléments de preuves utilisés pour obtenir ladite certification, auront des effets directs ou indirects sur les législations relatives à la phase d'utilisation. La place de ces preuves dans la réglementation sur l'IA ne peut donc être circonscrite à leur texte d'origine. Dans la mesure du raisonnable¹⁰⁷³, notre périmètre d'étude doit ainsi être étendu afin d'offrir une représentation la plus fidèle possible de la manière dont les preuves de conception d'un système d'IA à haut risque s'articulent avec les législations relatives à son utilisation. Les interactions entre les preuves de conception et les exigences juridiques applicables à l'utilisation de ces logiciels sont autant d'informations supplémentaires concernant les enjeux juridiques et la place occupée par ces éléments techniques de preuves. L'analyse portera particulièrement sur la possibilité de réutilisation de ces preuves au sein des législations relatives à l'utilisation des systèmes d'IA.

Annonce de plan. L'ampleur et la complexité des articulations de ces éléments de preuves dans l'encadrement de l'utilisation des systèmes d'IA à haut risque seront ainsi étudiées au travers des régimes de responsabilité *ex post* du fait de l'IA (**Titre 1**). En contribuant de façon déterminante à certifier un système d'IA à haut risque, les preuves de conception accroissent la confiance que les déployeurs et utilisateurs accordent aux résultats produits par ledit système.

¹⁰⁷² « Le caractère horizontal de la proposition requiert une cohérence parfaite avec la législation de l'Union existante applicable aux secteurs dans lesquels des systèmes d'IA à haut risque sont déjà utilisés ou sont susceptibles de l'être dans un avenir proche » - « Règlement sur l'intelligence artificielle », 2024. Version initiale proposée par la Commission européenne. P. 4.

¹⁰⁷³ La présente thèse n'a ni la vocation ni la prétention de proposer une description exhaustive du phénomène. L'ampleur de la tâche à accomplir ainsi que le caractère prospectif de cette recherche limite en pratique l'objet d'étude.

Afin d'explorer cet effet, il sera étudié la manière dont les preuves de conception peuvent s'articuler avec le traitement judiciaire de la preuve produite par un système d'IA (**Titre 2**).

TITRE I : LA PREUVE REUTILISEE CONTRE UN SYSTEME D'IA : LA RESPONSABILITE EX POST DU FAIT DE L'IA

Introduction. Au sein de l'ordre juridique, la responsabilité civile désigne « *toute obligation de répondre civilement du dommage que l'on a causé à autrui, c'est-à-dire de le réparer en nature ou par équivalent.* »¹⁰⁷⁴. Celle-ci se subdivise ensuite en différents régimes (la responsabilité contractuelle, la responsabilité délictuelle) qui possèdent eux-mêmes leurs propres distinctions (responsabilité du fait des choses inanimées, responsabilité du fait des parents, etc.)¹⁰⁷⁵. La responsabilité du fait de l'IA n'apparaît pas parmi ces catégories consacrées dans le code civil français¹⁰⁷⁶. Pour cause, l'expression, relativement récente, ne renvoie pas à un régime de responsabilité spécifique mais désigne l'ensemble des règles de responsabilité applicables lorsqu'un système d'IA cause un dommage. Outre la réparation du préjudice, la responsabilité civile présente une fonction préventive, en ce qu'elle incite les personnes à ne pas causer de dommages par crainte des sanctions pécuniaires¹⁰⁷⁷. Cette finalité prophylactique est également au cœur du règlement européen sur l'IA dont l'un des objectifs principaux est l'établissement d'un niveau élevé de protection de la santé, la sécurité et des droits fondamentaux contre les effets néfastes des systèmes d'IA¹⁰⁷⁸. Dès lors, le règlement sur l'IA et les régimes de responsabilité civile du fait de l'IA représentent deux instruments, l'un *ex ante* et l'autre *ex post*, au service d'un même objectif. Par conséquent, la responsabilité civile du fait de l'IA est un terrain propice aux interactions avec le règlement européen sur l'IA. Dès lors, les preuves techniques de certification, qui jouent déjà un rôle central dans le règlement européen sur l'IA, peuvent également avoir des conséquences sur la responsabilité civile du fait de l'IA. Plus précisément, il convient de s'interroger sur la place que peuvent occuper ces preuves dans le cadre des régimes de responsabilité du fait de l'IA.

¹⁰⁷⁴ CORNU Gérard, *Vocabulaire juridique*, 2022.

¹⁰⁷⁵ VINEY Genevieve, *Introduction à la responsabilité*, LGDJ, 2019.

¹⁰⁷⁶ Titre III : Des sources d'obligations (articles 1100 à 1303-4).

¹⁰⁷⁷ « La notion de « responsabilité » joue un rôle important dans notre vie quotidienne à un double titre : d'une part, elle garantit qu'une personne ayant subi un préjudice ou un dommage est en droit de demander réparation à la partie tenue pour responsable de ce préjudice ou de ce dommage, et, d'autre part, elle donne aux personnes des incitations économiques à ne pas causer de préjudice ou de dommage en premier lieu » - PARLEMENT EUROPÉEN, « Un régime de responsabilité civile pour l'intelligence artificielle », 2020. Considérant 1. Voir également, parmi d'autres : LE TOURNEAU Philippe, *Droit de la responsabilité et des contrats - Régimes d'indemnisation*, Dalloz, . P. 13.

¹⁰⁷⁸ « Règlement sur l'intelligence artificielle », 2024. Considérant 1.

Annonce du plan. L'étude des conséquences des preuves techniques de conception sur la responsabilité *ex post* de l'IA nécessite, au préalable, d'étudier les mécanismes d'interaction entre les exigences relatives à la responsabilité *ex ante* et celles relatives à la responsabilité *ex post* du fournisseur du système d'IA (**Chapitre 1**). Ensuite, il s'agira d'étudier la manière dont les éléments techniques de preuves de la conformité, considérés comme pertinents par les acteurs en charge de la mise en conformité des systèmes d'IA à haut risque, peuvent également être réutilisées comme éléments de preuves des faits constitutifs de ces régimes de responsabilité *ex post* (**Chapitre 2**).

CHAPITRE 1 : L'ARTICULATION ENTRE LE REGLEMENT EUROPEEN SUR L'IA ET LA RESPONSABILITE EX POST DU FOURNISSEUR DE SYSTEMES D'IA

Introduction. Pour le fournisseur d'un système d'IA à haut risque, l'obtention du certificat européen de conformité est une étape indispensable sans laquelle il ne peut ni commercialiser ni mettre en service son produit. Le règlement européen sur l'IA se superpose et s'articule avec d'autres régimes de responsabilité déjà existants. Dès lors, l'engagement de la responsabilité du fournisseur de système d'IA n'est pas limitée à l'obtention et à la conservation du marquage CE sur ses produits. Entre le droit national et le droit européen, les régimes de responsabilité *ex post* applicables aux systèmes d'IA sont nombreux. De plus, les nombreux contextes d'utilisation des systèmes d'IA, et donc de survenance du préjudice, accroissent encore davantage les régimes spécifiques de responsabilité applicables à l'IA. Traiter de l'application et l'articulation de l'ensemble de ces réglementations sur l'ensemble des acteurs intervenant dans le cycle de vie d'un logiciel d'IA se révélerait extrêmement difficile, si ce n'est impossible. En outre, l'analyse du règlement sur l'IA que nous avons précédemment réalisée prenait le parti de ne pas traiter des particularités propres à certains secteurs (la réglementation spécifique aux systèmes d'IA dans la santé, les transports, etc.). Dans un souci de cohérence, l'étude des régimes de responsabilité *ex post* reprend cette approche horizontale, également présente dans le règlement européen sur l'IA. En effet, il s'agit, tout d'abord, d'écarter du périmètre de recherche les régimes de responsabilité *ex post* propres à certains secteurs d'application. Ensuite, l'approche horizontale implique d'intégrer dans notre recherche la protection des droits fondamentaux des individus, enjeu transversal préalablement abordé au travers du règlement sur l'IA. Par conséquent, l'analyse se concentre sur l'articulation du régime de responsabilité *ex ante* de l'IA avec les régimes de responsabilité *ex post* applicables aux systèmes d'IA et les droits fondamentaux des individus affectés par le déploiement du système d'IA.

Annonce de plan. En élaborant des exigences relatives au fonctionnement des systèmes d'IA à haut risque, le règlement européen va interagir avec les régimes de responsabilité *ex post* en cas de dommage causé par ces logiciels. L'obtention de la certification européenne a ainsi des conséquences directes sur la responsabilité *ex post* du fournisseur dudit système d'IA à haut risque (**Section I**). Au-delà des régimes de responsabilité applicables traditionnellement aux fabricants des produits, le fournisseur d'un système d'IA à haut risque voit également sa

responsabilité *ex ante* s'articuler de façon inhabituelle avec le régime de protection des droits fondamentaux (**Section II**).

Section I : Les effets de la conformité ex ante du système d'IA sur la responsabilité ex post de son fournisseur

Introduction. Les régimes de responsabilité *ex post* possèdent, aussi bien en droit français qu'en droit européen, de façon constante trois éléments constitutifs : le fait générateur, le préjudice et le lien de causalité entre les deux¹⁰⁷⁹. L'utilisation d'un système d'IA expose les personnes affectées à un risque de préjudice. Toutefois, la nature de ces potentiels préjudices n'est pas nouvelle au regard du droit, d'une part, et celle-ci n'est pas liée au fonctionnement du système d'IA utilisé mais davantage à sa destination, d'autre part. Par exemple, un système d'IA de reconnaissance d'image utilisé pour diriger un véhicule autonome peut causer un préjudice patrimonial en cas d'erreur d'identification d'un obstacle (atteinte aux biens et/ou aux personnes) parce qu'il est utilisé pour guider un véhicule. La qualité de fonctionnement du système, garantie par les éléments techniques de preuve de conception, a des conséquences sur la cause et la fréquence des accidents mais pas sur la nature des préjudices qui en découlent. Ensuite, le lien de causalité est un élément intéressant qui présente de nombreux défis dans le cadre de la responsabilité *ex post* du fait de l'IA. Néanmoins, ce-dernier ne présente pas non plus de liens avec la qualité de fonctionnement du système d'IA, celle-ci n'ayant de conséquences que sur la cause du dommage. Par conséquent, ces deux éléments constitutifs de la responsabilité *ex post* n'interagissent pas avec les éléments techniques de preuve relatifs au fonctionnement du logiciel et ne sont donc pas pertinents à aborder dans notre étude. La conformité du système d'IA, et les éléments techniques de preuve utilisés pour en attester, ont des conséquences importantes sur le fait générateur des régimes de responsabilité *ex post*. Afin de continuer notre analyse sur la place des preuves techniques de conception des systèmes d'IA à haut risque, il est donc nécessaire de s'intéresser à la manière dont ces éléments peuvent contribuer à identifier l'existence, ou l'absence, d'un fait générateur de responsabilité civile.

¹⁰⁷⁹ VINEY Genevieve, *Introduction à la responsabilité*, LGDJ, 2019. BACACHE-GIBEILI Mireille, *Droit civil - Les obligations, la responsabilité civile extracontractuelle*, Economica, 2021. CARVAL Suzanne, VINEY Genevieve et JOURDAIN Patrice, *Les conditions de la responsabilité*, LGDJ, 2013. VINEY Genevieve, *Les conditions de la responsabilité*, LGDJ, 2013.

Annnonce de plan. Afin d'étudier la manière dont les preuves de conception s'articulent avec les faits générateurs constitutifs des régimes de responsabilité *ex post*, il convient de distinguer au préalable la responsabilité contractuelle du fournisseur de système d'IA (**Sous-section 1**) de sa responsabilité délictuelle (**Sous-section 2**).

Sous-section 1 : La défaillance contractuelle du fournisseur du système d'IA

Annnonce de plan. Au-delà des exigences imposées par le règlement européen sur l'IA, le fournisseur d'un système d'IA à haut risque est soumis à des obligations contractuelles relatives à son produit que sont l'obligation de délivrer un système d'IA conforme (§1) et absent de vices (§2). Le manquement à ces obligations par le fournisseur constitue une inexécution contractuelle, premier élément constitutif de la responsabilité civile contractuelle¹⁰⁸⁰.

§1. La délivrance non-conforme du système d'IA

Annnonce de plan. À la charge du fournisseur du logiciel, l'obligation de délivrance conforme du système d'IA à haut risque impose que celui-ci soit conforme à tous égards¹⁰⁸¹. La satisfaction de cette obligation inclut nécessairement la conformité du système d'IA à haut risque au règlement européen sur l'IA (A) et s'étend à l'ensemble des spécifications contractuelles (B).

A- La conception conforme du système d'IA délivré

L'étendue de l'obligation de délivrance. L'obligation de délivrance conforme du logiciel comprend de nombreuses exigences dont certaines font écho à celles imposées aux systèmes d'IA à haut risque par le règlement européen sur l'IA. Par exemple, le fournisseur doit

¹⁰⁸⁰ VINEY Genevieve, *Introduction à la responsabilité*, LGDJ, 2019. VINEY Genevieve, *Les conditions de la responsabilité*, LGDJ, 2013. CHABAS François, MAZEAUD Henri, MAZEAUD Jean et al., *Traité théorique et pratique de la responsabilité civile délictuelle et contractuelle*, LGDJ, 2013.

¹⁰⁸¹ LE TOURNEAU Philippe, *Contrats du numérique ; informatiques et électroniques*, Dalloz, 2022. P. 340.

communiquer à son client la documentation associée au logiciel, à savoir les modes d'emploi, les instructions, etc¹⁰⁸². Issue du droit national¹⁰⁸³, cette obligation est similaire à l'obligation pour le fournisseur du système d'IA à haut risque de fournir une notice d'utilisation :

« 2. Les systèmes d'IA à haut risque sont accompagnés d'une notice d'utilisation dans un format numérique approprié ou autre, contenant des informations concises, complètes, exactes et claires, qui soient pertinentes, accessibles et compréhensibles pour les déployeurs. »¹⁰⁸⁴

Néanmoins, les informations devant être contenues dans la notice d'utilisation imposée par le règlement européen ne sont ni listées de manière limitative ni de manière exhaustive. Le droit français impose des contraintes différentes, et donc supplémentaires, à celles prévues par le règlement européen, notamment l'obligation d'employer la langue française pour communiquer lesdites informations¹⁰⁸⁵. De la même manière, l'obligation de délivrance conforme s'applique à tous les systèmes d'IA, indépendamment du niveau de risque. En revanche, les informations énoncées par le texte européen sont toutes nécessaires pour satisfaire l'obligation de délivrance conforme car elles sont indispensables à l'obtention de la certification européenne.

La nécessité de délivrer un système d'IA à haut risque certifié. L'obligation de délivrance conforme est bien plus large pour le fournisseur d'un système d'IA à haut risque que son seul respect des exigences relatives aux systèmes d'IA à haut risque énoncées dans le règlement européen sur l'IA. Cependant, l'obtention du certificat européen, condition *sine qua non* de la mise sur le marché du système d'IA, est logiquement la première des conditions à remplir pour espérer délivrer un logiciel conforme. Ainsi, dans le cadre d'un contrat de consommation, le droit français impose que système d'IA délivré par le fournisseur soit notamment :

« propre à l'usage habituellement attendu d'un bien de même type, compte tenu, s'il y a lieu, de toute disposition du droit de l'Union européenne et du droit national ainsi que de toutes les normes techniques ou, en l'absence de telles normes techniques, des codes de conduite spécifiques applicables au secteur concerné ; »¹⁰⁸⁶

Pour un déployeur souhaitant utiliser le système d'IA dans une finalité à haut risque, la présence du marquage CE sur un système d'IA à haut risque est donc une caractéristique essentielle du

¹⁰⁸² *Ibid.* P. 247.

¹⁰⁸³ « L'obligation de délivrer la chose comprend ses accessoires et tout ce qui a été destiné à son usage perpétuel. » - Article 1615 du code civil.

¹⁰⁸⁴ « Règlement sur l'intelligence artificielle », 2024. Article 13.2.

¹⁰⁸⁵ « Loi n° 94-665 du 4 août 1994 relative à l'emploi de la langue française. », 1994.

¹⁰⁸⁶ Article L. 217-5.I.1° du code de la consommation.

logiciel. L'obtention de la certification européenne nécessite le respect des obligations documentaires, processuelles et relatives à son fonctionnement¹⁰⁸⁷. Ainsi, la conception d'un système d'IA à haut risque qui ne présenterait pas un « *niveau approprié d'exactitude, de robustesse et de cybersécurité* »¹⁰⁸⁸, ou encore dont la présence de biais discriminatoires exposerait à des risques inacceptables selon le droit à la non-discrimination¹⁰⁸⁹, entraînerait un refus d'octroi de la certification, et par conséquent une non-conformité du logiciel lors de sa délivrance. Par conséquent, le fournisseur d'un système d'IA à haut risque qui apposerait le marquage CE sans respecter concrètement les exigences du règlement européen sur l'IA pourrait voir sa responsabilité contractuelle engagée pour délivrance non conforme du logiciel¹⁰⁹⁰. De façon indirecte, les éléments techniques de preuve du fonctionnement du système d'IA à haut risque contribuent donc à réduire les hypothèses de mise en œuvre de la responsabilité contractuelle du fournisseur.

B- La contractualisation des preuves techniques de conception

La référence par défaut aux normes techniques. Bien que le terme soit identique, l'obligation de délivrance conforme ne renvoie pas à la conformité légale de la chose mais à sa conformité au regard des spécifications contractuelles convenues entre les parties¹⁰⁹¹. Dans un premier temps, il est possible que les co-contractants n'aient pas inclus dans leur contrat les éléments constitutifs de la conformité de la chose. Dans cette situation, le code civil dispose que :

*« lorsque la qualité de la prestation n'est pas déterminée ou déterminable en vertu du contrat, le débiteur doit offrir une prestation de qualité conforme aux attentes légitimes des parties en considération de sa nature, des usages et du montant de la contrepartie. »*¹⁰⁹²

Par défaut, le juge évaluera la conformité du système d'IA délivré au regard notamment de l'état de technique au moment de la délivrance du logiciel. L'état de la technique est déterminé selon un ensemble de pratiques professionnelles à respecter, dont les normes techniques élaborées par les organismes de normalisation, qu'ils soient nationaux, européens ou

¹⁰⁸⁷ Voir Section II : Une conformité juridique ex ante pour les systèmes d'IA à haut risque.

¹⁰⁸⁸ « Règlement sur l'intelligence artificielle », 2024. Article 15.1.

¹⁰⁸⁹ *Ibid.* Article 9.5.

¹⁰⁹⁰ Cour de cassation, 10 octobre 2018, n° 15-26.093.

¹⁰⁹¹ LE TOURNEAU Philippe, *Contrats du numérique ; informatiques et électroniques*, Dalloz, 2022.

¹⁰⁹² Article 1166 du code civil.

internationaux. Le manquement à ces obligations implicites peut alors engager la responsabilité contractuelle du fournisseur¹⁰⁹³. Dans le cas où les normes techniques contiendraient des recommandations relatives à des procédures de tests à réaliser et/ou des résultats à obtenir, ces derniers contribueraient à constituer la notion de conformité au sens de l'obligation de délivrance conforme. Ainsi, les indicateurs et seuils probabilistes, contenus le cas échéant dans les normes techniques servant de référence, jouent aussi un rôle dans l'évaluation de la conformité des systèmes d'IA autres qu'à haut risque. En l'absence de spécifications contractuelles, le système d'IA à faible ou moyen risque doit lui aussi être délivré conformément à l'état de la technique. En outre, l'obligation de la délivrance conforme peut s'étendre au-delà des exigences prévues par le règlement européen sur l'IA, comme la durabilité environnementale par exemple¹⁰⁹⁴. En d'autres termes, le recours à la notion d'« état de la technique » pour évaluer la conformité du système d'IA délivré fait intervenir les normes techniques dans des situations plus nombreuses et diverses que celles prévues par le seul règlement européen sur l'IA.

L'utilisation des preuves techniques de fonctionnement comme spécifications contractuelles. Dépendamment des rapports de pouvoirs entre les co-contractants, le client a toujours intérêt à préciser ses besoins devant être satisfaits par le système d'IA objet du contrat, généralement au sein d'un cahier des charges. De plus, il est préférable que ces spécifications soient, d'une part, formalisées de façon à être aisément vérifiables au moment de la délivrance et, d'autre part, associées à des procédures de validations précises de ces spécifications¹⁰⁹⁵. Les preuves techniques de conception, telles que définies dans notre étude, correspondent exactement à ces deux aspects des spécifications. Par conséquent, il est très probable que certains déployeurs, lorsque leur pouvoir de négociation le permet, inscrivent dans le contrat de fourniture du système d'IA à haut risque des preuves techniques au sein des spécifications contractuelles. Dans un premier temps, le déployeur peut exiger que le niveau de qualité du système d'IA soit supérieur au niveau, approprié ou acceptable selon les exigences, imposé par le règlement européen sur l'IA. Par exemple, imaginons que les normes européennes harmonisées, produites par le CEN et le CENELEC, indiquent qu'un niveau d'exactitude

¹⁰⁹³ MAZEAU Laurène, « L'influence des sciences sur le droit de la responsabilité civile professionnelle », in *Qu'en est-il du droit de la recherche ?*, Presses de l'Université Toulouse 1 Capitole, 2018. P. 29.

¹⁰⁹⁴ « Règlement sur l'intelligence artificielle », 2024. Article 95. 2.b.

¹⁰⁹⁵ LE TOURNEAU Philippe, « Contrats informatiques », 2023.

approprié au sens du règlement sur l'IA est au minimum une exactitude moyenne de 70%¹⁰⁹⁶. Le déployeur peut alors parfaitement exiger du fournisseur que le système d'IA à haut risque qu'il achète doit atteindre un niveau d'exactitude de 80%, mais également que le taux de faux négatif du logiciel ne doit pas être inférieur à 98%. Dans une autre situation, un client pourrait aussi inscrire que le système d'IA présente un fonctionnement équitable, qu'il définirait comme une parité de résultats favorables entre les hommes et les femmes à 5% près. Le cas échéant, il définit donc l'indicateur relatif à l'exigence de biais, à savoir la parité statistique, et le seuil maximal d'acceptabilité, ici 5%.

Dans ces deux exemples, les spécifications contractuelles peuvent être plus précises, complémentaires ou alternatives avec les preuves techniques contenues, le cas échéant, dans les normes européennes harmonisées. Quel que soit le contenu des normes techniques ou leur niveau de précision, les parties à un contrat portant sur un système d'IA utiliseront, comme pour beaucoup de logiciel, des métriques et seuils statistiques pour définir le niveau de conformité attendue du produit. Deux conclusions peuvent alors être tirées de la contractualisation des preuves techniques de conception. Premièrement, dans le cas d'un logiciel (à l'inverse du progiciel), le contrat sera un lieu privilégié pour le choix des preuves techniques pertinentes aux yeux des parties, en particulier pour le déployeur. Deuxièmement, les éléments techniques de preuve du fonctionnement du système d'IA sont des éléments centraux dans l'évaluation du respect de l'obligation de délivrance conforme, et donc dans la responsabilité contractuelle du fournisseur du logiciel.

§2. La présence de vice cachés

Annonce de plan. La performance¹⁰⁹⁷ d'un système d'IA, qu'il soit à haut risque ou non, est un élément non négligeable de la vente (ou location) d'un logiciel. Pour autant, la possibilité d'un recours à la garantie des vices cachés par le client, en raison de la qualité de fonctionnement du logiciel, à l'encontre du fournisseur du système d'IA n'apparaît pas

¹⁰⁹⁶ Pour la clarté de l'exemple, le contexte d'utilisation du logiciel ne sera pas pris en compte. De la même manière, la spécification, inventée pour les besoins de l'illustration, contenue dans la norme technique ne précise pas de secteur d'utilisation particulier.

¹⁰⁹⁷ « la capacité d'un système d'IA à remplir sa destination » - « Règlement sur l'intelligence artificielle », 2024. Article 3.18).

évidente. Pour cela, l'article 1641 du code civil énonce que les potentiels vices doivent être cachés du client (A) mais aussi qu'ils rendent le système d'IA impropre à l'usage¹⁰⁹⁸ (B).

A- Un vice caché

Le vice relatif à la qualité de fonctionnement. Le caractère caché est un critère de qualification du vice exprimé de façon extrêmement claire par le code civil français : « *Le vendeur n'est pas tenu des vices apparents et dont l'acheteur a pu se convaincre lui-même.* »¹⁰⁹⁹. En ce qui concerne notre étude, le potentiel vice porterait sur la qualité de fonctionnement du logiciel. Or, l'une des problématiques centrales mise en avant par les textes éthiques et le droit positif est l'opacité de leur fonctionnement¹¹⁰⁰. Au premier abord, si un vice relatif à la qualité de fonctionnement d'un système d'IA venait à exister, il est très probable que celui-ci ne soit pas apparent pour le client, notamment si celui-ci n'est pas un professionnel du secteur de l'IA¹¹⁰¹. Cependant, les nouvelles exigences de transparence contenues dans le règlement européen sur l'IA pourraient remettre en question le caractère caché d'un vice relatif à la qualité de fonctionnement d'un système d'IA.

La communication au client des informations relatives à la qualité de fonctionnement du système. Le règlement européen sur l'IA impose plusieurs mécanismes favorisant la transparence des systèmes d'IA, notamment en ce qui concerne leur fonctionnement. Par exemple, l'élaboration d'une documentation technique¹¹⁰² ainsi que la mise en place de journaux informatiques¹¹⁰³ permettent, chacune à leur manière, de rendre apparent le fonctionnement du système d'IA à haut risque. En ce qui concerne l'information communiqué au client, ici le déployeur, le fournisseur doit lui accompagner son système d'IA d'une notice d'utilisation, contenant notamment :

¹⁰⁹⁸ « *Le vendeur est tenu de la garantie à raison des défauts cachés de la chose vendue qui la rendent impropre à l'usage auquel on la destine, ou qui diminuent tellement cet usage que l'acheteur ne l'aurait pas acquise, ou n'en aurait donné qu'un moindre prix, s'il les avait connus.* » - Article 1641 du code civil.

¹⁰⁹⁹ Article 1641 du code civil.

¹¹⁰⁰ BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016. Pour plus de détails sur la notion d'opacité, voir A— Le principe de transparence.

¹¹⁰¹ La qualité du client, à savoir qu'il soit un professionnel du même secteur ou non, a des conséquences sur la qualification du caractère caché du vice. En effet, le professionnel présente les compétences pour réaliser un examen visant à déceler certains vices que ne possède pas les autres clients : Cour de Cassation, Chambre commerciale, 4 mars 2003, n° 00-18.668.

¹¹⁰² « Règlement sur l'intelligence artificielle », 2024. Articles 11 et 18.

¹¹⁰³ *Ibid.* Articles 12 et 19.

« les caractéristiques, les capacités et les limites de performance du système d'IA à haut risque, notamment :

[...]

ii) le niveau d'exactitude, y compris les indicateurs utilisés, de robustesse et de cybersécurité visé à l'article 15 qui a servi de référence pour les tests et la validation du système d'IA à haut risque et qui peut être attendu ;

iii) toutes circonstances connues ou prévisibles liées à l'utilisation du système d'IA à haut risque conformément à sa destination ou dans des conditions de mauvaise utilisation raisonnablement prévisible, susceptibles d'entraîner des risques pour la santé et la sécurité ou pour les droits fondamentaux ; [...]

v) le cas échéant, sa performance en ce qui concerne des personnes ou groupes de personnes spécifiques à l'égard desquels le système est destiné à être utilisé ; »¹¹⁰⁴

Au travers de la notice d'utilisation, le déployeur est donc informé à la fois des conditions d'utilisation dans lesquelles le système d'IA à haut risque peut présenter des risques mais également de la qualité de fonctionnement du logiciel, à savoir les procédures de tests utilisées ainsi que les résultats de tests. Ainsi, si le système d'IA en question présente un niveau d'exactitude particulièrement faible pour un profil d'individus que le déployeur rencontre fréquemment, ce dernier en prend connaissance au plus tard lors de l'accès à la notice d'utilisation. Par conséquent, la notice d'utilisation accompagnant un système d'IA à haut risque rend immédiatement apparent tout défaut relatif à la qualité de fonctionnement du logiciel. À *contrario*, la présence d'informations erronées, ou insuffisamment claires pour rendre compte fidèlement des conséquences¹¹⁰⁵, dans la notice d'utilisation permet de considérer le défaut comme étant caché.

L'antériorité du vice de fonctionnement du système d'IA. La présence d'un vice de fonctionnement dans le système d'IA doit nécessairement être antérieure au transfert des risques entre les parties au contrat¹¹⁰⁶. Jugée de façon souveraine par les juges du fond¹¹⁰⁷, l'antériorité du vice doit être prouvée par le client¹¹⁰⁸, en l'espèce le déployeur du logiciel. Dans le contexte

¹¹⁰⁴ *Ibid.* Article 13.3.b)

¹¹⁰⁵ Pour être considéré comme apparent, l'acheteur doit être informé des vices « *dans leurs causes et leurs amplitudes* » (Cour de Cassation, Chambre civile 3, 14 juin 1989, n° 87-19.312.) et de leur « *étendue et la gravité* » (Cour de Cassation, Chambre civile 1, 7 juin 1995, n° 93-13.060.).

¹¹⁰⁶ Cour de cassation, Chambre civile 1, 1 mars 2017, n° 15-29.413 Cour de cassation, Chambre commerciale, 20 avril 2017, n° 15-21.877.

¹¹⁰⁷ Cour de Cassation, Chambre civile 1, 15 juillet 1999, n° 97-16.783.

¹¹⁰⁸ Cour de Cassation, Chambre commerciale, 18 janvier 1984, n° 82-14.977.

d'un contrat portant sur un système d'IA, cette condition d'antériorité se révèle particulièrement délicate au regard de la capacité d'évolution des technologies d'IA, en particulier les techniques d'apprentissage automatique. De manière générale, le fonctionnement d'un logiciel d'apprentissage automatique est « gelé » à la fin de la phase d'entraînement, et donc avant la mise sur le marché du système d'IA. Dans cette situation, la découverte d'un défaut dans le fonctionnement du logiciel est un indice significatif de l'antériorité du vice, à condition qu'aucune autre cause ne puisse expliquer le dommage causé¹¹⁰⁹. Toutefois, le système d'IA peut être mis sur le marché européen avec la possibilité d'évoluer, que ce soit par la possibilité pour le client de réentraîner le modèle, de paramétrer le système ou encore de le mettre à jour. Dans ce cas, l'antériorité d'un vice, telle que l'augmentation considérable du taux de biais discriminatoires dans le système d'IA à haut risque par rapport à la notice d'utilisation, se révélerait beaucoup plus difficile à prouver.

B- Un vice rendant le système d'IA impropre à l'usage

Les caractéristiques du vice. La notion juridique de vice désigne un défaut affectant une chose¹¹¹⁰. Ainsi le vice doit être « *nécessairement inhérent à la chose elle-même* », ici le système d'IA à haut risque. Cependant, tous les défauts ne constituent pas un vice au sens de la garantie légale des vices cachés. Le défaut doit être suffisamment important pour rendre la chose « *impropre à l'usage à laquelle on la destine, ou qui diminuent tellement cet usage que l'acheteur ne l'aurait pas acquise, ou n'en aurait donné qu'un moindre prix* »¹¹¹¹. Contrairement au défaut de conformité, la qualification du vice prend pour référence la finalité de la chose et non les spécifications contractuelles. Le vice peut porter aussi bien sur les caractéristiques économiques de la chose que sur ses qualités techniques¹¹¹². Par exemple, la présence d'un virus dans un logiciel ou son fonctionnement défectueux peuvent constituer un vice. Ainsi, une qualité de fonctionnement significativement moins haute que celle annoncée dans la notice d'utilisation communiquée au déployeur est possiblement de nature à constituer un vice.

L'importance du vice de fonctionnement pour un système d'IA. Lors de la conception de son système d'IA à haut risque, le fournisseur constitue, par le biais de tests, des preuves

¹¹⁰⁹ Cour de Cassation, Chambre civile 1, 5 novembre 1996, n° 93-21.762.

¹¹¹⁰ CORNU Gérard, *Vocabulaire juridique*, 2022.

¹¹¹¹ Article 1641 du code civil.

¹¹¹² LE TOURNEAU Philippe, *Contrats du numérique ; informatiques et électroniques*, Dalloz, 2022.

techniques relatives à la qualité de fonctionnement de son logiciel afin d'attester de sa conformité au règlement européen sur l'IA. Ces preuves doivent notamment être communiquées au déployeur, dans notre situation le client, au travers de la notice d'utilisation. Si les informations relatives à la qualité de fonctionnement du système d'IA à haut risque se révèlent erronées, dans quelle mesure cela peut-il constituer un vice ? Par exemple, le seuil statistique d'exactitude du système sur une catégorie d'individus est en réalité 25% plus bas que celui indiqué dans la notice d'utilisation¹¹¹³. Un second exemple pourrait être une erreur sur l'indicateur utilisé pour réaliser les tests de biais sur la performance du logiciel (une confusion entre l'égalité des chances et l'asymétrie des chances par exemple¹¹¹⁴). En d'autres termes, les preuves techniques relatives à la qualité de fonctionnement du système d'IA sont-elles des éléments pouvant faire l'objet d'un vice ?

Dans un premier temps, il faut considérer la situation où la performance moindre du système d'IA à haut risque a des conséquences sur sa certification au sens du règlement européen sur l'IA. C'est notamment le cas lorsque le système d'IA à haut risque doit faire l'objet d'un rappel, à savoir « *toute mesure visant à assurer le retour au fournisseur d'un système d'IA mis à la disposition des déployeurs ou à le mettre hors service ou à désactiver son utilisation* »¹¹¹⁵. À la suite d'un contrôle *ex post* réalisé par une autorité de surveillance du marché, celle-ci peut décider que le marquage CE du système a été apposé à tort sur le système d'IA car la performance réelle du logiciel n'est pas satisfaisante au regard des exigences prévues par le règlement européen sur l'IA¹¹¹⁶. Pour un déployeur souhaitant utiliser un système d'IA pour une finalité considérée à haut risque, la perte de certification du logiciel, du fait de sa performance moindre, le rend impropre à sa destination. Par conséquent, un défaut relatif à la performance du système d'IA tel qu'il entraîne un rappel ou un retrait du logiciel est suffisamment important pour être qualifié de vice.

Dans un second temps, il convient de s'interroger sur le cas où la performance moindre d'un système d'IA à haut risque, en comparaison des informations contenues dans la notice d'utilisation, n'a pas de conséquences sur la certification européenne du logiciel. En d'autres

¹¹¹³ Voir *supra*.

¹¹¹⁴ Pour rappel, la métrique de l'égalité des chances compare la proportion d'erreur du système d'IA entre deux catégories de population (hommes et femmes par exemple). La métrique de l'asymétrie des chances, quant à elle, représente la proportion d'erreurs dommageables et d'erreurs favorables entre les deux catégories de population. Malgré leur similarité, ces deux métriques ne formalisent pas la même conception des biais discriminatoires. Pour plus de détails, voir §1. La détermination de l'objet de la preuve : l'indicateur.

¹¹¹⁵ « Règlement sur l'intelligence artificielle », 2024. Article 3.16)

¹¹¹⁶ *Ibid.* Article 83.2.

termes, la seule découverte d'une qualité de fonctionnement moindre est-elle suffisante pour caractériser un vice ? Tout d'abord, si le vice peut porter sur les qualités techniques d'une chose, il présente toujours un caractère qualitatif et non quantitatif¹¹¹⁷. Or, les preuves techniques utilisées pour évaluer la qualité de fonctionnement du système d'IA sont des outils quantitatifs. La seule performance faible du logiciel n'est donc pas suffisante pour caractériser un vice. Ce dernier doit être suffisamment important pour, au minimum, diminuer grandement l'usage du système d'IA. Pour les systèmes d'IA à haut risque, il est difficile d'imaginer un tel scénario car la certification est également dépendante de la destination du système d'IA. Par conséquent, il semble peu probable qu'un défaut, lié à la qualité de fonctionnement, qui engendrerait de telles conséquences sur l'usage du logiciel sans pour autant remettre en question sa conformité puisse être considéré comme un vice¹¹¹⁸.

Conclusion de la sous-section. Au-delà du règlement européen sur l'IA, les éléments techniques de preuve relatifs à la qualité de fonctionnement des systèmes d'IA à haut risque ont une place significative dans la responsabilité contractuelle du fournisseur. D'une part, en tant qu'éléments attestant de la certification légitime indispensable du système d'IA à haut risque, ces preuves contribuent à garantir que le logiciel sera capable d'assurer sa fonction en toute légalité. D'autre part, il apparaît que les spécifications contractuelles sont un espace de discussion et de détermination des éléments techniques de preuve considérés comme pertinents. Selon le pouvoir de négociation de chacun, le fournisseur et son client peuvent choisir les preuves techniques qu'ils considèrent comme pertinentes au regard de leurs objectifs. Si les preuves techniques sont déterminées en amont par les acteurs énoncés dans le règlement sur l'IA (organismes de normalisation ou autorités de contrôle), alors elles contraindront, au moins partiellement, la liberté contractuelle sur le sujet. Dans le cas inverse, les preuves techniques de fonctionnement peuvent être déterminées par les parties au regard de leurs intérêts, notamment économiques.

Sous-section 2 : Le fait délictuel du fournisseur du système d'IA

Annonce de plan. La survenance d'un préjudice du fait d'un système d'IA peut avoir de nombreuses causes, dont son utilisation par le déployeur mais également le fonctionnement

¹¹¹⁷ LE TOURNEAU Philippe, « Contrats informatiques », 2023.

¹¹¹⁸ À ce jour, il n'est cependant pas possible d'exclure complètement cette possibilité.

même du système d'IA. Dans ce dernier cas, la responsabilité délictuelle du fournisseur peut être interrogée, en particulier au regard de sa responsabilité *ex ante*. L'analyse de la qualité de fonctionnement du système d'IA à haut risque, et donc des preuves en attestant, peut alors permettre de révéler la présence (ou l'absence) d'une faute commise par le fournisseur (§1) ou d'un défaut dans son produit (§2).

§1. La faute du fournisseur

La définition de la faute par la proposition de directive européenne sur la responsabilité en matière d'IA. « *Tout fait quelconque de l'homme, qui cause à autrui un dommage, oblige celui par la faute duquel il est arrivé à le réparer.* ». Par cette disposition, l'article 1240 du code civil fonde le régime de responsabilité extracontractuelle du fait personnel. Pour reprendre la définition de Marcel Planiol, la notion de faute désigne le « *manquement à une obligation existante* »¹¹¹⁹. Le demandeur lésé qui souhaite agir sur le fondement de la responsabilité civile pour faute doit identifier, tout d'abord, les opérations que le défendeur aurait dû réaliser. La faute peut résulter aussi bien d'un acte positif que d'une omission¹¹²⁰. Dans le cadre d'une action en réparation contre un fournisseur de système d'IA à haut risque, il est évident que les obligations imposées par le règlement européen sur l'IA constituent une référence essentielle pour caractériser la faute. Ce lien entre les règles de responsabilité *ex ante* des fournisseurs de système d'IA à haut risque et leur responsabilité extracontractuelle pour faute est explicité au sein de la proposition de directive européenne sur la responsabilité en matière d'IA :

*« la condition énoncée au paragraphe 1, point a) [la faute du fournisseur d'un système d'IA à haut risque], n'est remplie que si le plaignant a démontré que le fournisseur ou, le cas échéant, la personne soumise aux obligations qui incombent au fournisseur n'a pas respecté l'une des exigences suivantes, énoncées dans lesdits chapitres, compte tenu des mesures prises dans le cadre du système de gestion des risques et des résultats de ce système conformément à [l'article 9 et à l'article 16, point a), de la législation sur l'IA]: »*¹¹²¹

¹¹¹⁹ PLANIOL Marcel, *Traité élémentaire de droit civil*, 1905. Dans BRUN Philippe, *Responsabilité civile extracontractuelle*, LEXISNEXIS, 2023. P. 201.

¹¹²⁰ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Considérant 22.

¹¹²¹ *Ibid.* Article 4.2.

En d'autres termes, le comportement du fournisseur d'un système d'IA à haut risque constitue une faute uniquement lorsqu'il vient à manquer aux exigences contenues dans le règlement européen sur l'IA. Plus précisément, la directive européenne liste un total de cinq obligations du fournisseur de système d'IA à haut risque :

« (a) le système d'IA est un système qui fait appel à des techniques qui impliquent l'entraînement de modèles au moyen de données et qui n'a pas été développé sur la base de jeux de données d'entraînement, de validation et de test qui satisfont aux critères de qualité visés à [l'article 10, paragraphes 2 à 4, de la législation sur l'IA];

(b) le système d'IA n'a pas été conçu et développé d'une manière qui réponde aux exigences de transparence énoncées à [l'article 13 de la législation sur l'IA];

(c) le système d'IA n'a pas été conçu et développé de manière à permettre un contrôle effectif par des personnes physiques pendant la période d'utilisation du système d'IA conformément à [l'article 14 de la législation sur l'IA];

(d) le système d'IA n'a pas été conçu et développé de manière à atteindre, compte tenu de sa destination, un niveau approprié d'exactitude, de robustesse et de cybersécurité conformément à [l'article 15 et à l'article 16, point a), de la législation sur l'IA]; ou

(e) les mesures correctives nécessaires n'ont pas été prises immédiatement pour mettre le système d'IA en conformité avec les obligations énoncées au [titre III, chapitre 2, de la législation sur l'IA] ou pour le retirer ou le rappeler, selon le cas, conformément à [l'article 16, point g), et à l'article 21 de la législation sur l'IA]. »¹¹²²

Les obligations mentionnées ci-dessus ne correspondent pas à l'ensemble des exigences applicables aux systèmes d'IA à haut risque¹¹²³. De façon logique, seules les obligations destinées à prévenir un dommage sont concernées¹¹²⁴. Par exemple, le manquement par le fournisseur à son obligation de mettre en place des « journaux »¹¹²⁵ tout au long du cycle de vie du système d'IA à haut risque¹¹²⁶ ne constitue pas une faute pouvant engager sa responsabilité extracontractuelle. Ainsi, ce sont toutes les obligations relatives au fonctionnement du système

¹¹²² Ibid.

¹¹²³ « Règlement sur l'intelligence artificielle », 2024. Intitulé de la Section 2 du chapitre III.

¹¹²⁴ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Considérant 22.

¹¹²⁵ Les journaux sont un mécanisme d'enregistrement automatique des événements au cours de l'utilisation du système d'IA.

¹¹²⁶ « Règlement sur l'intelligence artificielle », 2024. Article 12.

d'IA à haut risque qui sont visées par la directive. Par conséquent, il existe un lien direct mais partiel entre la conformité du système d'IA à haut risque mis sur le marché et la responsabilité pour faute de son fournisseur.

Un comportement fautif dépassant les seules obligations spécifiques du règlement européen sur l'IA. Toutefois, la proposition de directive européenne sur la responsabilité en matière d'IA ne limite pas la faute du fournisseur d'un système à haut risque au seul manquement au règlement sur l'IA. En premier lieu, le texte énonce que :

*« la faute du défendeur ou d'une personne dont le comportement relève de sa responsabilité, consistant en un manquement à un devoir de vigilance prévu par le droit de l'Union ou le droit national visant directement à protéger contre le dommage survenu ; »*¹¹²⁷

La notion de devoir de vigilance est définie par la proposition de directive comme :

*« la norme de conduite requise, fixée par le droit national ou le droit de l'Union, pour éviter de porter atteinte aux intérêts juridiques reconnus au niveau du droit national ou du droit de l'Union, notamment la vie, l'intégrité physique, les biens et la protection des droits fondamentaux. »*¹¹²⁸

Par conséquent, le fournisseur d'un système d'IA à haut risque réalise également une faute s'il vient à manquer à des obligations énoncées par des réglementations sectorielles, telle que le règlement (CE) n° 300/2008 relatif à l'instauration de règles communes dans le domaine de la sûreté de l'aviation civile¹¹²⁹. De la même manière, les devoirs de vigilance qui seraient énoncés par le droit national sont à prendre en compte dans la caractérisation de la faute du fournisseur du système d'IA à haut risque. Par ailleurs, la Cour de cassation semble considérer depuis peu¹¹³⁰ que le défaut de sécurité constitue également une faute permettant d'engager la responsabilité délictuelle du producteur, sans que cette faute ne doive être distincte du défaut de sécurité¹¹³¹. Si cette récente jurisprudence doit encore être précisée, elle renforce, d'une part,

¹¹²⁷ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 4.1.a)

¹¹²⁸ Ibid. Article 2(9).

¹¹²⁹ « Règlement (CE) n° 300/2008 du Parlement européen et du Conseil du 11 mars 2008 relatif à l'instauration de règles communes dans le domaine de la sûreté de l'aviation civile et abrogeant le règlement (CE) n° 2320/2002 (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008.

¹¹³⁰ Cour de cassation, 15 novembre 2023, n° 22-21.174 ; Cour de cassation, 15 novembre 2023, n° 22-21.178 ; Cour de cassation, 15 novembre 2023, n° 22-21.179 ; Cour de cassation, 15 novembre 2023, n° 22-21.180.

¹¹³¹ JOURDAIN Patrice, « La responsabilité du producteur peut être engagée en cas de faute », *RTD Civ.*, 2024. Une interprétation plus nuancée peut être trouvée dans : RADÉ Christophe, « La responsabilité du producteur résultant d'un défaut de suivi de ses produits », *Responsabilité civile et assurances*, 2024.

le constat que la faute du fournisseur d'un système d'IA à haut risque ne sera pas limitée à certains manquements au règlement sur l'IA et, d'autre part, les liens entre la notion de faute et la conception (défectueuse) du système d'IA¹¹³².

Contrairement à la résolution du parlement européen de 2020¹¹³³, la proposition de directive européenne sur la responsabilité en matière d'IA ne vise pas à créer un nouveau régime de responsabilité spécifique aux dommages causés par les systèmes d'IA¹¹³⁴. Son principal apport est la facilitation de la preuve pour le demandeur qui souhaiterait prouver la faute extracontractuelle du fournisseur du système d'IA¹¹³⁵. En ce qui nous concerne, elle permet d'affirmer que la conformité du système d'IA à haut risque limite, au moins partiellement, la responsabilité extracontractuelle de son fournisseur. Par conséquent, les preuves techniques de conception servant à attester cette conformité joueront un rôle déterminant dans la caractérisation de la faute extracontractuelle du fournisseur.

§2. Le défaut du système d'IA

Annonce de plan. Issue de la directive européenne de 1985¹¹³⁶, la responsabilité du fait des produits défectueux est un régime de responsabilité privilégié pour la réparation d'un dommage causé du fait d'un système d'IA¹¹³⁷. Dans le contexte de l'IA, l'attention portée à ce régime de responsabilité *ex post* s'explique par son caractère favorable en tant que régime de responsabilité objective mais surtout par sa primauté sur d'autres régimes de responsabilité

¹¹³² Pour les liens entre le défaut de sécurité, la conception d'un système d'IA et les preuves technique de conception, voir B- L'appréciation du défaut.

¹¹³³ PARLEMENT EUROPÉEN, « Un régime de responsabilité civile pour l'intelligence artificielle », 2020.

¹¹³⁴ PLEDEL Chloé, GALBOIS-LEHALLE Diane et CASSAR Bertrand, « L'articulation du projet de règlement sur l'intelligence artificielle avec le droit du numérique européen », 2023.

¹¹³⁵ GRYNBAUM Luc, LE GOFFIC Caroline et PAILLER Ludovic, *Droit des activités numériques*, Dalloz, 2023. Chapitre 3 : Responsabilité du fait d'une intelligence artificielle.

¹¹³⁶ « Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux », 210, 1985.

¹¹³⁷ BACACHE-GIBEILI Mireille, *Droit civil - Les obligations, la responsabilité civile extracontractuelle*, Economica, 2021. CAILLÉ Catherine, « Responsabilité du fait des produits défectueux », *Répertoire de droit civil*, 2018. CHABAS François, MAZEAUD Henri, MAZEAUD Jean et al., *Traité théorique et pratique de la responsabilité civile délictuelle et contractuelle*, LGDJ, 2013.

existants en droit positif¹¹³⁸. Par conséquent, il est indispensable d'étudier la manière avec laquelle les preuves techniques utilisées pour attester de la conformité d'un système d'IA à haut risque peuvent s'articuler, aujourd'hui, avec l'identification d'un défaut de sécurité **(A)** et dans quelle mesure cette interaction peut évoluer au regard de la proposition de directive européenne sur la responsabilité du fait des produits défectueux **(B)**.

A- La notion de défaut

La définition de la notion de défaut. Le terme de « défaut » se révèle particulièrement ambigüe et peut entraîner une confusion avec des notions voisines telles que le vice ou le défaut de conformité. Selon l'article 1245-3 du code civil, est défectueux le produit qui « *n'offre pas la sécurité à laquelle on peut légitimement s'attendre* ». Ainsi, l'identification d'un défaut nécessite la présence de deux critères : l'existence d'un défaut de sécurité, d'une part, et la sécurité à laquelle le produit manque doit être celle qui peut être légitimement attendue, d'autre part.

Au sens de la responsabilité du fait des produits défectueux, le défaut est donc entendu comme celui qui compromet la sécurité de l'utilisateur du produit¹¹³⁹. Ainsi, le défaut de sécurité se différencie de la non-conformité du produit en ce qu'il ne se réfère pas aux spécifications contractuelles convenues entre les parties. De la même manière, le défaut de sécurité s'écarte du vice en ce qu'il ne nécessite pas de rendre le produit impropre à l'usage. Dans le cadre des systèmes d'IA, l'un des deux objectifs poursuivis par le règlement européen sur l'IA est la garantie d'un « *niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux [...]* »¹¹⁴⁰. Il apparaît donc un lien clair entre la responsabilité *ex ante* du fournisseur d'un

¹¹³⁸ Le régime de responsabilité du fait des choses, inscrit à l'article 1242 du code civil, se révèle être un régime de responsabilité objective bien plus intéressant et adapté pour le demandeur lésé. En effet, ce régime n'est pas limité aux seuls préjudices matériels et ne nécessite aucun fait générateur pour engager la responsabilité du gardien de la chose. Cependant, les jurisprudences nationale et européenne s'accordent sur le fait que le régime de responsabilité du fait des produits défectueux est impératif et prime sur la responsabilité du fait des choses : Cour de cassation, Chambre civile 1, 11 juillet 2018, n° 17-20.154. Cour de justice des Communautés européennes, 25 avril 2002, n° C-52/00 Cour de justice des Communautés européennes, 25 avril 2002, n° C-154/00.

¹¹³⁹ LE TOURNEAU Philippe, *Droit de la responsabilité et des contrats - Régimes d'indemnisation*, Dalloz, .

¹¹⁴⁰ « L'objectif du présent règlement est d'améliorer le fonctionnement du marché intérieur et de promouvoir l'adoption de l'intelligence artificielle (IA) axée sur le facteur humain et digne de confiance, tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux consacrés dans la charte des droits fondamentaux, notamment la démocratie, l'État de droit et la protection de l'environnement, contre les effets néfastes des systèmes d'intelligence artificielle (ci-après dénommés "systèmes d'IA") dans l'Union, ainsi que de soutenir l'innovation. » - « Règlement sur l'intelligence artificielle », 2024. Article 1.1.

système d'IA, en l'espèce à haut risque, et la caractérisation d'un défaut de sécurité contenu dans le logiciel.

Les catégories de défauts. Au fil des décisions judiciaires sur la notion de défaut, la doctrine a mis en avant trois catégories de défauts : les défauts d'information, les défauts de fabrication et les défauts de conception¹¹⁴¹. Premièrement, le défaut d'information renvoie à l'importance de l'information communiquée à l'utilisateur dans l'appréciation de la « sécurité légitimement attendue ». En d'autres termes, un système d'IA à haut risque qui ne contiendrait aucune anomalie matérielle peut toutefois présenter un défaut dès lors que l'information fournie par le fournisseur-producteur n'est pas suffisante¹¹⁴². L'absence d'anomalie matérielle dans le produit conduit à écarter cette catégorie de défaut de notre analyse car elle ne présente aucun lien avec la qualité de fonctionnement du système d'IA.

Deuxièmement, le défaut de fabrication désigne la situation où un seul élément d'un lot est différent du reste des produits, cette différence entraînant alors un défaut de sécurité. Troisièmement et dernièrement, le défaut de conception est le résultat d'un processus de conception défectueux qui n'a pas permis d'assurer la qualité du produit final. Dans cette situation, tous les exemplaires du produit du fabricant présentent le défaut de sécurité. Le défaut de fabrication et le défaut de conception, aussi appelés « défauts intrinsèques »¹¹⁴³, présentent tous les deux des liens avec la conformité d'un système d'IA¹¹⁴⁴ et seront donc au centre de l'analyse.

B- L'appréciation du défaut

Les critères législatifs du défaut. Tous les défauts de sécurité d'un système d'IA ne permettront pas d'engager la responsabilité extracontractuelle du fournisseur-producteur d'un système d'IA à haut risque. Le défaut identifié doit encore être suffisamment important pour diminuer la sécurité de l'utilisateur du produit en dessous de ses attentes légitimes. La notion

¹¹⁴¹ BERTOLINI Andrea, « Artificial Intelligence and Civil Liability », Parlement Européen, 2020. P. 52.

¹¹⁴² Voir notamment : Cour de cassation, Chambre civile 1, 4 février 2015, n° 13-19.781. Cour de cassation, Chambre civile 1, 27 novembre 2019, n° 18-16.537.

¹¹⁴³ CAILLÉ Catherine, « Responsabilité du fait des produits défectueux », *Répertoire de droit civil*, 2018.

¹¹⁴⁴ BOINE Claire, « La responsabilité civile liée à l'intelligence artificielle », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général. Perspectives comparées*, Bruylant, 2023.

de « sécurité à laquelle on peut légitimement s'attendre » est interprétée notamment au travers des critères complémentaires énoncés à l'article 1245-3 alinéa 2 du code civil :

« Dans l'appréciation de la sécurité à laquelle on peut légitimement s'attendre, il doit être tenu compte de toutes les circonstances et notamment de la présentation du produit, de l'usage qui peut en être raisonnablement attendu et du moment de sa mise en circulation. »¹¹⁴⁵

Cette liste, non-exhaustive, est la transposition exacte de la directive européenne relative à la responsabilité du fait des produits défectueux datant de 1985¹¹⁴⁶. Une proposition de directive européenne relative à la responsabilité du fait des produits défectueux, destinée à actualiser celle de 1985, a été proposée par la Commission européenne en 2022¹¹⁴⁷. Si elle ne change pas substantiellement la définition du défaut, cette proposition énonce de nouveaux critères d'appréciation de la défectuosité, dont certains ont été pensés explicitement pour les systèmes d'IA¹¹⁴⁸ :

« Un produit est considéré comme défectueux lorsqu'il n'offre pas la sécurité à laquelle le grand public peut légitimement s'attendre compte tenu de toutes les circonstances, y compris :

[...]

c) l'effet sur le produit de toute capacité à poursuivre son apprentissage après le déploiement ;

d) l'effet sur le produit d'autres produits dont on peut raisonnablement s'attendre à ce qu'ils soient utilisés en même temps que le produit ;

[...]

f) les exigences en matière de sécurité des produits, y compris les exigences de cybersécurité pertinentes pour la sécurité ;

¹¹⁴⁵ « Ordonnance n° 2016-131 du 10 février 2016 portant réforme du droit des contrats, du régime général et de la preuve des obligations », 2016. Article 2.

¹¹⁴⁶ « Un produit est défectueux lorsqu'il n'offre pas la sécurité à laquelle on peut légitimement s'attendre compte tenu de toutes les circonstances, et notamment : a) de la présentation du produit ; b) de l'usage du produit qui peut être raisonnablement attendu ; c) du moment de la mise en circulation du produit. » - « Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux », 210, 1985. Article 6.

¹¹⁴⁷ Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 28 septembre 2022.

¹¹⁴⁸ MANGEMATIN Céline, « Produits défectueux - Responsabilité du fait des produits défectueux et intelligence artificielle : une proposition presque parfaite », *Responsabilité civile et assurances*, 2023. P. 10.

g) toute intervention d'une autorité de réglementation ou d'un opérateur économique visé à l'article 7 en ce qui concerne la sécurité des produits ;

h) les attentes spécifiques des utilisateurs finaux auxquels le produit est destiné. »¹¹⁴⁹

Les liens entre le défaut et la conformité du système d'IA. Parmi les éléments d'appréciation contenus dans la proposition de directive européenne, la Commission européenne met en avant l'importance du respect des exigences en matière de sécurité des produits. De plus, le texte énonce une présomption de défectuosité lorsque « *le produit n'est pas conforme aux exigences de sécurité obligatoires prévues par le droit de l'Union ou le droit national qui sont destinées à protéger contre le risque de dommage survenu ;* »¹¹⁵⁰. De plus, un système d'IA à haut risque doit se voir apposer le marquage CE lors de sa mise sur le marché. S'il s'avère que ledit logiciel manque à une ou plusieurs obligations de fonctionnement contenues dans le règlement européen sur l'IA, alors le marquage CE correspond à une fausse présentation du produit¹¹⁵¹. Par conséquent, un tel système d'IA serait qualifié de défectueux et une telle caractérisation engagerait ainsi la responsabilité de son fournisseur. Du fait de leur rôle central dans la conformité du système d'IA, les preuves techniques de conception jouent donc un rôle indirect dans la caractérisation du défaut du logiciel.

Les liens entre le défaut et les preuves techniques de conception. Si la non-conformité du système d'IA à haut risque conduit à grandement envisager sa défectuosité, l'inverse n'est pas vrai. Autrement dit, la conformité légale du système d'IA à haut risque ne garantit pas son absence de défaut :

*« Le producteur peut être responsable du défaut alors même que le produit a été fabriqué dans le respect des règles de l'art ou de normes existantes ou qu'il a fait l'objet d'une autorisation administrative. »*¹¹⁵²

En l'espèce, il s'agit d'interroger la possibilité pour un système d'IA à haut risque, satisfaisant les exigences du règlement européen sur l'IA, d'être qualifié de défectueux, plus précisément un défaut de sécurité intrinsèque. En d'autres termes, les preuves techniques, ayant permis

¹¹⁴⁹ Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 28 septembre 2022.

¹¹⁵⁰ Ibid. Article 9.2.b)

¹¹⁵¹ ALHELT K., « The Applicability of the EU Product Liability Directive to Software », *Comp. & Int'l L.J. S. Afr.*, 34, 2001. P. 204. Le critère d'appréciation de la défectuosité du produit est celui relatif à sa présentation.

¹¹⁵² Article 1245-9 du code civil. Voir également : Cour d'appel de d'Aix-en-PROVENCE - ch. 01 C, 16 mars 2017 / n° 2017/238.

d'attester de la conformité du logiciel, peuvent-elles être réutilisées pour prouver la présence ou l'absence d'un défaut intrinsèque du logiciel ?

Ainsi, la Cour d'appel de Toulouse qualifie en 2018 le médicament « Médiator » comme un produit défectueux en raison « *des risques graves liés à l'utilisation du produit que ne justifie pas le bénéfice qui en est attendu.* »¹¹⁵³. S'appuyant sur un ensemble de données et d'articles scientifiques relatifs au rapport bénéfice/risque, les juges caractérisent la défectuosité malgré les décisions d'autorisation remises par les autorités de santé au produit au moment des faits¹¹⁵⁴. De façon similaire, la Cour d'appel de Lyon caractérise un défaut intrinsèque d'une prothèse de hanche¹¹⁵⁵. Avancé par le fabricant, le taux de fracture des prothèses est réutilisé en sa défaveur par les juges de la Cour d'appel (sans pour autant fonder leurs conclusions sur cet élément) :

*« Le moyen invoqué par la société assimilant le taux de fracture, minime (0,022 %) à un aléa thérapeutique n'est pas opérant : le fabricant admettant ainsi implicitement qu'une fraction des prothèses qu'il produit peut céder, pour des raisons purement intrinsèques. »*¹¹⁵⁶

En dépit de la conformité du produit, au regard de l'état de l'art ou du fait d'une autorisation administrative, les juges ont, au sein de ces deux arrêts, caractérisé l'existence d'un défaut intrinsèque du produit. Pour cela, ils se sont appuyés notamment sur des éléments techniques de preuve, que sont le rapport bénéfice/risque et le taux de fracture du produit, pour qualifier la défectuosité du produit. Par analogie, il est donc possible d'envisager qu'un défaut d'un système d'IA puisse être caractérisé en s'appuyant sur le niveau d'exactitude, ou encore le taux de défaillance auquel expose le logiciel.

Cependant, il convient de nuancer cette conclusion pour deux raisons. Premièrement, les décisions relatives à des produits certifiés présentant des défauts de sécurité intrinsèques sont rares. Deuxièmement, les produits mentionnés ne sont pas des logiciels mais des objets matériels. Cette différence doit être prise en compte afin de relativiser la portée de l'analogie. Par conséquent, la réutilisation des preuves techniques de conception pour caractériser le défaut

¹¹⁵³ Cour d'appel de Toulouse - ch. 01 sect. 01, 13 août 2018 / n° 16/01919.

¹¹⁵⁴ Le Médiator a obtenu une AMM le 16 juillet 1974 et a été commercialisé à compter du 27 août 1976. Pour autant, il a fait l'objet d'une suspension d'AMM le 24 novembre 2009 suivie de son retrait définitif le 20 juillet 2010. Toutefois, les faits se sont déroulés de juillet 2001 à mai 2009.

¹¹⁵⁵ Cour d'appel de Lyon - 1ère chambre civile A, 12 octobre 2023 / n° 21/02991.

¹¹⁵⁶ *Ibid.*

intrinsèque d'un système d'IA à haut risque est possible théoriquement mais difficile à envisager concrètement à ce jour.

CONCLUSION DE LA SECTION I :

La responsabilité *ex post* du fournisseur d'un système d'IA à haut risque est intimement liée à la conformité de son produit au règlement européen sur l'IA. La mise sur le marché européen d'un tel logiciel sans certificat européen, ou avec un certificat apposé à tort, est constitutive d'un fait générateur aussi bien pour la responsabilité contractuelle que délictuelle du fournisseur. Dès lors, du fait de la place centrale des preuves techniques dans l'obtention du certificat européen, ces preuves ont des conséquences indirectes importantes sur la responsabilité *ex post* du fournisseur du système d'IA à haut risque. L'utilisation des preuves techniques de conception, qu'ils s'agissent des métriques ou des seuils probabilistes (relatives au niveau de risques, à l'exactitude, au niveau de biais, etc.), pour caractériser un fait générateur de responsabilité est incertaine, à l'exception des spécifications contractuelles. Si elle est possible théoriquement, la réutilisation de ces éléments pour engager la responsabilité du fournisseur du système d'IA est difficilement envisageable. En d'autres termes, il est aujourd'hui difficile d'imaginer qu'un juge qualifie un système d'IA à haut risque de produit défectueux en raison de son fonctionnement alors que celui-ci est conforme au règlement européen sur l'IA. À ce jour, il semble que les preuves techniques de conception aient des conséquences directes faibles sur la responsabilité du fournisseur du système d'IA à haut risque. Ainsi, le certificat européen a donc des effets directs bien plus importants, sur la responsabilité *ex post* du fournisseur d'un système d'IA à haut risque, que les preuves techniques qui ont permis d'obtenir ledit certificat.

Section II : L'articulation inhabituelle avec les droits et libertés fondamentaux

Annnonce de plan. Comme énoncé dans le règlement européen sur l'IA, le système d'IA est un produit au sens du règlement 2019/1020¹¹⁵⁷. Dès lors, il n'est pas surprenant que la conformité

¹¹⁵⁷ « Le règlement (UE) 2019/1020 s'applique aux systèmes d'IA relevant du présent règlement. Aux fins du contrôle effectif de l'application du présent règlement :[...] b) toute référence à un produit en vertu du règlement (UE) 2019/1020 s'entend comme incluant tous les systèmes d'IA relevant du champ d'application du présent règlement. » - « Règlement sur l'intelligence artificielle », 2024. Article 74.

du système d'IA à haut risque aux exigences européennes entraîne des conséquences sur la responsabilité de son fournisseur. Ces interactions entre la responsabilité *ex ante* des fournisseurs et leur responsabilité *ex post* en cas de dommage causé lors de l'utilisation sont traditionnelles dans le cas des produits. Cependant, les systèmes d'IA se distinguent des autres produits réglementés par l'Union européenne en raison du lien unique qu'ils entretiennent avec le régime juridique des droits et libertés fondamentaux (**Sous-section 1**). Dès lors, il est pertinent d'analyser comment les droits et libertés fondamentaux interagissent avec la conformité nouvelle des systèmes d'IA à haut risque mis en place par le règlement européen sur l'IA (**Sous-section 2**).

Sous-section 1 : Le lien particulier entre le fonctionnement du système d'IA et les droits et libertés fondamentaux

Annonce de plan. Les principaux risques entourant le recours aux technologies d'IA portent sur des atteintes aux droits et libertés fondamentaux des déployeurs et des utilisateurs des systèmes d'IA. Si ce constat sonne comme une évidence depuis plusieurs années, le fait qu'un produit puisse engendrer des atteintes importantes aux droits et libertés fondamentaux demeure extrêmement rare d'un point de vue juridique. Au travers notamment du règlement sur l'IA, le législateur européen a reconnu et pris en compte ce lien particulier entre les systèmes d'IA et les droits et libertés fondamentaux des individus (§1). Outre la phase d'utilisation, les potentielles atteintes aux droits et libertés fondamentaux trouvent leur origine également dans le fonctionnement même des systèmes d'IA (§2).

§1. Un lien nouveau reconnu entre les systèmes d'IA et les droits fondamentaux

Les risques traditionnels associés aux produits. Afin de comprendre l'originalité du lien entre les systèmes d'IA et les droits et libertés fondamentaux, il est nécessaire d'analyser les autres produits réglementés de façon similaire à l'IA par l'Union européenne. L'étude se concentre sur les textes contenus dans l'annexe I du règlement européen sur l'IA, à savoir un total de vingt textes. Outre l'objectif relatif au fonctionnement du marché intérieur, ces textes

d'harmonisation poursuivent systématiquement un objectif de protection de « *la santé et la sécurité* » des consommateurs. La conformité est alors entendue uniquement au travers des notions de « sécurité » et de « santé », les exigences essentielles énoncées par les textes se limitant alors à ces deux objectifs¹¹⁵⁸. De façon spécifique, les textes relatifs à l'encadrement de véhicules sur le marché européen prennent en compte la protection de l'environnement¹¹⁵⁹. En dépit du nombre de textes visés, aucun d'entre eux ne mentionne d'objectif de protection des droits fondamentaux dans le corps du texte. Les deux règlements européens relatifs aux dispositifs médicaux de 2017 font figure d'exception en mentionnant les droits fondamentaux de la Charte européenne à deux reprises, et uniquement dans leurs considérants¹¹⁶⁰. Parmi l'ensemble des produits pouvant intégrer des systèmes d'IA comme composants, aucun d'entre eux n'a été perçu par le législateur européen comme pouvant engendrer des risques aux droits et libertés fondamentaux.

La séparation traditionnelle entre les produits réglementés et les atteintes aux droits fondamentaux est confirmée par le règlement (UE) 2019/1020 sur la surveillance du marché et la conformité des produits, commun à tous les textes listés précédemment :

« L'objectif du présent règlement est [...] de veiller à ce que seuls les produits conformes qui répondent aux exigences garantissant un haut niveau de protection des intérêts publics tels que la santé et la sécurité en général, la santé et la sécurité sur le lieu de travail, la protection des consommateurs, la protection de l'environnement et de la sécurité publique et de tout autre intérêt public protégé par ladite législation, soient mis sur le marché de l'Union. »¹¹⁶¹

¹¹⁵⁸ De façon générale, ce constat s'illustre dans l'intitulé même des dispositions relatives aux exigences essentielles, alors nommées « *Exigences essentielles de santé et de sécurité* ». À titre d'exemple, voir « Règlement (UE) 2016/425 du Parlement européen et du Conseil du 9 mars 2016 relatif aux équipements de protection individuelle et abrogeant la directive 89/686/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. « Directive 2006/42/CE du Parlement européen et du Conseil du 17 mai 2006 relative aux machines et modifiant la directive 95/16/CE (refonte) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2006.

¹¹⁵⁹ « Règlement (UE) n° 168/2013 du Parlement européen et du Conseil du 15 janvier 2013 relatif à la réception et à la surveillance du marché des véhicules à deux ou trois roues et des quadricycles Texte présentant de l'intérêt pour l'EEE », OJ L, 2013. « Règlement (UE) n° 167/2013 du Parlement européen et du Conseil du 5 février 2013 relatif à la réception et à la surveillance du marché des véhicules agricoles et forestiers Texte présentant de l'intérêt pour l'EEE », OJ L, 2013.

¹¹⁶⁰ « Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017. Considérants 67 et 89. « Règlement (UE) 2017/746 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic in vitro et abrogeant la directive 98/79/CE et la décision 2010/227/UE de la Commission (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017. Considérants 68 et 89.

¹¹⁶¹ « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019.

Sans pour autant exclure la protection des droits et libertés fondamentaux de son champ d'application¹¹⁶², le règlement européen se concentre principalement sur les risques traditionnels des produits, à savoir la sécurité et la santé des individus. Les droits et libertés fondamentaux y sont mentionnés uniquement comme contraintes à respecter par les autorités de surveillance du marché lors de la mise en application du règlement¹¹⁶³.

La spécificité des risques engendrés par les systèmes d'IA. Le règlement européen sur l'IA fait alors figure d'exception dans le corpus législatif d'harmonisation de l'Union européenne :

« L'objectif du présent règlement est d'améliorer le fonctionnement du marché intérieur et de promouvoir l'adoption de l'intelligence artificielle (IA) axée sur le facteur humain et digne de confiance, tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux consacrés dans la Charte des droits fondamentaux [...] »¹¹⁶⁴

Pour la première fois, une réglementation européenne établissant un régime de conformité du produit énonce explicitement la protection des droits fondamentaux parmi ses objectifs¹¹⁶⁵. Ces droits sont ainsi mentionnés à une centaine de reprises dans le texte¹¹⁶⁶, les rendant ainsi formellement omniprésents. Tout d'abord, le règlement européen sur l'IA pose des exigences spécifiques pour leur protection à l'image de l'analyse d'impact des systèmes d'IA à haut risque sur les droits fondamentaux. Imposée à certains déployeurs¹¹⁶⁷, cette analyse d'impact vient s'ajouter à celle, plus large, réalisée par le fournisseur du logiciel au sein de son système de gestion de risques¹¹⁶⁸. Ensuite, les droits fondamentaux sont aussi également utilisés dans le texte pour préciser la signification de certaines exigences essentielles. Par exemple, l'obligation, imposée au fournisseur d'un système d'IA à haut risque, de réaliser un examen des

¹¹⁶² L'expression « tout autre intérêt public protégé par ladite législation » indique le caractère non exhaustif de cette liste.

¹¹⁶³ « Les autorités de surveillance du marché exercent les pouvoirs énoncés au présent article [...] conformément au droit de l'Union et au droit national, y compris aux principes de la Charte des droits fondamentaux de l'Union européenne [...] » - « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 14.2.

¹¹⁶⁴ « Règlement sur l'intelligence artificielle », 2024. Article 1. Emphase ajoutée par l'auteur.

¹¹⁶⁵ RENONDIN DE HAUTECLOCQUE Stanislas, *Intelligence artificielle, la recherche d'un régime juridique*, Université Pais Cité, 2024.

¹¹⁶⁶ L'expression « droits fondamentaux » est employée 55 fois dans les considérants, 46 fois dans le corps du texte et une fois dans l'annexe IV.

¹¹⁶⁷ « les déployeurs qui sont des organismes de droit public ou des entités privées fournissant des services publics et les déployeurs de systèmes d'IA à haut risque visés à l'annexe III, point 5, b) et c), effectuent une analyse de l'impact sur les droits fondamentaux [...] » - « Règlement sur l'intelligence artificielle », 2024. Article 27.1.

¹¹⁶⁸ Le système de gestion de risques comprend notamment : « l'identification et l'analyse des risques connus et raisonnablement prévisibles que le système d'IA à haut risque peut poser pour la santé, la sécurité ou les droits fondamentaux lorsque le système d'IA à haut risque est utilisé conformément à sa destination ; » - Ibid. Article 9.2.a)

biais sur les données de conception s'est précisée progressivement au fur et à mesure des versions du texte. À l'origine, la proposition de la Commission européenne imposait « *un examen permettant d'identifier de potentiels biais* »¹¹⁶⁹. Or, la notion de biais est extrêmement large du point de vue technique. Désormais, la version finale formule cette obligation de la manière suivante :

*« un examen permettant de repérer d'éventuels biais qui sont susceptibles de porter atteinte à la santé et à la sécurité des personnes, d'avoir une incidence négative sur les droits fondamentaux ou de se traduire par une discrimination interdite par le droit de l'Union [...] »*¹¹⁷⁰

Ainsi, la notion de biais a été limitée, notamment au travers de la référence aux droits fondamentaux. Enfin, la prise en compte de la protection des droits fondamentaux s'illustre par l'organisation du contrôle *ex post* de la conformité des systèmes d'IA à haut risque. À la différence des autres textes européens d'harmonisation, le règlement européen sur l'IA confère des prérogatives de contrôle aux autorités de surveillance du marché mais également aux autorités chargées de la protection des droits fondamentaux¹¹⁷¹. L'implication de ces nouveaux acteurs aux côtés des autorités traditionnelles de contrôle des produits met en lumière, d'une part, l'importance du lien entre les systèmes d'IA et les droits fondamentaux et, d'autre part, l'originalité d'un tel lien avec un produit.

§2. Un lien entre le fonctionnement du système d'IA et l'atteinte aux droits fondamentaux

Le lien entre les systèmes algorithmiques et les droits fondamentaux. Le risque d'atteinte aux droits fondamentaux par un système algorithmique n'est pas un élément unique au règlement européen sur l'IA. Par exemple, le règlement relatif aux services numériques (DSA ci-après) contient également des dispositions qui articulent systèmes algorithmiques et

¹¹⁶⁹ *Ibid.* Article 10.2.f). Version de la proposition de la Commission européenne du 21 avril 2021.

¹¹⁷⁰ *Ibid.* Article 10.2.f).

¹¹⁷¹ *Ibid.* Article 77. Pour plus de détails sur les acteurs en charge du contrôle *ex post* de la conformité des systèmes d'IA à haut risque, voir Section II : Le contrôle *ex post* de la conception du système d'IA.

protection des droits fondamentaux¹¹⁷². Cette interaction s'illustre en particulier au travers de la notion, originale, de « risques systémiques », à savoir :

- a) la diffusion de contenus illicites par l'intermédiaire de leurs services;*
- b) tout effet négatif réel ou prévisible pour l'exercice des droits fondamentaux, [...] ;*
- c) tout effet négatif réel ou prévisible sur le discours civique, les processus électoraux et la sécurité publique ;*
- d) tout effet négatif réel ou prévisible lié aux violences sexistes et à la protection de la santé publique et des mineurs et les conséquences négatives graves sur le bien-être physique et mental des personnes. »¹¹⁷³*

Le DSA impose ainsi aux fournisseurs de très grandes plateformes en ligne ou de très grands moteurs de recherche en ligne une obligation de gestion des risques systémiques¹¹⁷⁴. Le législateur européen considère que les risques systémiques peuvent découler notamment des systèmes algorithmiques, du fait de leur conception ou de leur utilisation¹¹⁷⁵. Dès 2022, le DSA met donc en avant le lien clair entre les systèmes algorithmiques et l'atteinte aux droits fondamentaux.

Le lien entre le fonctionnement du système d'IA et les droits fondamentaux. Malgré cela, le lien, entre les systèmes d'IA et les droits fondamentaux, réalisé au sein du règlement sur l'IA conserve une originalité importante. Là où le DSA se concentre sur les services numériques, le règlement européen sur l'IA s'intéresse principalement aux produits que sont les systèmes d'IA. Ce dernier se concentre davantage sur la conception des systèmes d'IA que sur leur utilisation, en mettant notamment en place des exigences de conformité et une certification obligatoire des

¹¹⁷² « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. Article 1.

¹¹⁷³ *Ibid.* Article 34.1.

¹¹⁷⁴ *Ibid.* Articles 34 et suivants.

¹¹⁷⁵ « [...] tout risque systémique au sein de l'Union découlant de la conception ou du fonctionnement de leurs services et de leurs systèmes connexes, y compris des systèmes algorithmiques, ou de l'utilisation faite de leurs services. » ; « Lorsqu'ils procèdent à des évaluations des risques, les fournisseurs de très grandes plateformes en ligne et de très grands moteurs de recherche en ligne examinent notamment si et comment les facteurs suivants influencent les risques systémiques visés au paragraphe 1 et en tiennent compte: a) la conception de leurs systèmes de recommandation et de tout autre système algorithmique pertinent; b) leurs systèmes de modération des contenus; c) les conditions générales applicables et leur mise en application; d) les systèmes de sélection et de présentation de la publicité; e) les pratiques du fournisseur en matière de données. » - *Ibid.* Article 34.

systèmes d'IA à haut risque¹¹⁷⁶. Ainsi, le législateur européen instaure un lien clair entre le fonctionnement du produit et le risque d'atteinte aux droits fondamentaux.

L'articulation entre le fonctionnement d'un système d'IA et les droits fondamentaux est parfaitement illustré par un arrêt de la CJUE du 26 avril 2022 opposant la République de Pologne au Parlement européen et au Conseil de l'Union européenne¹¹⁷⁷. La décision fait suite à un recours en annulation, initié par la Pologne, de dispositions contenues dans la directive (UE) 2019/790 sur le droit d'auteur et les droits voisins dans le marché unique numérique¹¹⁷⁸. Plus précisément, la Pologne demande, à titre principal, d'annuler les points b) et c) de l'article 17 paragraphe 4 relatif à l'utilisation de contenus protégés par des fournisseurs de services de partage de contenus en ligne :

« 1. Les États membres prévoient qu'un fournisseur de services de partage de contenus en ligne effectue un acte de communication au public ou un acte de mise à la disposition du public aux fins de la présente directive lorsqu'il donne au public l'accès à des œuvres protégées par le droit d'auteur ou à d'autres objets protégés qui ont été téléversés par ses utilisateurs.

[...]

4. Si aucune autorisation n'est accordée, les fournisseurs de services de partage de contenus en ligne sont responsables des actes non autorisés de communication au public, y compris la mise à la disposition du public, d'œuvres protégées par le droit d'auteur et d'autres objets protégés, à moins qu'ils ne démontrent que :

a) ils ont fourni leurs meilleurs efforts pour obtenir une autorisation ; et

b) ils ont fourni leurs meilleurs efforts, conformément aux normes élevées du secteur en matière de diligence professionnelle, pour garantir l'indisponibilité d'œuvres et autres objets protégés spécifiques pour lesquels les titulaires de droits ont fourni aux fournisseurs de services les informations pertinentes et nécessaires ; et en tout état de cause

c) ils ont agi promptement, dès réception d'une notification suffisamment motivée de la part des titulaires de droits, pour bloquer l'accès aux œuvres et autres objets protégés faisant l'objet de la notification ou pour les retirer de

¹¹⁷⁶ RENONDIN DE HAUTECLOCQUE Stanislas, *Intelligence artificielle, la recherche d'un régime juridique*, Université Pais Cité, 2024.

¹¹⁷⁷ Cour de justice de l'Union européenne, 26 avril 2022, n° C-401/19. BENSAMOUN Alexandra, « Filtrage des contenus par les plateformes contributives et liberté d'expression », *Légipresse*, 2021.

¹¹⁷⁸ « Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019.

leurs sites internet, et ont fourni leurs meilleurs efforts pour empêcher qu'ils soient téléversés dans le futur, conformément au point b). »¹¹⁷⁹

Au travers de ces dispositions visant à protéger le droit d'auteur, la directive européenne introduit un régime de responsabilité pour les fournisseurs de service de partage de contenus¹¹⁸⁰. Afin de bénéficier de l'exonération de responsabilité de l'article 17 paragraphe 4, ces acteurs doivent mettre en place des outils de filtrage automatique des contenus téléversés par les utilisateurs¹¹⁸¹. Selon la Pologne, le recours à ces systèmes algorithmiques pour contrôler préventivement le contenu téléversé représente une « *ingérence particulièrement grave dans le droit à la liberté d'expression et d'information des utilisateurs* »¹¹⁸². La République de Pologne reproche ainsi au législateur européen de ne pas respecter un « *juste équilibre* » entre, d'un côté, le droit d'auteur des titulaires de droit et, de l'autre côté, le droit à la liberté d'expression et d'information des utilisateurs¹¹⁸³. Dans ce contexte de mise en balance entre deux droits fondamentaux¹¹⁸⁴, les juges européens reconnaissent l'existence d'une limitation du droit à la liberté d'expression et d'information¹¹⁸⁵ et sont donc amenés à s'interroger sur la proportionnalité de cette atteinte. À ce titre, la qualité de fonctionnement du système de filtrage apparaît être une condition pour caractériser la proportionnalité de l'atteinte à la liberté d'expression :

« un système de filtrage qui risquerait de ne pas suffisamment distinguer entre un contenu illicite et un contenu licite, de telle sorte que son déploiement pourrait avoir pour effet d'entraîner le blocage de

¹¹⁷⁹ Ibid. Article 17.

¹¹⁸⁰ « Le fournisseur d'un service de la société de l'information dont l'objectif principal ou l'un des objectifs principaux est de stocker et de donner au public l'accès à une quantité importante d'œuvres protégées par le droit d'auteur ou d'autres objets protégés qui ont été téléversés par ses utilisateurs, qu'il organise et promeut à des fins lucratives. » - Ibid. Article 2.6).

¹¹⁸¹ Ce constat est partagé par la CJUE : « [...] pour pouvoir effectuer un tel contrôle préalable, les fournisseurs de services de partage de contenus en ligne sont, en fonction du nombre de fichiers téléversés et du type d'objet protégé en question, dans les limites énoncées à l'article 17, paragraphe 5, de la directive 2019/790, contraints de recourir à des outils de reconnaissance et de filtrage automatiques. » - Cour de justice de l'Union européenne, 26 avril 2022, n° C-401/19. Par. 54.

¹¹⁸² Ibid. Par. 41.

¹¹⁸³ « le législateur de l'Union aurait méconnu le juste équilibre entre la protection des titulaires de droits et celle des utilisateurs de services de partage de contenus en ligne [...] » - Ibid. Par. 61.

¹¹⁸⁴ « La propriété intellectuelle est protégée. » - « Charte des droits fondamentaux de l'Union européenne », OJ C, 2012. Article 17.2. « Toute personne a droit à la liberté d'expression. Ce droit comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontières. » - Ibid. Article 11.1.

¹¹⁸⁵ « il y a lieu de conclure que le régime de responsabilité spécifique, instauré à l'article 17, paragraphe 4, de la directive 2019/790 pour les fournisseurs de services de partage de contenus en ligne, comporte une limitation de l'exercice du droit à la liberté d'expression et d'information des utilisateurs de ces services de partage, garanti à l'article 11 de la Charte. » - Cour de justice de l'Union européenne, 26 avril 2022, n° C-401/19. Par. 58.

communications à contenu licite, serait incompatible avec le droit à la liberté d'expression et d'information »¹¹⁸⁶

En d'autres termes, un système algorithmique, ayant pour finalité la détection et le blocage de contenus illicites, qui présenterait un taux d'erreur (plus précisément un taux de faux positifs) trop important est incompatible avec la Charte européenne des droits fondamentaux. Pour reprendre l'expression de l'avocat général, il est nécessaire de « *minimiser l'effet collatéral d'une mesure de filtrage et de blocage* »¹¹⁸⁷. En vertu de la jurisprudence de la CJUE¹¹⁸⁸ et de la CEDH¹¹⁸⁹, l'avocat général précise qu'il s'agit d'une exigence distincte et cumulative avec les garanties procédurales¹¹⁹⁰. Par conséquent, la qualité de fonctionnement d'un tel système algorithmique est une condition *sine qua non* de la proportionnalité de l'atteinte au droit fondamental qu'est le droit à la liberté d'expression et d'information. Dès lors, il existe un lien direct entre l'exactitude d'un système d'IA et le droit fondamental à la liberté d'expression et d'information. Ainsi, il apparaît que les caractéristiques techniques relatives au fonctionnement d'un système d'IA peuvent avoir des conséquences sur la légalité des atteintes aux droits fondamentaux du fait de l'utilisation dudit logiciel.

Sous-section 2 : Les interactions entre le fonctionnement du système d'IA à haut risque et les droits et libertés fondamentaux

Annonce de plan. Il existe donc un lien entre le fonctionnement du système d'IA et le respect des droits fondamentaux des utilisateurs du logiciel. Cette interaction entre des éléments issus de la responsabilité *ex ante* et de la responsabilité *ex post* du fait de l'IA prend deux formes. Dans un premier temps, les éléments relatifs à la qualité de fonctionnement du système d'IA à haut risque, déjà exigés par le règlement européen sur l'IA, présentent une utilité importante pour caractériser une atteinte aux droits fondamentaux (§1). Dans un second temps, la

¹¹⁸⁶ *Ibid.* Par. 86. La CJUE reprend un argument déjà énoncé en 2012 dans l'arrêt Cour de justice de l'Union européenne, 16 février 2012, n° C-360/10. Par. 50 et 51.

¹¹⁸⁷ Cour de justice de l'Union européenne, 15 juillet 2021, n° C-401/19.

¹¹⁸⁸ Cour de justice, 15 septembre 2016, n° C-484/14 Cour de justice de l'Union européenne, 27 mars 2014, n° C-314/12.

¹¹⁸⁹ « Any indiscriminate blocking measure which interferes with lawful content Or, websites as a collateral effect of a measure aimed at illegal content Or, websites amounts to arbitrary interference with the rights of owners of such websites » - Cour européenne des droits de l'Homme, 16 novembre 2020, n° 10795/14. Par. 46.

¹¹⁹⁰ « L'existence de telles garanties procédurales ne dispense pas les autorités publiques de veiller à minimiser l'effet collatéral d'une mesure de filtrage et de blocage. Il s'agit là d'exigences distinctes, et cumulatives. » - Cour de justice de l'Union européenne, 15 juillet 2021, n° C-401/19. Par. 181.

qualification d'une violation d'un ou plusieurs droits fondamentaux peut amener à interroger la conformité du système d'IA à haut risque, et donc la responsabilité *ex post* de son fournisseur (2).

§1. La réutilisation des exigences réglementaires de conception du système d'IA à haut risque : l'exemple de la discrimination algorithmique

Annonce de plan. Dans le cas d'un système d'IA à haut risque, les caractéristiques de fonctionnement du logiciel pertinentes pour caractériser une atteinte aux droits fondamentaux sont similaires à celles posées par le règlement européen sur l'IA. Dès lors, ces éléments, déjà contrôlés dans le cadre de la responsabilité *ex ante* du fournisseur, peuvent être employés une seconde fois pour attester de l'atteinte aux droits fondamentaux lors de l'utilisation du logiciel par le déployeur. Afin d'illustrer cette réutilisation, il sera développé l'exemple d'une discrimination indirecte causée par une décision prise par un système d'IA à haut risque. Ainsi, les caractéristiques de fonctionnement du logiciel sont des éléments de preuve pertinents à la fois pour le demandeur lors de l'identification d'une différence de traitement (A) mais également pour le défendeur (le déployeur) lors de la justification de cette différence de traitement (B).

A- La caractérisation d'une différence de traitement

La preuve d'une différence de traitement subie par le demandeur. En droit européen de la non-discrimination, la preuve d'une discrimination, en particulier indirecte, peut être apportée par tous moyens¹¹⁹¹. Au niveau national, le principe de liberté de la preuve est maintenu, le

¹¹⁹¹ « L'appréciation des faits qui permettent de présumer l'existence d'une discrimination directe ou indirecte appartient à l'instance judiciaire nationale ou à une autre instance compétente, conformément au droit national ou aux pratiques nationales, qui peuvent prévoir, en particulier, que la discrimination indirecte peut être établie par tous moyens, y compris sur la base de données statistiques ». - « Directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail », 303, 2000. Considérant 15. Directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en œuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique, 180 2000. Considérant 15.

droit pénal et le droit civil ne prévoyant pas de règles de preuves spécifiques¹¹⁹². Dans le cas d'une discrimination indirecte, la preuve de la différence de traitement est apportée principalement de façon statistique¹¹⁹³, aussi bien en droit national qu'en droit européen¹¹⁹⁴. Le recours à des statistiques permet au demandeur de démontrer qu'il subit un désavantage spécifique en dépit d'une égalité formelle de traitement¹¹⁹⁵.

La preuve de ce désavantage peut reposer d'une part sur la comparaison statistique entre deux groupes de population¹¹⁹⁶. Tout d'abord, il faut identifier un groupe de « référence » puis un groupe « d'arrivée », qui est le groupe ayant subi la décision ou l'acte discriminant. Les individus de ces groupes sont répartis en deux classes, la classe A qui est la classe dominante ou majoritaire et la classe B qui est la classe minoritaire et généralement défavorisée. L'objectif est de comparer la proportion entre les classes A et B avant et après le recours à la décision ou à l'acte potentiellement discriminant. Prenons l'exemple de l'élaboration de lignes directrices pour le recrutement dans une entreprise. Le groupe de « référence » représente le groupe de candidats qui se présente en entretien et le groupe « d'arrivée » représente le groupe d'individus sélectionnés pour un poste. Dans cette illustration, la classe A représente les hommes tandis que la classe B désigne les femmes. Si la proportion de femmes parmi les candidats est significativement plus faible que la proportion de femmes parmi les personnes sélectionnées, alors cela peut entraîner la présomption que ces lignes directrices de recrutement, bien qu'en apparence neutres, désignent des critères de sélection qui défavorisent les femmes par rapport aux hommes.

La logique de comparaison statistique entre deux groupes d'individus dans le règlement européen sur l'IA. Au sein du processus de preuve de la différence de traitement, la logique de comparaison entre deux groupes de population fait écho avec plusieurs exigences du règlement européen sur l'IA, en particulier celles relatives aux biais. Tout d'abord, les

¹¹⁹² En droit civil, voir l'article 1358 du code civil : « *Hors les cas où la loi en dispose autrement, la preuve peut être apportée par tout moyen.* ». Pour les litiges pénaux, voir l'article 427 du code de procédure pénale : « *Hors les cas où la loi en dispose autrement, les infractions peuvent être établies par tout mode de preuve et le juge décide d'après son intime conviction.* ».

¹¹⁹³ La preuve statistique n'est toutefois pas indispensable. Il est possible de démontrer que la décision est susceptible, par nature ou intrinsèquement, de défavoriser les personnes appartenant à une catégorie protégée contre la discrimination.

¹¹⁹⁴ « *The recent case-law of the Court of Justice of the European Communities [...] shows that it permits claimants to rely on statistical evidence and the national courts to take such evidence into account where it is valid and significant.* » - Cour européenne des droits de l'Homme, 13 novembre 2007, n° 57325/00. Par. 187

¹¹⁹⁵ Il s'agit d'une condition *sine qua non* à l'existence d'une discrimination indirecte. Si ce critère n'est pas satisfait, alors il s'agit d'une discrimination directe.

¹¹⁹⁶ DE SCHUTTER Olivier et SCHUTTER Olivier De, « Prouver la discrimination », *ERA Forum*, 4, 2003. P.47.

dispositions relatives aux données de conception imposent au fournisseur d'un système d'IA à haut risque d'évaluer ses jeux de données notamment sous le prisme de catégories d'individus :

« Les jeux de données d'entraînement, de validation et de test sont pertinents, suffisamment représentatifs et, dans toute la mesure possible, exempts d'erreurs et complets au regard de la destination. Ils possèdent les propriétés statistiques appropriées, y compris, le cas échéant, en ce qui concerne les personnes ou groupes de personnes à l'égard desquels le système d'IA à haut risque est destiné à être utilisé. »¹¹⁹⁷

Non définie par le texte, l'exigence de représentativité des jeux de données renvoie à l'importance que l'échantillon de données utilisé par le fournisseur du système d'IA à haut risque corresponde aux futures personnes affectées par le logiciel¹¹⁹⁸. Par exemple, il n'est pas souhaitable qu'un système d'IA destiné à être utilisé sur une proportion très largement féminine (70%) soit conçu à l'aide de jeux de données composés à 90% de données issues d'hommes. Les jeux de données de conception ne seraient alors pas suffisamment représentatifs. Il apparaît clairement ici la même logique de comparaison statistique de groupes qu'énoncé précédemment pour prouver la différence de traitement subie par un individu. En effet, le groupe de « référence » est ici les futures personnes affectées et le groupe « d'arrivée » serait les jeux de données de conception. La proportion d'hommes (classe A) et de femmes (classe B) dans chacun de ces groupes permet d'évaluer la représentativité des jeux de conception utilisés.

La réutilisation de l'exigence d'exactitude pour prouver la différence de traitement. Si elle permet d'illustrer clairement la logique de comparaison statistique de groupes, la représentativité des jeux de données de conception n'est pas une caractéristique de conception permettant de prouver une différence de traitement. En effet, l'absence d'une représentativité suffisante ne signifie pas que le demandeur a subi un désavantage spécifique du fait du recours au système d'IA à haut risque.

Toutefois, l'exigence d'exactitude imposée au fournisseur d'un système d'IA à haut risque reprend cette même logique de comparaison statistique mais d'une manière bien plus pertinente pour attester de la différence de traitement subie par le demandeur. Pour rappel, le législateur européen impose que le système d'IA à haut risque possède « *un niveau approprié d'exactitude [...]* »¹¹⁹⁹. L'exactitude du logiciel comprend, d'une part, le niveau global d'exactitude et,

¹¹⁹⁷ « Règlement sur l'intelligence artificielle », 2024. Article 10.

¹¹⁹⁸ LAROUSSE Éditions, « Définitions ».

¹¹⁹⁹ « Règlement sur l'intelligence artificielle », 2024. Article 15.1.

d'autre part, « *le degré d'exactitude pour des personnes ou des groupes de personnes spécifiques à l'égard desquels le système est destiné à être utilisé [...]* »¹²⁰⁰. En d'autres termes, le fournisseur du système d'IA à haut risque doit mesurer le niveau d'exactitude du logiciel pour différents groupe d'individus. On retrouve alors la logique de comparaison statistique entre des groupes d'individus. Il apparaît deux groupes de population avec, d'un côté, les individus pour lesquels le logiciel a produit un résultat correct et, de l'autre côté, les individus pour lesquels le logiciel s'est trompé. À nouveau, chacun de ces groupes peut être divisé en deux classes avec les hommes et les femmes. Ainsi, les informations communiquées par le fournisseur pourront permettre de comparer la proportion théorique d'hommes faisant l'objet d'erreurs par rapport à celle de femmes. L'évolution du taux d'erreur selon la population à laquelle appartient un individu est alors une information particulièrement pertinente pour prouver une différence de traitement. Par conséquent, l'exactitude d'un système d'IA à haut risque, caractéristique technique exigée par le règlement européen sur l'IA, peut être réutilisée comme élément de preuve de la présence, ou de l'absence, d'une différence de traitement causée par le recours au logiciel.

B- La justification de la discrimination

Les preuves de la justification objective de la différence de traitement. Une fois que le demandeur parvient à prouver l'existence de traitement, la charge de la preuve est renversée¹²⁰¹. Dès lors, c'est au défendeur de contester la différence de traitement avancée ou démontrer que celle-ci est justifiée objectivement « *par un but légitime et que les moyens pour parvenir à ce but ne soient appropriés et nécessaires* »¹²⁰². Pour être considérée comme justifiée, une

¹²⁰⁰ *Ibid.* Annexe IV.3.

¹²⁰¹ Cet allègement de la charge de la preuve a été mis en place au niveau européen par les différentes directives relatives à l'égalité (« Directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en oeuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique », 180, 2000 « Directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail », 303, 2000 « Directive 2004/113/CE du Conseil du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans l'accès ` des biens et services et la fourniture de biens et services », 373, 2004 « Directive 2006/54/CE du Parlement européen et du Conseil du 5 juillet 2006 relative à la mise en œuvre du principe de l'égalité des chances et de l'égalité de traitement entre hommes et femmes en matière d'emploi et de travail (refonte) », 204, 2006.) et transposées en droit français par la loi du 27 mai 2008 (« LOI n° 2008-496 du 27 mai 2008 portant diverses dispositions d'adaptation au droit communautaire dans le domaine de la lutte contre les discriminations (1) », 2008-496, 2008.)

¹²⁰² « Directive 2004/113/CE du Conseil du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans l'accès ` des biens et services et la fourniture de biens et services », 373, 2004. Article 2.b).

discrimination doit satisfaire un test de proportionnalité composé de quatre étapes : (a) l'existence d'un but légitime ; (b) le caractère adapté du critère utilisé ; (c) la nécessité dudit critère ; et (d) la proportionnalité de l'utilisation du critère¹²⁰³. Parmi ces étapes, l'évaluation de l'adéquation et la nécessité du critère litigieux s'articulent particulièrement bien avec les exigences de conformité contenues dans le règlement européen sur l'IA.

La preuve de l'adéquation du critère neutre utilisé. En droit européen, le défendeur peut justifier le recours au critère neutre, dont les impacts sont discriminants, en démontrant dans un premier temps que ce dernier est adapté au but légitime poursuivi. Cette condition d'adéquation du critère neutre implique donc qu'il doit être efficace dans la réalisation de l'objectif¹²⁰⁴. Pour prouver cela, les affirmations du défendeur sont insuffisantes aux yeux de la CJUE qui impose l'apport de preuves concrètes¹²⁰⁵. Pour rappel, un système d'IA à haut risque certifié signifie que le logiciel présente un niveau approprié d'exactitude¹²⁰⁶. Dans le contexte d'une décision discriminante produite par un système d'IA à haut risque, l'efficacité du critère neutre se rapproche donc sensiblement du niveau d'exactitude du logiciel. Le déployeur du système d'IA à haut risque, à savoir le défendeur dans notre illustration, peut alors s'appuyer sur les informations contenues dans la notice d'utilisation que lui a communiquée le fournisseur du logiciel. Celle-ci contient notamment « *le niveau d'exactitude, y compris les indicateurs utilisés [...]* »¹²⁰⁷ du logiciel ainsi que « *sa performance en ce qui concerne des personnes ou groupes de personnes spécifiques à l'égard desquels le système est destiné à être utilisé* »¹²⁰⁸. En d'autres termes, les informations relatives à l'exactitude et, plus généralement, à la performance du logiciel peuvent être réutilisées pour attester de la pertinence du critère neutre utilisé par le système d'IA ayant produit la décision discriminante.

La nécessité du critère neutre litigieux employé par le système d'IA. La nécessité de la pratique supposée discriminante s'évalue en observant s'il n'existe aucune autre méthode qui présenterait une efficacité égale (ou supérieure) tout en étant moins discriminante. L'obligation d'identifier une alternative aussi efficace met à nouveau en avant l'exigence d'exactitude et de

¹²⁰³ Philipp Hacker, « Teaching Fairness to Artificial Intelligence : Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law » (2018) 55 Common Market Law Review 1143-1186 (2018) (Common Market Law Review) 35.

¹²⁰⁴ *Ibid.*

¹²⁰⁵ « *De simples affirmations générales concernant l'aptitude d'une mesure déterminée à promouvoir l'embauche ne suffisent pas pour faire apparaître que l'objectif de la règle litigieuse est étranger à toute discrimination fondée sur le sexe ni à fournir des éléments permettant raisonnablement d'estimer que les moyens choisis étaient aptes à la réalisation de cet objectif.* » Cour de justice de l'Union européenne, 9 février 1999, n° C-167/97. Par. 76.

¹²⁰⁶ « Règlement sur l'intelligence artificielle », 2024. Article 15.1.

¹²⁰⁷ *Ibid.* Article 13.3.b).ii)

¹²⁰⁸ *Ibid.* Article 13.3.b).v)

performance du système d'IA à haut risque. Dans un article de 2018, le professeur de droit Philipp Hacker identifie seulement deux situations dans lesquelles un système d'IA discriminant peut ne pas être nécessaire¹²⁰⁹. Premièrement, il s'agit du cas où la discrimination est causée par des données d'entraînement biaisées et qu'un jeu de données d'entraînement équivalent et au même coût mais moins biaisé existait et n'a pas été choisi par le fabricant. Deuxièmement, il existe un modèle alternatif de système d'IA qui présente les mêmes performances que le modèle litigieux mais avec une discrimination moins importante. Selon l'auteur, ces deux situations seront en pratique très rares dès lors que le fabricant a pris un tant soit peu en considération la question des biais. Outre les renseignements sur l'exactitude du logiciel, la notice d'utilisation contient « [...] toute autre information pertinente concernant les jeux de données d'entraînement, de validation et de tests utilisés, compte tenu de la destination du système d'IA à haut risque »¹²¹⁰. À nouveau, le déployeur peut réutiliser les informations relatives aux exigences du règlement européen sur l'IA pour démontrer cette étape du test de proportionnalité du critère.

§2. De l'atteinte aux droits fondamentaux à la conformité du système d'IA

La remise en question de la conformité du système d'IA à haut risque. Afin de démontrer qu'il respecte les droits fondamentaux, le déployeur d'un système d'IA à haut risque peut réutiliser des éléments de la conformité du logiciel. Cependant, la certification d'un système d'IA à haut risque ne signifie pas nécessairement que son fonctionnement garantisse un respect absolu des droits fondamentaux¹²¹¹. Le travail réalisé lors de la mise en conformité *ex ante* du logiciel est donc évalué par le juge, dans les limites de l'action en réparation.

À titre d'exemple, prenons le cas d'un système d'IA destiné à évaluer des candidats pour un recrutement¹²¹² qui entraînerait un désavantage particulier pour le demandeur. Comme vu

¹²⁰⁹ HACKER Philipp, « Teaching Fairness to Artificial Intelligence », 55 *Common Market Law Review* 1143-1186 (2018), 2018.

¹²¹⁰ « Règlement sur l'intelligence artificielle », 2024. Article 13.3.b).vi)

¹²¹¹ Le cas où l'atteinte à un droit fondamental soit causée par l'utilisation du système d'IA à haut risque est exclu de la présente analyse. Bien qu'elle soit possible, cette situation n'entre pas dans le périmètre de l'étude car elle ne présente aucun lien avec la phase de conception du logiciel.

¹²¹² Le système d'IA est donc un système d'IA à haut risque au sens du règlement européen sur l'IA : « systèmes d'IA destinés à être utilisés pour le recrutement ou la sélection de personnes physiques, en particulier pour [...] évaluer les candidat ; » - « Règlement sur l'intelligence artificielle », 2024. Annexe III.4.a).

précédemment, pour démontrer l'efficacité de la méthode utilisée (ici le système d'IA), le déployeur peut mettre en avant le niveau d'exactitude du logiciel. Dans le cas où le juge considère que l'exactitude du logiciel n'est pas suffisante pour considérer le système d'IA à haut risque comme une méthode efficace pour atteindre l'objectif de sélection des candidats, deux conséquences sont envisageables. Premièrement, la différence de traitement n'est pas justifiée objectivement. Par conséquent, le déployeur porte une atteinte disproportionnée au droit à la non-discrimination du demandeur. Deuxièmement, une telle décision judiciaire pose une seconde question, plus indirecte : si le niveau d'exactitude d'un système d'IA n'est pas suffisant pour être qualifié d'efficace au sens du droit à la non-discrimination, peut-il être qualifié comme « *approprié* » au sens du règlement européen sur l'IA ? De façon plus générale, si l'utilisation d'un système d'IA à haut risque porte atteinte, du fait de son fonctionnement, à un droit fondamental, le niveau des risques résiduels du logiciel peut-il être considéré comme « *acceptable* » au sens du règlement européen sur l'IA ?¹²¹³

Selon la cause de l'atteinte, la caractérisation d'une violation d'un droit fondamental par un système d'IA à haut risque pourrait donc engager, d'une part, la responsabilité *ex post* du fournisseur d'IA pour faute et, d'autre part, sa responsabilité *ex ante* en remettant en cause de façon globale la certification de son logiciel. En d'autres termes, l'évaluation de l'atteinte à un droit fondamental par un système d'IA à haut risque peut conduire à questionner la conformité réglementaire de ce même logiciel.

L'articulation des droits fondamentaux avec les exigences de fonctionnement des systèmes d'IA à haut risque. L'évaluation d'une atteinte aux droits fondamentaux peut générer des conséquences au-delà de la conformité du système d'IA à haut risque litigieux. De façon plus globale, l'appréciation par le juge des caractéristiques techniques du système d'IA, établies par le règlement européen sur l'IA, peut remettre en question les exigences de conformité énoncées par le texte européen, ou du moins l'interprétation qui en a été faite par les acteurs impliqués. Par exemple, en décidant que le niveau d'exactitude atteint par le système d'IA à haut risque certifié n'est pas suffisant pour être considéré comme « efficace », au sens du droit à la non-discrimination, le juge caractérise aussi l'insuffisance de ce seuil d'exactitude pour tous les systèmes d'IA commercialisés pour une finalité similaire. Par conséquent, l'ensemble des fournisseurs de systèmes d'IA similaires devront concevoir leur logiciel en prenant en compte

¹²¹³ « Les mesures de gestion des risques visées au paragraphe 2, point d), sont telles que le risque résiduel pertinent associé à chaque danger ainsi que le risque résiduel global lié aux systèmes d'IA à haut risque sont jugés acceptables. » - Ibid. Article 9.5.

cette décision judiciaire. Le juge contribue donc à préciser la notion « appropriée » pour l'ensemble des fournisseurs du système d'IA, dans un contexte donné. Au travers du prisme des droits fondamentaux, le juge peut à la fois engager la responsabilité *ex post* du fournisseur du système d'IA à haut risque et, indirectement, se prononcer sur l'interprétation des exigences constitutives du régime de responsabilité *ex ante* des fournisseurs.

CONCLUSION DE LA SECTION II :

Avec le règlement européen sur l'IA, le législateur européen reconnaît l'existence d'un lien unique entre les droits fondamentaux et ce nouveau produit qu'est le système d'IA. Dès lors, il apparaît une interaction entre certaines exigences de conception, issues du régime de responsabilité *ex ante* des fournisseurs de système d'IA à haut risque, et le régime de protection des droits fondamentaux, dont le manquement peut engager la responsabilité *ex post* des fournisseurs. Cette articulation entre les règles et notions provenant de ces deux régimes de responsabilité est à double sens. Tout d'abord, le déployeur, ayant utilisé le système d'IA litigieux, peut employer certains éléments de conformité du logiciel pour contester ou justifier l'atteinte avancée par le demandeur. Par cette action, il soumet à l'évaluation du juge les résultats du processus de mise en conformité du système d'IA à haut risque. Si le juge caractérise tout de même la violation en raison du fonctionnement du système d'IA, la conformité logiciel peut alors être interrogée. Ainsi, la caractérisation d'une atteinte aux droits fondamentaux, causée par un résultat unique du système d'IA, peut avoir des conséquences dépassant le litige. Le fournisseur dudit système d'IA pourrait alors voir sa responsabilité *ex post* engagée sur le fondement de la responsabilité délictuelle, voire également sur la base de sa responsabilité contractuelle. Cette articulation à double sens est alors l'occasion d'interroger l'interprétation, réalisée *ex ante*, de certaines exigences et notions juridiques par le biais du choix des preuves techniques. Au travers des droits fondamentaux, le juge peut alors venir préciser ou s'opposer à ces interprétations. De façon indirecte, il pourra alors se prononcer sur l'adéquation des processus de mise en place de ces exigences, telle que la quantification de certaines notions juridiques en seuils statistiques et en métriques.

CONCLUSION DU CHAPITRE 1 :

Comme de nombreuses législations encadrant les produits, le régime de conformité élaboré par le règlement européen sur l'IA a des conséquences directes sur la caractérisation du fait générateur de responsabilité du fabricant, ici le fournisseur d'un système d'IA à haut risque. Classiquement, l'obtention de la certification européenne par le logiciel d'IA réduit grandement les possibilités d'engager la responsabilité *ex post* du fournisseur. Tant qu'elle n'impacte pas cette certification, la qualité de fonctionnement du système est rarement un élément décisif pour caractériser un fait générateur au regard des régimes de responsabilité traditionnels appliqués aux produits. Cependant, le règlement européen sur l'IA s'articule également avec un régime de responsabilité inhabituel pour un produit : le régime de protection des droits fondamentaux. Ainsi, le déploiement d'un système d'IA à haut risque peut entraîner une atteinte aux droits fondamentaux de certains utilisateurs, en raison de sa mauvaise utilisation mais également du fait de sa conception. Dans le second cas, la responsabilité *ex post* du fournisseur d'IA peut alors être engagée. De plus, en raison de l'obligation de minimisation des risques générés par son produit sur les droits fondamentaux prévue par le règlement européen sur l'IA, la caractérisation d'une violation d'un droit fondamental peut remettre en question la validité, totale ou partielle, de la certification du système d'IA à haut risque. Par conséquent, une atteinte aux droits fondamentaux peut avoir des conséquences à la fois sur la responsabilité *ex post* et *ex ante* du fournisseur d'un système d'IA à haut risque.

L'articulation entre le régime de responsabilité *ex ante* et les régimes de responsabilité *ex post* applicables au fournisseur d'un système d'IA à haut risque renforce davantage l'importance des preuves techniques de conception. Déjà centrales dans l'interprétation des exigences substantielles du règlement européen sur l'IA, ces preuves pourraient donc être réutilisées, autant par le défendeur que le demandeur lésé, dans le cadre d'un dommage causé par un système d'IA à haut risque.

CHAPITRE 2 : LA REUTILISATION DES PREUVES DE CONCEPTION LORS D'UN DOMMAGE CAUSE PAR UN SYSTEME D'IA

Introduction. En raison de l'articulation entre le régime de responsabilité *ex ante* et les régimes de responsabilité civile du fait de l'IA, les éléments de preuve permettant d'attester du respect des obligations de conception sont destinés à jouer aussi un rôle dans la démonstration des faits constitutifs de la responsabilité *ex post*. En d'autres termes, le recours à la conformité du système d'IA à haut risque par un acteur pour tenter de s'exonérer de sa responsabilité (ou par un demandeur pour engager la responsabilité du fournisseur) se traduit par une réutilisation des éléments de preuves de conception dans le contexte d'un régime de responsabilité *ex post*. Les preuves techniques de conception, qui occupent déjà une place centrale dans le règlement européen sur l'IA, peuvent donc entraîner des conséquences dans le cadre d'une action en réparation. L'analyse de la place de ces preuves de conception réutilisées sera ainsi analysée sous le prisme de l'accès à la preuve des faits constitutifs de la responsabilité du fait de l'IA. L'accès à la preuve est une composante du droit à la preuve, ce dernier étant lui-même un principe issu du droit fondamental à un procès équitable¹²¹⁴. Pour rappel, le droit à la preuve désigne, d'une part, le droit d'une partie de soumettre au juge les preuves à sa disposition et, d'autre part, son droit de solliciter le juge pour accéder aux preuves qu'elle ne détient pas, par le biais des mesures d'instruction ou de production forcée¹²¹⁵. Dans le cadre d'une action en réparation du fait de l'IA, un accès facilité ou difficile aux preuves techniques de conception accroît ou diminue leur réutilisation par les parties. Par conséquent, les modalités d'accès à la preuve permettent d'étudier l'interaction, sous le prisme des éléments de preuve, entre le droit des systèmes d'IA et la responsabilité *ex post* du fait de l'IA.

Annonce de plan. L'analyse de la réutilisation des éléments de preuves issus du règlement européen sur l'IA dans le cadre d'une action en réparation nécessite, tout d'abord, d'identifier lesdits éléments de preuves (**Section I**) avant de s'intéresser aux mécanismes juridiques permettant leur réutilisation dans ce contexte (**Section II**).

¹²¹⁴ « Tout accusé a droit notamment à : [...] d) interroger ou faire interroger les témoins à charge et obtenir la convocation et l'interrogation des témoins à décharge dans les mêmes conditions que les témoins à charge ; » - CONSEIL DE L'EUROPE, « Convention européenne des droits de l'Homme », 1950. Article 6 alinéa 3. Le droit à la preuve est un des nombreux principes qui composent le droit à un procès équitable (principe d'égalité des armes, principe du contradictoire), voir : Cour européenne des droits de l'Homme, 23 juin 1993, n° 12952/87.

¹²¹⁵ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022 BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010.

Section I : L'apparition de nouvelles preuves issues de la conformité du système d'IA

Annonce de plan. Les preuves disponibles pour la personne lésée sont déterminantes pour sa capacité à, premièrement, constater qu'elle a fait l'objet d'une décision dommageable et, secondement, obtenir réparation pour le préjudice subi. Ainsi, l'individu lésé dispose d'un panel de mécanismes juridiques lui permettant de rassembler des éléments de preuve relatifs à la décision préjudiciable prise par un système d'IA. Toutefois, le fonctionnement des systèmes d'IA présente des particularités qui les différencient des produits traditionnels. Par conséquent, les besoins probatoires du demandeur, ou du demandeur potentiel, présentent certaines spécificités qui doivent être mises en lumière (**Sous-section 1**). Ensuite, il conviendra d'analyser l'accessibilité des différents éléments de preuves existants et leur pertinence à la lumière des besoins probatoires de l'individu lésé (**Sous-section 2**).

Sous-section 1 : Le besoin d'accès la preuve en cas de dommage causé par un système d'IA

Annonce de plan. Les technologies d'IA présentent des spécificités techniques qui viennent complexifier la constitution de preuves pertinentes dans le cadre d'une action en indemnisation pour un dommage causé par le résultat d'un système d'IA. Ces caractéristiques distinguent les systèmes d'IA des autres produits existants (**§1**), créant alors pour le demandeur un besoin de preuves concernant la décision individuelle dommageable et le fonctionnement global du logiciel (**§2**).

§1. Les particularités des systèmes d'IA dans la réalisation du dommage

La particularité des systèmes d'IA en tant que produit. La première particularité d'un système d'IA est sa capacité à produire des résultats. Le terme de « résultat » désigne la sortie de l'algorithme tandis que celui de « décision » renvoie à la décision produite, à la suite de l'ensemble du processus décisionnel, qui affecte l'individu. Si le processus décisionnel est

réalisé exclusivement par un système d'IA, alors le résultat et la décision sont le même élément. Toutefois, si le processus décisionnel n'est que partiellement réalisé par un système d'IA, alors la décision n'est pas nécessairement identique au résultat. Dans ce contexte, le résultat du système d'IA correspond la décision finale seulement s'il se trouve à la fin du processus décisionnel. Si le résultat est modifié par un intervenant humain, alors le résultat contribue à la décision mais ne constitue pas la décision finale.

La capacité des systèmes d'IA à produire un résultat est une particularité très importante dans la réalisation d'un dommage. En effet, la cause directe du préjudice subi par le demandeur ne sera jamais le système d'IA lui-même mais le ou les résultats qu'il a produits. Par exemple, un système d'IA de recrutement ne cause pas de dommage en tant que tel. Ce sont les résultats qu'il produit, en particulier les refus, qui peuvent causer un préjudice aux candidats. Il existe donc deux éléments importants pour le demandeur lors de sa recherche d'éléments de preuve : les informations concernant le système d'IA et celles relatives à sa décision dommageable. Cette dichotomie différencie considérablement les systèmes d'IA des autres produits traditionnels pour lesquels le dommage est directement causé par le produit (jouets, machines, aviation, etc.).

La particularité de certains systèmes d'IA en tant que logiciel. La capacité des systèmes d'IA à produire des résultats en fait des produits particuliers mais pas uniques. Les systèmes informatiques dits « décisionnels », à savoir les logiciels de prise de décision ou d'aide à la prise de décision, possèdent aussi cette caractéristique. Au regard de la définition définitive retenue par le législateur européen, la notion de système d'IA semble désigner un grand nombre de logiciels. Toutefois, certains de ces systèmes d'IA présentent un fonctionnement particulièrement opaque et imprévisible, en particulier les logiciels utilisant des technologies d'apprentissage automatique. Cette opacité et imprévisibilité de fonctionnement permet de distinguer ces systèmes d'IA d'autres systèmes plus traditionnels mais pose également des difficultés supplémentaires pour comprendre le fonctionnement du logiciel, notamment lors de la survenance d'un dommage.

Dans un logiciel classique, la transformation des données d'entrée en résultat par le système est très souvent compréhensible, à défaut d'être toujours claire. Par conséquent, chaque résultat produit par le logiciel peut être expliqué au regard du fonctionnement global du système. En revanche, certains systèmes d'IA étant opaques et imprévisibles, leur fonctionnement n'est pas toujours connaissable et très rarement compréhensible. De ce fait, la logique de fonctionnement

du système d'IA ne peut pas être utilisée pour anticiper et comprendre le résultat produit. Deux résultats similaires peuvent être justifiés par une combinaison différente de critères. De la même manière, une légère variation sur les données d'entrées peut avoir des conséquences très importantes sur le résultat produit¹²¹⁶. Par conséquent, il est d'autant plus crucial de distinguer entre les informations relatives au fonctionnement du système d'IA litigieux et celles concernant le résultat et/ou la décision dommageable rendu.

L'évolution des éléments de preuves disponibles. Là où un logiciel informatique traditionnel ne permet qu'une automatisation de l'exécution du processus de prise de décision, l'autonomie décisionnelle et l'opacité des systèmes d'IA entraînent une véritable délégation d'une partie du raisonnement décisionnel¹²¹⁷. À l'exception des situations où le système d'IA est utilisé comme un outil purement supplémentaire par rapport aux processus décisionnel existant, le recours à cet outil entraîne nécessairement une réduction de la place des individus dans le processus. Par conséquent, l'exécution du processus décisionnel génère moins d'éléments de preuves directement issus d'un être humain. En d'autres termes, l'usage des systèmes d'IA comme outil de prise de décision conduit à une réduction de la proportion des sources humaines de preuves qui existaient jusqu'alors. Par exemple, si une partie du processus de recrutement est réalisée par un système d'IA, les salariés des ressources humaines ne pourront pas témoigner pour expliquer les raisons derrière le refus, l'acceptation ou l'évaluation des candidats durant la phase automatisée. Ensuite, le recours aux systèmes d'IA entraîne une évolution de la nature des éléments de preuves disponibles dans le cadre de l'action en réparation. Si les preuves d'origine humaine se réduisent, elles sont remplacées par des preuves d'origine technique. Ainsi, là où un témoignage permet de comprendre le raisonnement d'un recruteur, c'est la documentation ou encore les conclusions d'un audit informatique qui rempliront cette fonction lorsque la tâche a été déléguée à un système d'IA. Pour les parties au litige, l'évolution de la nature des éléments de preuve disponibles rend aussi plus difficile leur compréhension du fait de leur technicité. Les individus avec peu ou pas de compétences informatiques font alors face à une couche supplémentaire d'opacité¹²¹⁸.

Cette évolution des types d'éléments de preuves disponibles est d'autant plus importante lorsque la prise de décision est entièrement réalisée par un système d'IA. L'évolution des

¹²¹⁶ Le critère de robustesse discuté précédemment vise notamment à éviter ce type de situations sans pour autant pouvoir l'empêcher complètement.

¹²¹⁷ SARANGI Saswat et SHARMA Pankaj, *Artificial Intelligence*, Routledge India, 2018.

¹²¹⁸ BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016.

éléments de preuves disponibles met donc en avant la place des éléments techniques de preuves lors d'une action en réparation du fait de l'IA.

§2. Le besoin d'informations relatives au résultat individuel dommageable

L'existence d'un lien de causalité double. Au-delà d'un état de fait, la dichotomie entre le système d'IA et le résultat dommageable a des effets directs sur les exigences probatoires en matière de responsabilité lors d'un dommage causé par l'IA. L'existence de ces deux éléments implique en effet l'existence d'un lien de causalité plus étendu entre le fait générateur et le dommage subi. Premièrement, il existe un lien de causalité entre le fait générateur de responsabilité, en l'espèce lié au fonctionnement du système d'IA, et le résultat produit par le logiciel. Secondement, il est nécessaire de démontrer le lien de causalité entre le résultat produit par le système d'IA et la décision préjudiciable pour le demandeur. L'existence de ce double lien de causalité entre le fait générateur et le préjudice est par ailleurs reconnue par le législateur européen. En effet, le texte propose une présomption de lien de causalité lorsque trois conditions sont satisfaites :

« « [...] le lien de causalité entre la faute du défendeur et le résultat produit par le système d'IA ou l'incapacité de celui-ci à produire un résultat, lorsque toutes les conditions suivantes sont remplies :

- a) [la faute du défendeur] ;*
- b) Il peut être considéré comme raisonnablement probable, compte tenu des circonstances de l'espèce, que la faute a influencé le résultat du système d'IA ou l'incapacité de celui-ci à produire un résultat ;*
- c) Le demandeur a démontré que le résultat du système d'IA ou l'incapacité de celui-ci à produire un résultat est à l'origine du dommage. »¹²¹⁹*

¹²¹⁹ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 4.1.

L'existence de ce lien de causalité double peut être illustré par l'exemple suivant. Prenons un processus décisionnel partiellement automatisé destiné à évaluer la solvabilité d'un demandeur de prêt. Pour les besoins de l'illustration, le choix est fait de simplifier le processus décisionnel en considérant que la décision finale produite est binaire : soit les demandeurs de prêt reçoivent une décision positive, soit ils subissent un refus. Le score de solvabilité est une variable continue qui s'étend de 0 à 100. Si le score est strictement inférieur à 70, alors le prêt est refusé. Si le score est supérieur ou égal à 70, alors la demande de prêt est acceptée. Dans cet exemple, le résultat du système d'IA est le score de solvabilité, la décision est l'octroi ou non du prêt. Le système d'IA peut être dysfonctionnel sans pour autant que le résultat local (le score de l'individu) soit affecté. C'est le cas lorsque le système d'IA surestime grandement le score de solvabilité en raison d'un critère (discriminant ou non) mais que le demandeur ne possède pas ledit critère responsable de la défaillance. Et même si le score de récidive est affecté par le mauvais fonctionnement du système d'IA (qu'il soit dû à son utilisation ou à sa conception), le demandeur doit également apporter la preuve que le résultat erroné est déterminant dans la décision dommageable. Le dommage, quant à lui, est l'impossibilité d'obtenir un prêt. La distinction entre le système d'IA, son résultat et la décision apparaît alors clairement. Dans un premier temps, le demandeur a donc besoin d'accéder aux informations relatives au fonctionnement global du système d'IA afin de prouver le fait générateur. Dans un second temps, il doit également avoir accès aux informations relatives au résultat local afin de constater, d'une part, le lien de causalité entre le dysfonctionnement du système d'IA et le résultat généré et, d'autre part, le lien de causalité entre ce résultat et la décision finale préjudiciable.

Sous-section 2 : Les nouveaux éléments de preuve issus du règlement sur l'IA

Annonce du plan. Face à ces problématiques probatoires spécifiques aux systèmes d'IA, les parties à une action en réparation du fait d'un système d'IA ont besoin d'accéder à des informations relatives à la fois au fonctionnement global du système et au résultat litigieux. Par le biais du règlement européen sur l'IA, le législateur européen impose aux fournisseurs et aux déployeurs la mise en place de mécanismes de constitution et d'accès à de telles informations. Cependant, l'apparition de nouveaux éléments de preuve potentiels ne garantit pas qu'ils seront

accessibles au demandeur (ou demandeur potentiel). Outre la description de ces informations, il est donc nécessaire d'analyser également leur accessibilité pour chacune des parties. Ces éléments de preuves doivent ainsi être distingués selon qu'ils apportent des renseignements sur le fonctionnement global du système d'IA (§1) ou sur le résultat individuel préjudiciable pour le demandeur (§2).

§1. Les éléments de preuves relatifs au fonctionnement global du système d'IA

Annonce de plan. Les informations relatives au fonctionnement global du système d'IA sont une source de preuves particulièrement importante aux yeux du législateur qui met en place différents mécanismes juridiques pour accroître la quantité d'informations disponibles. En matière de responsabilité du fait d'un résultat du système d'IA, ce type d'information est particulièrement intéressant car il permet d'établir, ou du moins de contribuer à établir, l'existence du fait générateur de la responsabilité. Toutefois, la quantité accrue d'informations que les acteurs techniques, du fournisseur à l'utilisateur, doivent générer dans le cadre de leurs activités n'implique pas leur disponibilité pour le demandeur. Ainsi si le demandeur dispose désormais d'une plus grande quantité d'informations **(A)**, les éléments de preuve les plus pertinents dans le cadre d'une action en réparation lui demeurent inaccessibles sans la coopération du défendeur ou des tiers impliquées **(B)**.

A- Les informations indirectes sur le fonctionnement d'un système d'IA

Les attestations de conformité des systèmes d'IA à haut risque. La logique de certification souhaitée dans le règlement européen sur l'IA conduit à l'élaboration de mécanismes permettant aux fournisseurs de systèmes d'IA à haut risque de montrer publiquement que leur produit est conforme aux exigences juridiques. À l'image des autres produits encadrés sur le marché européen¹²²⁰, le fournisseur d'un système d'IA à haut risque doit apposer un marquage CE « de

¹²²⁰ Ainsi les ascenseurs, les explosifs à usage civil, les articles pyrotechniques sont des produits devant faire l'objet d'un marquage CE pour attester de leur conformité aux réglementations européennes. Pour la liste complète des produits soumis au marquage CE dans le marché européen, voir : « Le marquage « CE » | entreprises.gouv.fr ».

façon visible, lisible et indélébile »¹²²¹. Toutefois, le marquage CE ne signifie pas que le produit respecte effectivement les réglementations d’harmonisation associées mais seulement que son fabricant « *assume la responsabilité de la conformité du produit avec toutes les exigences applicables [...]* »¹²²². En d’autres termes, le fournisseur du système d’IA à haut risque indique qu’il a respecté les obligations qui lui incombent au regard des procédures de mise en conformité prévues dans le règlement sur l’IA. La présence du marquage CE a pour objectif, d’une part, de faciliter la circulation intra-européenne des produits concernés et, d’autre part, de créer la confiance entre les consommateurs, les États membres et le fabricant au sujet de la qualité du produit en question¹²²³. Du point de vue de la responsabilité du fait d’un résultat d’IA, l’apposition du marquage sur un système d’IA à haut risque par le fournisseur est un élément de preuve de l’absence de faute de sa part¹²²⁴ et participe à écarter partiellement l’hypothèse d’un défaut du système d’IA¹²²⁵. Toutefois, il s’agit d’un élément de preuve documentaire, dont la crédibilité dépend de la bonne foi de celui qui l’appose ou de la connaissance effective du fonctionnement du logiciel.

Au-delà du risque d’apposition volontairement illicite du marquage¹²²⁶, le système d’IA à haut risque peut avoir évolué (réapprentissage, modification manuelle, etc.) vers une version qui n’est plus, soit totalement soit partiellement, conforme au règlement sur l’IA. De plus, le marquage CE garantit la conformité du système d’IA uniquement aux exigences du règlement sur l’IA et non aux autres obligations pertinentes. Il est important de se rappeler que l’évaluation de la proportionnalité de l’atteinte aux droits fondamentaux sera réalisée le plus souvent par le fournisseur lui-même¹²²⁷, sur la base d’instructions produites par des acteurs privés non formés à ce genre d’enjeux. La présence du marquage CE ne garantit pas totalement que le fonctionnement du système d’IA soit effectivement conforme à la réglementation. Par conséquent, s’il peut être considéré comme un élément de preuve utilisable par le défendeur

¹²²¹ « Règlement sur l’intelligence artificielle », 2024. Article 48.3.

¹²²² « Règlement (CE) n° 765/2008 du Parlement européen et du Conseil du 9 juillet 2008 fixant les prescriptions relatives à l’accréditation et à la surveillance du marché pour la commercialisation des produits et abrogeant le règlement (CEE) n° 339/93 du Conseil (Texte présentant de l’intérêt pour l’EEE) », OJ L, 2008. Article 30.3.

¹²²³ *Ibid.*

¹²²⁴ Pour rappel, la version actuelle de la proposition de directive sur la responsabilité en matière d’IA définit limitativement la faute du fournisseur du système d’IA à haut risque comme le manquement à certaines obligations de conformité définies dans le règlement européen sur l’IA. Pour plus de détails sur le sujet, voir §1. La faute du fournisseur.

¹²²⁵ Contrairement à la faute du fournisseur du système d’IA à haut risque, la défectuosité d’un produit ne se limite pas à sa conformité avec les exigences juridiques en vigueur.

¹²²⁶ Cette situation ouvre par ailleurs un droit à la réparation de la part de l’acheteur :

¹²²⁷ Certains dépoyeurs de systèmes d’IA à haut risque doivent également réaliser une analyse d’impact sur les droits fondamentaux avant la mise en service du logiciel. « Règlement sur l’intelligence artificielle », 2024. Article 27.

dans une action en réparation, il serait malavisé de le considérer par principe comme une preuve convaincante de l'absence de fait générateur pour le défendeur. En revanche, l'absence de marquage CE constitue un des éléments de preuve pertinents pour le demandeur. Dans le cas de l'apposition trompeuse d'un marquage CE sur le système d'IA à haut risque litigieux, le demandeur qui souhaiterait engager la responsabilité du fournisseur doit donc passer outre le marquage pour prouver le défaut du produit ou la faute du fournisseur. Bien qu'accessible au demandeur, la présence d'un marquage CE ne constitue pas un élément de preuve pertinent pour identifier la présence, ou l'absence, d'un fait générateur de responsabilité du fournisseur.

Les normes techniques de conformité du système d'IA litigieux. Pour leur part, les normes techniques relatives aux systèmes d'IA permettront de préciser les caractéristiques de fonctionnement des logiciels. Les normes européennes harmonisées seront particulièrement visées ici car la présomption de conformité au règlement européen sur l'IA qu'entraîne leur respect en fait un outil très utilisé par les opérateurs économiques. Ainsi dans la grande majorité des situations, le demandeur qui souhaite prouver la faute devra avant tout apporter des éléments de preuves pertinents laissant penser que le système d'IA à haut risque litigieux n'est pas conforme aux normes européennes harmonisées comme il prétend l'être. En d'autres termes, l'accès aux normes européennes harmonisées est essentiel pour parvenir à renverser la présomption de conformité au règlement sur l'IA, et donc contribuer à caractériser la faute du défendeur. De plus, l'accès aux normes européennes harmonisées et aux méthodologies d'évaluation des risques liés aux biais discriminatoires permet d'apporter des éléments de preuves de l'incohérence potentielle des indicateurs et seuils statistiques exigés au regard du droit à la non-discrimination. Le standard que le système d'IA litigieux est censé respecter est donc sans aucun doute un élément de preuve pertinent pour dans le cadre d'une action en réparation.

Les normes techniques, qu'elles soient ou non publiées dans le journal officiel de l'UE, sont accessibles à tous. La particularité de ces éléments de preuve était que leur accès n'était pas gratuit pour le demandeur. Cependant, cette spécificité a récemment changé depuis une décision du 5 mars 2024 rendu par la Cour de justice de l'Union européenne¹²²⁸. Dans cet arrêt, la juridiction européenne déclare qu'en tant que partie intégrante du droit de l'Union, les normes européennes harmonisées doivent respecter l'exigence d'accès libre au droit prévue par l'article

¹²²⁸ Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P. Par. 81.

2 du TUE¹²²⁹. Plus précisément, la Cour caractérise l'accès aux normes européennes harmonisées comme un « *intérêt public supérieur* »¹²³⁰ justifiant la divulgation des normes harmonisées en question¹²³¹. Cette ouverture gratuite¹²³² de ces normes techniques garantit un accès à ces éléments de preuves pour le demandeur. Selon le niveau de détails en termes de preuves, l'accès à une norme européenne harmonisée permet au demandeur d'obtenir des informations plus ou moins précises sur les exigences minimales de fonctionnement que le système d'IA litigieux devrait présenter. Néanmoins, si l'accès à ces preuves est indispensable pour l'individu lésé, il n'est en rien suffisant pour démontrer la présence d'une faute ou d'un défaut du produit. Premièrement, le fournisseur du système d'IA à haut risque peut avoir choisi une autre norme, issue d'un autre organisme de normalisation ou élaborée par lui-même, pour mettre en conformité son logiciel. Deuxièmement, même si le fournisseur garantit sa conformité par son respect d'une norme européenne harmonisée, il s'agit seulement d'un référentiel de fonctionnement. La norme technique ne fournit aucune information sur le fonctionnement effectif du système d'IA litigieux.

B- La documentation technique inaccessible au demandeur

L'équilibre entre l'accès aux preuves et la confidentialité des informations. L'amélioration de la redevabilité des acteurs économiques du numérique est un enjeu central de l'encadrement des nouvelles technologies, et les technologies d'IA ne dérogent pas à cette volonté de l'Union

¹²²⁹ « L'Union est fondée sur les valeurs de respect de la dignité humaine, de liberté, de démocratie, d'égalité, de l'État de droit, ainsi que de respect des droits de l'homme, y compris des droits des personnes appartenant à des minorités. Ces valeurs sont communes aux États membres dans une société caractérisée par le pluralisme, la non-discrimination, la tolérance, la justice, la solidarité et l'égalité entre les femmes et les hommes. » - « Traité sur l'Union européenne (version consolidée) », OJ C, 2012. Article 2. Sur le lien entre l'État de droit et le libre accès au droit, voir notamment : Cour de justice, 22 février 2022, n° C-160/20. Par. 41.

¹²³⁰ « Les institutions refusent l'accès à un document dans le cas où sa divulgation porterait atteinte à la protection : - des intérêts commerciaux d'une personne physique ou morale déterminée, y compris en ce qui concerne la propriété intellectuelle, - des procédures juridictionnelles et des avis juridiques, - des objectifs des activités d'inspection, d'enquête et d'audit, à moins qu'un intérêt public supérieur ne justifie la divulgation du document visé. » - « Règlement (CE) n° 1049/2001 du Parlement européen et du Conseil du 30 mai 2001 relatif à l'accès du public aux documents du Parlement européen, du Conseil et de la Commission », OJ L, 2001. Article 4, paragraphe 2, dernier membre de phrase. Emphase ajoutée par l'auteur.

¹²³¹ Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P. Par. 85.

¹²³² « L'accès aux documents s'exerce soit par consultation sur place, soit par délivrance d'une copie, y compris, le cas échéant, une copie électronique, selon la préférence du demandeur. Le coût de la réalisation et de l'envoi des copies peut être mis à la charge du demandeur. Il ne peut excéder le coût réel de la réalisation et de l'envoi des copies. La gratuité est de règle en cas de consultation sur place ou lorsque le nombre de copies n'excède pas 20 pages A4, ainsi qu'en cas d'accès direct sous forme électronique ou par le registre. » - « Règlement (CE) n° 1049/2001 du Parlement européen et du Conseil du 30 mai 2001 relatif à l'accès du public aux documents du Parlement européen, du Conseil et de la Commission », OJ L, 2001. Article 10.1. Emphase ajoutée par l'auteur.

européenne¹²³³. Cette poursuite de l'objectif de redevabilité s'est accompagnée d'une demande de création de nouveaux éléments documentaires envers les opérateurs économiques concernés. La confidentialité des informations visées par ces obligations de redevabilité a entraîné une mise en balance immédiate avec les intérêts économiques des acteurs au regard des secrets protégés par la loi, en particulier le secret des affaires¹²³⁴. La transparence qu'implique le besoin de redevabilité expose à des risques de fuite de données, une potentielle perte de compétitivité pour les acteurs ou encore un risque de tromperie du raisonnement du système par les personnes affectées¹²³⁵. Dès lors, une grande partie des nouveaux éléments de preuves produits par le fournisseur ou l'utilisateur du système d'IA voient leur accès grandement limité, matérialisant alors le premier degré d'opacité des technologies d'IA¹²³⁶. La limitation de la transparence s'opère à la fois au travers des personnes en capacité d'accéder à ces informations et des modalités de cet accès.

L'accès à la documentation technique limité aux acteurs du contrôle de conformité. Le règlement européen sur l'IA impose aux fournisseurs de systèmes d'IA à haut risque la production d'une documentation technique contenant une description détaillée du système d'IA et de son processus de développement. Le contenu de cette documentation technique se révèle très prometteur en tant qu'élément de preuve, que l'action en réparation se fonde sur les régimes traditionnels de responsabilité du fait d'un produit ou sur une atteinte aux droits fondamentaux. En effet, les informations détaillées contenues dans la documentation technique (procédures d'évaluation et de tests du système d'IA, informations sur le niveau de contrôle humain nécessaire, justification des choix techniques dont les compromis des métriques, etc.¹²³⁷), associés aux normes techniques et au règlement européen, peuvent permettre de caractériser la présence d'une faute du fournisseur ou d'un défaut de sécurité dans le produit. Néanmoins, le contenu de la documentation technique n'est accessible qu'à une liste restreinte d'acteurs dont

¹²³³ La responsabilisation est une des sept exigences de la commission européenne pour renforcer la confiance dans l'IA : *Communication de la Commission européenne au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions - Renforcer la confiance dans l'intelligence artificielle axée sur le facteur humain*, 2019. Cette exigence est ainsi reprise dans le livre blanc de l'Union européenne publié l'année suivante en 2020 : COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020.

¹²³⁴ KROLL Joshua A., HUEY Joanna, BAROCAS Solon *et al.*, « Accountable Algorithms », *University of Pennsylvania Law Review*, 165, 2017.

¹²³⁵ DE LAAT Paul B., « Algorithmic Decision-Making Based on Machine Learning from Big Data », *Philos. Technol.*, 31, 2018.

¹²³⁶ BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016. Les autres degrés d'opacité sont la complexité technique et l'opacité intrinsèque propre aux technologies d'IA.

¹²³⁷ « Règlement sur l'intelligence artificielle », 2024. Annexe IV.2.

les autorités de surveillance du marché ainsi que les autorités ou organismes publics chargés de la protection des droits fondamentaux :

« les fournisseurs accordent aux autorités de surveillance du marché un accès complet à la documentation ainsi qu'aux jeux de données d'entraînement, de validation et de test utilisés pour le développement des systèmes d'IA à haut risque [...] »¹²³⁸

« Les autorités ou organismes publics nationaux qui supervisent ou font respecter les obligations au titre du droit de l'Union visant à protéger les droits fondamentaux, y compris le droit à la non-discrimination, en ce qui concerne l'utilisation des systèmes d'IA à haut risque visés à l'annexe III sont habilités à demander toute documentation créée ou conservée en vertu du présent règlement [...] »¹²³⁹

Lorsque le fournisseur d'un système d'IA à haut risque est une autorité répressive ou une autorité compétente en matière d'asile, la documentation technique n'est même pas communiquée par principe aux autorités de surveillance du marché. En effet, la documentation technique doit demeurer dans les locaux des fournisseurs. Si les autorités de surveillance du marché souhaitent accéder à la documentation technique, elles doivent en faire la demande et seuls les membres du personnel habilités peuvent consulter les informations qui y sont contenues¹²⁴⁰.

La documentation technique, au sens du règlement européen sur l'IA, contient des informations particulièrement précises sur le fonctionnement global du système d'IA à haut risque concerné. Il s'agit donc d'une preuve extrêmement importante dans le cadre d'une action en réparation pour un dommage causé par un système d'IA à haut risque. Cependant, le niveau de détail de la documentation technique en fait aussi une source d'informations sensibles pour le fournisseur du logiciel. Par conséquent, la documentation technique est une nouvelle preuve pertinente pour le demandeur lésé auquel il ne peut accéder sans l'intervention du juge.

L'accès à ces éléments de preuves limité au juge. Le recours au juge pour accéder aux éléments de preuve pertinents dans le litige est un mécanisme traditionnel constitutif du droit à la preuve. Toutefois, la protection du secret des affaires doit être prise en compte par le juge. Les directives proposées par la Commission européenne, relatives à la responsabilité du fait des

¹²³⁸ *Ibid.* Article 74.12

¹²³⁹ *Ibid.* Article 77.1

¹²⁴⁰ *Ibid.* Article 70.2

produits défectueux¹²⁴¹ et en matière de responsabilité civile extracontractuelle¹²⁴², énoncent toutes deux l'importance du respect du secret des affaires et autres informations confidentielles par les juridictions¹²⁴³. La conciliation entre la demande de production forcée du demandeur pour obtenir ces nouveaux éléments de preuve et l'invocation de l'empêchement légitime du défendeur ou du tiers concerné, dont la violation entacherait la preuve d'illicéité, est réalisée par le juge lui-même¹²⁴⁴. En d'autres termes, ce n'est pas au demandeur mais au juge de vérifier la nécessité et la proportionnalité de la mesure de production forcée de preuve (ou la mesure d'instruction) et les intérêts invoqués par le destinataire de la mesure. Le secret des affaires, comme tous les secrets protégés par la loi, est donc un obstacle important à l'accès par le demandeur aux nouveaux éléments de preuves relatifs au fonctionnement du système d'IA litigieux.

§2. Les éléments de preuves relatifs au résultat individuel dommageable

Annnonce de plan. La preuve de l'existence d'un lien de causalité entre le fait générateur et le préjudice réparable impose d'accéder à des informations relatives au résultat du système d'IA à l'origine du litige. Ce besoin de renseignements spécifiques à la situation du demandeur est une particularité des systèmes d'IA qui produisent des décisions opaques s'adaptant à chaque situation. Ainsi, le règlement européen sur l'IA prévoit plusieurs mécanismes d'information représentant des éléments de preuve prometteurs pour le demandeur, à savoir les journaux d'enregistrement **(A)** et la possibilité d'une explication individuelle **(B)**.

¹²⁴¹ COMMISSION EUROPÉENNE, *Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux*, 28 septembre 2022.

¹²⁴² *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022.

¹²⁴³ *Ibid.*

¹²⁴⁴ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 299.

A- Les journaux d'enregistrement des systèmes d'IA

Définition et objectifs du journal d'enregistrement. Le règlement européen sur l'IA impose une obligation pour le fournisseur d'un système d'IA à haut risque d'intégrer des « *fonctionnalités permettant l'enregistrement automatique des événements (« journaux ») pendant le fonctionnement de ces systèmes* »¹²⁴⁵. Généralement appelés par leur nom anglais « log », ce les journaux désignent un enregistrement numérique de l'activité d'un ordinateur¹²⁴⁶. En d'autres termes, il s'agit d'un journal de bord horodaté qui liste l'ensemble des différents événements qui se sont produits sur un ordinateur, un serveur, etc. Ils sont généralement utilisés pour comprendre les raisons d'une panne informatique ou pour identifier des menaces en termes de cybersécurité¹²⁴⁷.

Dans le contexte du règlement sur l'IA, l'objectif de ces journaux est de garantir la traçabilité du fonctionnement du système d'IA. La traçabilité est une composante centrale du principe de redevabilité qui est présente au sein des réflexions de l'Union européennes depuis les premiers travaux sur l'IA¹²⁴⁸. Dans la version initiale du règlement proposé par la Commission européenne, la tenue des logs avait trois objectifs principaux. Dans un premier temps, les logs doivent permettre de surveiller et d'identifier l'apparition d'un risque, du fait de l'utilisation du système d'IA, pour la sécurité, la santé ou la protection des droits fondamentaux des individus¹²⁴⁹. Dans un deuxième temps, la Commission européenne a pensé l'élaboration de ce journal d'enregistrement comme un moyen de s'assurer que le système d'IA n'a pas évolué de façon substantielle en réaction à une mise en situation particulière. Dans ce cas, l'objectif est de vérifier si le système d'IA à haut risque a besoin de faire l'objet d'une nouvelle évaluation de conformité, par le fournisseur du système d'IA lui-même ou par un organisme notifié¹²⁵⁰. Enfin, la dernière finalité attribuée à ces logs était la « *facilitation de la surveillance après*

¹²⁴⁵ COMMISSION EUROPÉENNE, *Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux*, 28 septembre 2022. Article 8.3. *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022. Article 3.4.

¹²⁴⁶ Traduit de l'anglais par l'auteur : « *a written Or, digital record of activity on a computer Or, phone line* » - « Log_1 noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com ».

¹²⁴⁷ « Log », publié le 9 janvier 2019.

¹²⁴⁸ Voir notamment : COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020 SERVICE DE RECHERCHE DU PARLEMENT EUROPÉEN, « A governance framework for algorithmic accountability and transparency - Think Tank », Parlement Européen, 2019.

¹²⁴⁹ « Règlement sur l'intelligence artificielle », 2024. Article 65.1.

¹²⁵⁰ *Ibid.* Article 43.4.

commercialisation »¹²⁵¹, à savoir la surveillance du fonctionnement de systèmes d'IA afin de s'assurer que celui-ci respecte toujours les exigences de conception imposées par le règlement européen tout au long de son cycle de vie.

A la suite des discussions et des amendements par le Parlement européen, le journal d'enregistrement s'est vu attribué d'autres raisons d'être, notamment au regard des enjeux énergétiques. Dans sa version du 9 mai 2023, la proposition de règlement européen sur l'IA fait de ces fonctionnalités un outil de traçabilité du coût écologique¹²⁵². Cependant, cet aspect n'a pas été conservé dans la version finale du règlement européen qui conserve les trois finalités proposées initialement par la Commission européenne.

Contenu des journaux d'enregistrement d'un système d'IA. À ce jour, il n'est pas possible de décrire avec précision le contenu des journaux d'enregistrement associés avec les systèmes d'IA à haut risque, et ce pour plusieurs raisons. Premièrement, la structure et les informations globales devant être contenues ne sont décrites ni par le règlement sur l'IA ni par ses annexes. Comme de nombreux points du règlement européen, la détermination du contenu et des autres exigences relatives aux journaux d'enregistrement sera réalisée par les organismes de normalisation et communiquée aux acteurs sous la forme de normes techniques¹²⁵³.

Deuxièmement, le contenu des journaux d'enregistrement varie en fonction des finalités poursuivies. Or, ces besoins dépendent eux-mêmes en partie de l'utilisation du système d'IA. Par exemple, si le journal d'enregistrement est associé à un serveur dans le cadre d'une application en ligne, il enregistrera des informations relatives à l'activité sur les sites web en question. Dans le cas où le fichier est lié à une application locale installée sur un ordinateur, les informations enregistrées seront différentes. Par conséquent, le contenu des *logs* est contextuel et il n'aurait pas été pertinent pour le législateur européen d'essayer de lister exhaustivement les éléments qui devraient apparaître.

Le règlement sur l'IA énonce toutefois quelques éléments sur le contenu potentiel de ces journaux d'enregistrement. Le législateur européen fournit ainsi une liste non limitative des informations obligatoires que doivent contenir les journaux d'enregistrement des systèmes d'IA

¹²⁵¹ *Ibid.* Article 12.3.

¹²⁵² « *High-risk AI systems shall be designed and developed with the logging capabilities enabling the recording of energy consumption, the measurement Or, calculation of resource use and environmental impact of the high-risk AI system during all phases of the system's lifecycle* » - *Ibid.* Version du 9 mai 2023 proposée par le Parlement européen.

¹²⁵³ COMMISSION EUROPÉENNE, « Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence », 2022. Annexe I.

utilisés à des fins d'identification biométrique à haut risque¹²⁵⁴. Ces informations sont les suivantes :

« (a) l'enregistrement de la période de chaque utilisation du système (date et heure de début et de fin pour chaque utilisation) ;

(b) la base de données de référence utilisée par le système pour vérifier les données d'entrée ;

(c) les données d'entrée pour lesquelles la recherche a abouti à une correspondance ;

(d) l'identification des personnes physiques participant à la vérification des résultats, visées à l'article 14, paragraphe 5.

Cette disposition illustre bien le critère contextuel des journaux d'enregistrement en mettant en avant des éléments importants spécifiques pour le cadre d'utilisation que sont les systèmes d'IA d'identification biométrique. L'obligation selon laquelle l'identification réalisée par le système d'IA soit vérifiée par deux êtres humains est en effet une particularité de cette finalité qui ne se retrouve pas pour les autres systèmes d'IA à haut risques¹²⁵⁵.

La disposition met également en avant un point essentiel dans l'analyse des *logs* comme éléments de preuves en cas de dommage du fait d'un système d'IA : celui de la variété des informations pouvant être contenues. Le journal d'enregistrement contient au minimum l'évènement exécuté, la date et l'heure de l'évènement et enfin l'information sur l'utilisateur du système (adresse IP, adresse locale de l'ordinateur, etc.)¹²⁵⁶. Mais les journaux d'enregistrement peuvent également contenir les données utilisées en entrées. De plus, les informations qu'ils contiennent peuvent être analysées globalement afin de retracer un historique des évènements associés à une prise de décision pour un individu, il est alors question d'historique de *logs*. Ces informations peuvent représenter des éléments de preuves intéressants en cas de dommage causé par un résultat de systèmes d'IA.

¹²⁵⁴ « Règlement sur l'intelligence artificielle », 2024. Annexe III.1.a. En raison des divergences entre les différentes versions du texte sur les systèmes d'IA biométriques, l'expression est laissée volontairement large en l'espèce.

¹²⁵⁵ *Ibid.* Article 14.

¹²⁵⁶ « Fichier LOG : définition, ouverture et utilisation | Adobe », en ligne : <<https://www.adobe.com/fr/acrobat/resources/document-files/text-files/log-file.html>>.

L'intérêt en tant qu'éléments de preuve en cas de dommage du fait de l'IA. Tout d'abord, le contenu d'un journal d'enregistrement sera extrêmement pratique pour attester de la survenance d'une erreur de fonctionnement du système d'IA. Ainsi les *logs* associés au résultat dommageable pourront être un élément attestant la présence ou l'absence de « *dysfonctionnement manifeste* » du système d'IA, au sens de la proposition de directive sur la responsabilité du fait des produits défectueux¹²⁵⁷, permettant de présumer le défaut du produit. De la même manière, si les données d'entrée sont conservées dans les journaux d'enregistrement, cela facilitera le travail d'évaluation de leur pertinence mais aussi de la personne qui contrôle lesdites données¹²⁵⁸. Dans la situation où le système d'IA à haut risque a été non-conforme temporairement au règlement européen sur l'IA (mauvaise mise à jour du logiciel par exemple), l'enregistrement de la date et de l'heure de l'utilisation du système d'IA sont des éléments, à la disposition de l'utilisateur ou du fournisseur, pouvant caractériser l'existence d'une simultanéité entre le fait générateur et le dommage. Ces différents exemples démontrent que les journaux d'enregistrement peuvent représenter un outil de preuve très prometteur pour le demandeur, le défendeur et même le juge dans les litiges relatifs aux dommages du fait d'un système d'IA. De ce point de vue, il est clair qu'ils remplissent leur fonction d'aide à la traçabilité des décisions algorithmiques et de facilitation du règlement des litiges en matière de responsabilité.

Limites des journaux d'enregistrement en tant qu'outil de preuve de responsabilité. Malgré leur intérêt probatoire significatif en cas de dommage du fait d'un système d'IA, les journaux d'enregistrement présentent certaines limites. La première limite se rapporte à la conservation des informations dans les journaux d'enregistrement. Les informations qui ne seront pas considérées comme essentielles à conserver par les normes techniques, qu'ils s'agissent des normes européennes harmonisées ou d'autres standards privés, ne seront pas enregistrées par l'utilisateur et/ou le fournisseur. Ce constat semble relever de l'évidence mais il signifie que ces informations qui n'auront pas été jugées comme pertinentes seront bien plus difficiles, voire impossibles, à retrouver *a posteriori* au moment du litige. En effet, pour des raisons de place de stockage mais aussi de réglementation relative à la protection des données

¹²⁵⁷ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux*, 28 septembre 2022 Article 9.2.c)

¹²⁵⁸ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022.

personnelles, seules les données jugées nécessaires seront stockées¹²⁵⁹. En tant que potentiel élément de preuve, les informations des journaux d'enregistrement doivent être conservées pendant une longue durée. Au regard de la vitesse de production de résultat par un système d'IA, la quantité d'informations à conserver peut vite devenir une problématique technique et environnementale de taille. Par conséquent, même si les journaux d'enregistrement peuvent contenir des informations pertinentes en cas de résultat dommageable, il ne faut pas oublier que leur objectif est d'abord d'aider à vérifier la conformité du système d'IA. De ce fait, ce sont avant tout les éléments nécessaires à cette conformité qui seront conservés. Si ces informations peuvent être réutilisées en tant qu'élément de preuve en cas de résultat dommageable, leur pertinence et leur suffisance doivent être questionnées.

La seconde limite importante des journaux d'enregistrement est structurelle. Ce sont des outils de preuve qui portent certes sur le résultat individuel mais ils retranscrivent avant tout le fonctionnement global du système d'IA. En d'autres termes, ce sont des éléments de preuve permettant d'attester du fonctionnement correct du système d'IA mais pas de la qualité du résultat en tant que tel. Par exemple, l'analyse des *logs* peut révéler un dysfonctionnement du logiciel mais pas un résultat erroné. La victime d'une mauvaise décision par un système d'IA fonctionnant correctement ne pourra pas prouver que le résultat produit est biaisé, par exemple, à l'aide des informations contenues dans les journaux d'enregistrement.

B- L'explication de la décision dommageable

Un droit à l'explication de la décision individuelle prise sur le fondement d'un système d'IA à haut risque. Dans la version originelle d'avril 2021 du règlement européen sur l'IA proposée par la Commission européenne, le texte ne comprenait quasi ment aucune disposition offrant des droits aux personnes affectées. À l'exception de son article 52 sur les systèmes d'IA présentant un risque spécifique, les droits des individus étaient un sujet laissé aux réglementations spécifiques à l'exécution des systèmes d'IA (RGPD, régimes de responsabilité, droits fondamentaux, etc.). Le compromis élaboré par le Parlement européen change cette situation en proposant d'accorder plusieurs droits (droit de déposer une plainte auprès d'une

¹²⁵⁹ Le contenu des journaux d'enregistrement constitue généralement une donnée à caractère personnel au sens du règlement européen 2016/679. Ces données sont donc soumises aux principes généraux du RGPD, à savoir le principe de minimisation des données traitées et au principe de conservation proportionnée à la finalité du traitement.

autorité nationale de contrôle, droit à un recours effective contre une autorité de contrôle). Parmi ces nouvelles prérogatives se trouve le droit à l'explication lors d'une prise de décision individuelle. Cette disposition a été conservée dans la version finale du règlement européen :

« Toute personne concernée faisant l'objet d'une décision prise par un déployeur sur la base du résultat d'un système d'IA à haut risque mentionné à l'annexe III, à l'exception des systèmes énumérés au point 2 de ladite annexe, et qui produit des effets juridiques ou affecte significativement cette personne de façon similaire d'une manière qu'elle considère comme ayant des conséquences négatives sur sa santé, sa sécurité ou ses droits fondamentaux a le droit d'obtenir du déployeur des explications claires et pertinentes sur le rôle du système d'IA dans la procédure décisionnelle et sur les principaux éléments de la décision prise. »¹²⁶⁰

Le règlement européen sur l'IA consacre ainsi un « droit à l'explication » pour certaines décisions prises sur le fondement d'un système d'IA à haut risque listé à l'annexe III du texte. Au-delà de la limitation aux systèmes d'IA à haut risque, l'utilisateur du système d'IA peut seulement exercer son droit seulement si la décision *« produit des effets juridiques ou affecte significativement cette personne de façon similaire d'une manière qu'elle considère comme ayant des conséquences négatives sur sa santé, sa sécurité ou ses droits fondamentaux »¹²⁶¹*. La formulation de la condition d'accès au droit à l'explication renvoie clairement à l'une des conditions du droit de la personne concernée de ne pas faire l'objet d'une décision individuelle automatisée énoncée à l'article 22 du règlement (UE) 2016/679 relatif à la protection des données à caractère personnel (RGPD ci-après)¹²⁶². Cependant, cette condition de criticité semble être redondante avec la limitation du droit à l'explication aux seuls systèmes d'IA à haut risque. En effet, les systèmes d'IA à haut risque sont déjà caractérisés par les risques importants qu'ils entraînent pour les individus. Pour rappel, le règlement européen sur l'IA permet au fournisseur d'un système d'IA, listé dans l'annexe III du texte, de déroger à la qualification de système d'IA à haut risque *« s'il ne présente pas de risque important de préjudice pour la santé, la sécurité ou les droits fondamentaux des personnes physiques, y compris en n'ayant pas d'incidence significative sur le résultat de la prise de décision. »¹²⁶³*. Par conséquent, il semble

¹²⁶⁰ « Règlement sur l'intelligence artificielle », 2024. Article 86.1.

¹²⁶¹ *Ibid.*

¹²⁶² « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

¹²⁶³ « Règlement sur l'intelligence artificielle », 2024. Article 6.3.

peu probable qu'un système d'IA à haut risque puisse produire un résultat qui ne permettrait pas à l'individu affecté d'exercer son droit à bénéficier d'une explication.

Le choix d'une explication individuelle. Lorsque l'explication porte sur le fonctionnement du système d'IA, on parle alors d'explication globale tandis que si l'explication porte sur un résultat particulier, il s'agit d'une explication dite « locale »¹²⁶⁴. Jusqu'à présent, les réglementations adressant spécifiquement la question des décisions prises ou assistées par un système algorithmique ont généralement privilégié l'imposition d'une explication globale. Au niveau européen, le RGPD renvoie à la « *logique sous-jacente*¹²⁶⁵ » de la décision automatisée. Le récent règlement européen sur les services numériques impose plusieurs obligations relatives aux systèmes algorithmiques utilisés par les plateformes en ligne¹²⁶⁶. À l'international, le projet de loi canadien relatif aux données et à l'IA prévoit par exemple la publication « *en langage clair* » de plusieurs informations relatives au fonctionnement global des systèmes d'IA ayant une incidence élevée¹²⁶⁷.

Pour autant, le règlement sur l'IA confère un droit à une explication individuelle, centrée sur la décision reçue par l'individu. À l'image de la loi française pour une république numérique¹²⁶⁸ et la directive canadienne sur la prise de décision automatisée¹²⁶⁹, l'explication proposée par le règlement européen doit contenir des informations sur le rôle du système d'IA à haut risque dans le processus décisionnel et les principaux éléments ayant influencé la décision¹²⁷⁰.

L'explication individuelle comme mécanisme d'accès à des preuves pour le demandeur.

Imposé par le règlement sur l'IA, le déployeur du système d'IA à haut risque doit communiquer à la personne affectée lésée des informations pouvant se révéler intéressantes pour la

¹²⁶⁴ Ce type d'explication est parfois désigné sous le nom de « explication centrée sur l'individu », voir EDWARDS Lilian et VEALE Michael, « Slave to the Algorithm? », *Duke Law Technology Review*, 16, 2017.

¹²⁶⁵ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

¹²⁶⁶ « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. Articles 34, 35, 40 et 69.

¹²⁶⁷ « La personne qui rend disponible un système à incidence élevée publie, sur un site Web accessible au public, selon toute modalité fixée par règlement, une description, en langage clair, du système qui prévoit, notamment, les éléments suivants : a) l'utilisation visée; b) le contenu qu'il est censé générer, les prédictions ou recommandations qu'il est censé faire ou les décisions qu'il est censé prendre; c) les mesures d'atténuation établies à son égard au titre de l'article 8; d) tout autre renseignement prévu par règlement. » - Loi sur l'intelligence artificielle et les données, 16 juin 2022. Article 11.

¹²⁶⁸ « LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique », 2016.

¹²⁶⁹ « Directive sur la prise de décision automatisée (Canada) », 2019.

¹²⁷⁰ « Règlement sur l'intelligence artificielle », 2024. Article 86.1.

constatation d'un préjudice. L'élément le plus prometteur de ce droit à l'explication est l'obtention par la personne affectée des principaux éléments utilisés pour prendre la décision individuelle. La notion de « principaux éléments » n'est pas définie ou illustrée dans le corps du texte ou dans les considérants. Il est donc trop tôt pour deviner le contenu précis du droit à l'explication. Pour autant, il est important d'évaluer l'intérêt, en termes d'accès à la preuve, de ce droit à l'explication pour la personne affectée lésée. Dans un premier temps, il est clair que l'explication ne pourra pas permettre à l'individu de constater le défaut du système d'IA à haut risque ou la non-conformité du système d'IA aux exigences du règlement européen sur l'IA car cela relève du fonctionnement global du logiciel. En revanche, l'explication sur le rôle joué par le logiciel peut révéler des informations potentiellement pertinentes pour démontrer le lien de causalité entre le résultat produit par le système d'IA à haut risque et la décision dommageable rendue.

Dans un second temps, il est important de préciser que l'explication porte sur la décision prise et non sur le résultat produit par le système d'IA. Il faut distinguer selon que la décision est prise exclusivement sur le fondement du résultat du système d'IA à haut risque ou si le résultat n'est qu'un élément parmi d'autres sur lesquels se fonde la décision.

Dans le cas où la décision a été prise sur le fondement partiel d'un résultat produit par un système d'IA à haut risque, alors il n'est pas garanti que la personne affectée obtienne des informations relatives au résultat produit. Cette interprétation nous semble confortée par le changement de termes utilisés par le législateur européen entre la version proposée par le Parlement européen et la version définitive adoptée. Dans sa proposition, le Parlement européen employait le terme de « *principaux paramètres* » tandis que la version définitive utilise celui de « *principaux éléments* ». Le terme de « paramètre » présente un aspect plus technologique que celui « d'éléments ». Par exemple, la législation française relative aux systèmes algorithmiques utilisés dans l'administration impose une obligation de communiquer à l'administré « *les paramètres de traitement [algorithmique], et le cas échéant leur pondération, appliqués à la situation de l'intéressé* »¹²⁷¹. En préférant le terme « d'éléments », le législateur européen privilégie un mot plus neutre techniquement. Par conséquent, dans le cas d'une décision prise partiellement avec un système d'IA à haut risque, il paraît improbable que la personne affectée reçoive des informations sur les paramètres pris en compte par le système d'IA

¹²⁷¹ « Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique », 2017-330, 2017. Codifié à l'article R. 311-3-2-1 du code des relations entre le public et l'administration.

pour produire son résultat. Dans le meilleur des cas, il pourra avoir accès au résultat produit par le système d'IA, pour autant que le déployeur considère qu'il s'agit d'un élément majeur dans la prise de décision.

Dans le cas où la décision a été prise sur le fondement exclusif d'un résultat produit par un système d'IA, il sera possible pour la personne affectée d'obtenir davantage d'informations sur le raisonnement que le système d'IA a appliqué à sa situation. Par exemple, une explication individuelle d'un résultat produit par un système d'IA permet d'identifier la présence d'une différence de traitement potentiellement discriminant de façon plus efficace qu'une explication sur le fonctionnement global du logiciel¹²⁷². Cependant, il convient de nuancer l'utilité d'une telle explication pour l'accès à des éléments de preuve pertinents. En effet, l'intérêt du droit à l'explication est grandement dépendant de la qualité de l'explication fournie par le déployeur. À l'image des exigences substantielles du règlement européen sur l'IA, la pertinence d'une explication peut être mesurée, et donc interprétée, de très nombreuses manières. À titre d'exemple, Facebook avait interprété l'expression « paramètres principaux » comme désignant les paramètres les plus fréquemment utilisés par le logiciel pour sélectionner une publicité ciblée. Par cette interprétation, l'explication fournie par Facebook à ces utilisateurs sur les raisons pour lesquelles ils voyaient une publicité omettait les paramètres les plus influents pour se concentrer sur les plus fréquents¹²⁷³. En d'autres termes, le déployeur a la possibilité d'éviter que l'explication fournie, même de manière automatique, puisse potentiellement engager sa responsabilité ou celle du fournisseur¹²⁷⁴.

Dans le meilleur des cas, l'individu lésé pourrait alors obtenir des éléments de preuves pertinents pour démontrer l'existence d'un fait générateur engageant la responsabilité du fournisseur du logiciel ou d'un lien de causalité avec le dommage. Par exemple, face à des difficultés excessives rencontrées par le demandeur, la proposition de directive européenne sur la responsabilité du fait des produits défectueux énonce une présomption de défaut et/ou de lien de causalité lorsque :

« le demandeur a démontré, sur la base d'éléments de preuve suffisamment pertinents, ce qui suit :

¹²⁷² DODGE Jonathan, LIAO Q. Vera, ZHANG Yunfeng *et al.*, Explaining models: an empirical study of how explanations impact fairness judgment, *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Association for Computing Machinery, 2019.

¹²⁷³ ANDREOU Athanasios, VENKATADRI Giridhari, GOGA Oana *et al.*, Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations, 2018.

¹²⁷⁴ MERRER Erwan Le et TRÉDAN Gilles, « Le problème du videur », *Interstices*, 2021.

a) le produit a contribué au dommage ; et

b) il est probable que le produit était défectueux ou que sa défectuosité est une cause probable du dommage, ou les deux. »¹²⁷⁵

Par conséquent, si le contenu de l'explication n'est pas suffisant pour démontrer clairement les faits constitutifs de la responsabilité du fournisseur du système d'IA, les éléments communiqués pourront tout de même permettre au demandeur de bénéficier des présomptions probatoires mises en place par la proposition de directive sur la responsabilité en matière d'IA¹²⁷⁶ et la proposition de directive sur la responsabilité du fait de produits défectueux¹²⁷⁷.

CONCLUSION DE LA SECTION I :

Entre les obligations documentaires, la mise en place de mécanismes de traçabilité des résultats et le droit à une explication pour la personne affectée, le règlement européen sur l'IA est à l'origine de l'apparition de nombreuses informations concernant les systèmes d'IA à haut risque. À l'exception du droit à l'explication, chacun de ces éléments sert un objectif probatoire afin de faciliter le contrôle du fonctionnement dudit logiciel au regard des exigences du règlement européen. Cependant, au vu des interactions que le texte entretient avec les régimes de responsabilité *ex post*, en particulier la responsabilité du fournisseur, ces éléments de preuves apparaissent particulièrement intéressants dans le cadre d'une action en réparation pour un dommage causé par un système d'IA. L'articulation constatée entre les exigences peut donc être envisagée également au niveau des éléments de preuve. Toutefois, l'individu lésé seul n'a accès qu'à une très faible partie des éléments de preuves existants. De plus, les potentiels éléments de preuve disponibles pour le futur demandeur présentent un intérêt probatoire très limité en comparaison de ceux que possèdent le fournisseur ou les autorités de contrôle. Par conséquent, si ces preuves semblent à première vue pertinentes pour engager, ou exclure, la responsabilité *ex post* du fournisseur, leur réutilisation dans ce contexte dépendra significativement des mécanismes d'accès à ces preuves offerts au demandeur (ou au demandeur potentiel).

¹²⁷⁵ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux*, 28 septembre 2022. Article 9.4.

¹²⁷⁶ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022.

¹²⁷⁷ COMMISSION EUROPÉENNE, *op. cit.*

Section II : Les mécanismes de réutilisation des preuves de conception du système d'IA

Annnonce de plan. Malgré l'apparition de ces nombreux éléments de preuves potentiels, leur accès par l'individu lésé nécessite l'intervention de l'opérateur économique défendeur et/ou celle du juge. Devant cette asymétrie d'information, le législateur européen propose plusieurs dispositions relatives à l'accès aux preuves par le défendeur, mettant en avant les preuves utilisées lors de la certification du système d'IA à haut risque (**Sous-section 1**). Néanmoins, l'importance de la réutilisation de ces preuves, produites lors de la certification du logiciel, dépend également de l'utilité probatoire que les juges y accorderont (**Sous-section 2**).

Sous-section 1 : La mise en avant des preuves de conception par le législateur européen

Annnonce de plan. Dans le cadre d'un dommage causé par un résultat de systèmes d'IA, l'asymétrie d'informations entre les deux parties est incontestable. Ce constat pose une problématique d'égalité entre le demandeur et le défendeur que le législateur tente de régler au travers de différentes dispositions énoncées dans ses deux récentes propositions de directives¹²⁷⁸. Si la démarche est plus que bienvenue, elle opère une influence forte sur la réutilisation des preuves de conception produite en tant qu'éléments de preuve pertinents pour constater les faits constitutifs de responsabilité. Cette dynamique d'articulation est observable au travers de deux mécanismes. Dans un premier temps, le demandeur, à plusieurs reprises, est incité à exploiter les éléments de preuves détenus par le défendeur (**§1**) et, dans un second temps, le législateur européen souhaite accorder des prérogatives au juge qui semblent favoriser la divulgation des preuves existantes (**§2**).

¹²⁷⁸ *Ibid.*

§1. Les incitations à collecter préalablement des preuves auprès du défendeur

Annnonce de plan. L'allègement de la preuve pour le demandeur lésé est un objectif spécifique poursuivi à la fois par la proposition de directive relative à la responsabilité en matière d'IA¹²⁷⁹ et par la proposition de directive sur la responsabilité du fait des produits défectueux¹²⁸⁰. Au travers de ces textes, le législateur européen encourage le demandeur à réutiliser les éléments de preuve issus de la conformité aux normes techniques du système d'IA litigieux pour étayer ses prétentions. Ces incitations prennent la forme de conditions pour pouvoir réaliser une demande d'accès à des preuves auprès du juge **(A)** et de conditions pour bénéficier de présomptions probatoires **(B)**.

A- Une condition à la sollicitation du juge

Une condition préalable à toute demande de divulgation de preuves. Tout d'abord, la nouvelle proposition de directive européenne sur la responsabilité en matière d'IA subordonne toute sollicitation du juge à un travail préalable de collecte d'éléments de preuves de la part du demandeur :

« Dans le cadre d'une action en réparation, la juridiction nationale n'ordonne la divulgation des éléments de preuve par l'une des personnes visées au paragraphe 1 [le fournisseur ou le déployeur] que si le demandeur a mis en œuvre tous les moyens proportionnés pour collecter les éléments de preuve pertinents auprès du défendeur. »¹²⁸¹

¹²⁷⁹ « À cet égard, la présente directive allège la charge de la preuve de manière très ciblée et proportionnée en recourant à la divulgation et aux présomptions réfragables » - COMMISSION EUROPÉENNE, Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. P. 13.

¹²⁸⁰ « En particulier, la présente proposition vise à : [...] alléger la charge de la preuve dans les cas complexes et assouplir les restrictions à l'introduction de demandes en réparation » - COMMISSION EUROPÉENNE, Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 28 septembre 2022. P. 2.

¹²⁸¹ COMMISSION EUROPÉENNE, Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 3.2.

Cette disposition permettrait alors d'éviter une sollicitation excessive des juges à ordonner la production forcée ou une mesure d'instruction d'une ou plusieurs preuves que le défendeur serait prêt à communiquer directement au demandeur. En droit civil français, cette exigence rejoint les conditions de nécessité de la mesure de production forcée par rapport à l'exercice du droit à la preuve et sa proportionnalité par rapport aux intérêts en jeu¹²⁸². La sollicitation du juge n'est donc pas nécessaire si l'alternative de la production volontaire est disponible, ou du moins si elle n'a pas été explorée par le demandeur. De ce fait, la question de la proportionnalité ne se pose même pas. Une logique similaire s'applique dans le cas d'une demande d'instruction qui ne peut pas être ordonnée si le demandeur dispose déjà des éléments suffisants pour prouver le fait en question :

« Une mesure d'instruction ne peut être ordonnée sur un fait que si la partie qui l'allègue ne dispose pas d'éléments suffisants pour le prouver.

*En aucun cas une mesure d'instruction ne peut être ordonnée en vue de suppléer la carence de la partie dans l'administration de la preuve. »*¹²⁸³

Cette subsidiarité des mesures d'instruction¹²⁸⁴ est réaffirmée au travers de la condition d'opportunité de la mesure, une évaluation laissée à l'appréciation souveraine du juge en droit français¹²⁸⁵. Le demandeur doit donc exploiter les modalités d'accès aux preuves pertinentes à sa disposition avant de solliciter l'aide de la juridiction.

Les conséquences sur la réutilisation des éléments issus du régime de responsabilité *ex ante*. La présence de cette condition impose donc au demandeur d'exploiter les éléments de preuves à disposition du défendeur. Ce n'est que dans un second temps que leur pertinence sera appréciée par le juge. Les preuves à disposition du défendeur seront celles qu'il détient au moment de la sollicitation par le défendeur. Par conséquent, il s'agira principalement de preuves que le défendeur a préconstitué ou qui lui ont été fournies par un autre opérateur économique. Par exemple, cela peut être les informations que le fournisseur du système d'IA à haut risque a communiquées au déployeur du logiciel¹²⁸⁶. Dès lors, il s'agit principalement des éléments de

¹²⁸² LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. PP.299-300

¹²⁸³ Article 146 du code de procédure civile.

¹²⁸⁴ Pour plus d'informations sur cette subsidiarité, voir CADIET Loïc et JEULAND Emmanuel, *Droit judiciaire privé*, LexisNexis, 2020.

¹²⁸⁵ L'étude de la notion de pertinence dans l'articulation entre les différents éléments de preuve sera plus amplement discutée dans Sous-section 2 : L'évaluation incertaine de l'utilité probatoire par le juge.

¹²⁸⁶ En particulier la notice d'utilisation dont la communication au déployeur est imposée au fournisseur. « Règlement sur l'intelligence artificielle », 2024. Article 13.2.

preuves issus des procédures de mise en conformité, du certificat de conformité européen obtenu aux résultats des tests de qualité réalisés par le fournisseur.

B- Une condition pour bénéficier des présomptions probatoires

Une condition aux présomptions pour le défendeur. Le besoin de recherche des éléments de preuve auprès du défendeur est également motivé par le bénéfice des présomptions probatoires. La directive européenne relative à la responsabilité en matière d'IA insiste à plusieurs reprises sur l'accès à la preuve, ou plus précisément sur la difficulté d'accès à la preuve, pour justifier de la nécessité de recourir à une présomption concernant l'existence d'un lien de causalité. Ainsi, pour un système d'IA qualifié à haut risque au sens du règlement européen sur l'IA, le défendeur peut empêcher le demandeur de bénéficier de cette présomption de lien causal s'il « *démontre que le demandeur peut raisonnablement accéder à une expertise et à des éléments de preuve suffisants pour prouver le lien de causalité [...]* »¹²⁸⁷. Dès lors que le défendeur accepte de communiquer certains éléments de preuves en rapport avec le dommage et de mettre à disposition une personne compétente sur le système d'IA litigieux, il deviendra difficile pour le demandeur de bénéficier de cette présomption. Par conséquent, le demandeur qui souhaite bénéficier de la présomption de lien de causalité doit démontrer que les informations et l'assistance fournie par le défendeur ne sont pas « *suffisantes* » pour apporter la preuve de l'existence d'un lien de causalité entre sa faute et le dommage subi. Ainsi, il apparaît particulièrement difficile pour le demandeur de convaincre la juridiction que l'insuffisance des informations communiquées est due à leur qualité et/ou leur quantité et non à l'absence d'un lien de causalité.

En ce qui concerne les systèmes d'IA à moyens ou faibles risques, la présomption de lien causal est accordée seulement « *si la juridiction nationale estime qu'il est excessivement difficile pour le demandeur de prouver le lien de causalité [...]* »¹²⁸⁸. À nouveau, la condition de difficulté excessive mentionnée ici renvoie à l'accès aux éléments de preuve par le demandeur. La demande de communication, auprès du défendeur, d'informations qui pourraient lui permettre

¹²⁸⁷ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA)*, 28 septembre 2022. Article 4.4.

¹²⁸⁸ *Ibid.* Article 4.5.

de prouver le lien de causalité est donc là aussi une étape préalable indispensable pour la victime d'un dommage causé par un système d'IA.

Une condition relative de présomption dans le cadre de la responsabilité du fait des produits défectueux. La dépendance du demandeur aux informations à la disposition du défendeur se retrouve également au sein de la proposition de directive relative à la responsabilité du fait des produits défectueux. Le lien de causalité et/ou le défaut peut être présumé par le juge national lorsque le demandeur rencontre des « *difficultés excessives, en raison d'une complexité technique ou scientifique* »¹²⁸⁹. La satisfaction de cette condition peut être contestée par le défendeur. En pratique, cela peut s'illustrer par l'apport d'éléments de preuves pertinents détenus par le défendeur et qui permettront d'attester concrètement de l'existence (ou de l'absence) d'un défaut et/ou du lien de causalité avec le dommage. Par conséquent, qu'il s'agisse de la juridiction nationale refusant la présomption¹²⁹⁰ ou du défendeur qui la conteste, le demandeur doit s'assurer d'avoir exploité toutes ses possibilités de réunir l'ensemble des preuves pertinentes pouvant lui être communiquées s'il souhaite profiter des mécanismes allégeant sa charge probatoire.

§2. La priorisation de la divulgation des preuves détenues par les acteurs de la conformité

Annonce de plan. Le dynamique de réutilisation des éléments de preuve issus de la mise en conformité des systèmes d'IA s'observe également au travers de la place du juge dans les propositions de directives européennes. Ainsi, le juge dispose de prérogatives à destination des opérateurs économiques **(A)**. Les mesures prévues par la Commission européenne se concentrent en priorité sur l'accès aux preuves déjà existantes en possession du défendeur **(B)**.

¹²⁸⁹ COMMISSION EUROPÉENNE, *Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux*, 28 septembre 2022. Article 9.4.

¹²⁹⁰ MANGEMATIN Céline, « Produits défectueux - Responsabilité du fait des produits défectueux et intelligence artificielle : une proposition presque parfaite », *Responsabilité civile et assurances*, 2023. P. 10.

A- Les destinataires des mesures prononcées par le juge

Le destinataire des mesures d'accès aux preuves dans les propositions de directives européennes. Notre analyse des preuves accessibles en cas de dommage causé par un système d'IA impose de s'attarder sur le destinataire des mesures d'instruction que peuvent prononcer les juridictions. Les propositions de directives européennes relative à la responsabilité civile extracontractuelle et relative à la responsabilité du fait des produits défectueux désignent explicitement le défendeur comme le destinataire central de ces mesures de production forcée de preuves. Au sein de la nouvelle directive sur la responsabilité du fait des produits défectueux, le défendeur est le seul acteur mentionné explicitement comme destinataire d'une demande de production de preuves :

« Les États membres veillent à ce que les juridictions nationales soient habilitées [...] à ordonner au défendeur de divulguer les éléments de preuve pertinents dont il dispose. »¹²⁹¹

« Les États membres veillent à ce que, lorsqu'il est ordonné à un défendeur de divulguer des informations [...]. »¹²⁹²

De manière similaire, la proposition de directive relative à la responsabilité en matière d'IA traite le défendeur comme la source principale de preuves pertinentes dans le cadre d'une action en réparation. Plus précisément, la proposition européenne désigne comme destinataire des mesures de divulgation de preuves : le fournisseur, la personne soumise aux obligations soumises aux obligations du fournisseur ou l'utilisateur du système d'IA à haut risque. L'article 3.1. dispose que :

« Les États membres veillent à ce que les juridictions nationales soient habilitées, soit à la demande d'un demandeur potentiel qui a précédemment demandé à un fournisseur, à une personne soumise aux obligations incombant au fournisseur en vertu de [l'article 24 ou de l'article 28, paragraphe 1, de la législation sur l'IA] ou à un utilisateur de divulguer les éléments de preuve pertinents qu'il détient concernant un système d'IA à haut risque spécifique soupçonné d'avoir causé un dommage, mais dont la

¹²⁹¹ *Ibid.* Article 8.1.

¹²⁹² *Ibid.* Article 8.4.

*demande a été refusée, soit à la demande d'un demandeur, à ordonner la divulgation de ces éléments de preuve par ces personnes. »*¹²⁹³

Contrairement à la responsabilité du fait des produits défectueux, la proposition de directive européenne en matière d'IA semble ne s'appliquer qu'aux systèmes d'IA qualifiés à haut risque au sens du règlement européen sur l'IA. La sollicitation par le juge de mesures de divulgation des preuves pertinentes à un tiers (une personne n'étant pas le défendeur dans le litige) est envisagée seulement en dernier recours, lorsque que toutes les tentatives auprès du défendeur ont échoué¹²⁹⁴.

Les destinataires de mesures d'accès aux preuves par le juge en droit français. En matière civile, dès lors que le juge est sollicité par une partie¹²⁹⁵, il peut adresser la mesure de production forcée de preuves pertinentes au défendeur mais également envers le(s) tiers au litige¹²⁹⁶. Le maintien de la possibilité d'ordonner la production d'informations à des tiers permet de diversifier les sources de preuves disponibles dans le contexte de l'IA. Il est ainsi possible de mettre en relation les informations fournies par le défendeur à d'autres renseignements détenus par des tiers afin de reconstituer plus aisément les faits déterminants dans le litige. Dans le cadre d'un dommage causé par un système d'IA à haut risque, il serait pertinent pour le juge de pouvoir obtenir des informations provenant de l'organisme notifié, le cas échéant, voire de l'autorité notifiante ou de l'autorité de surveillance du marché concernant par exemple les tests et contrôles que ces derniers ont réalisés sur le système d'IA à haut risque litigieux. Dans le contexte d'une potentielle atteinte aux droits fondamentaux, les informations détenues par les « *autorités ou organismes publics nationaux qui supervisent ou font respecter les obligations au titre du droit de l'Union visant à protéger les droits fondamentaux en ce qui concerne l'utilisation des systèmes d'IA [...]* »¹²⁹⁷ relatives au fonctionnement global du système d'IA

¹²⁹³ « Les États membres veillent à ce que les juridictions nationales soient habilitées, soit à la demande d'un demandeur potentiel qui a précédemment demandé à un fournisseur, à une personne soumise aux obligations incombant au fournisseur en vertu de [l'article 24 ou de l'article 28, paragraphe 1, de la législation sur l'IA] ou à un utilisateur de divulguer les éléments de preuve pertinents qu'il détient concernant un système d'IA à haut risque spécifique soupçonné d'avoir causé un dommage, mais dont la demande a été refusée, soit à la demande d'un demandeur, à ordonner la divulgation de ces éléments de preuve par ces personnes. » - Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022.

¹²⁹⁴ Ibid. Préambule, p. 15.

¹²⁹⁵ Cour de Cassation, Chambre civile 1, 21 octobre 1975, n° 74-12.73.

¹²⁹⁶ Voir notamment Cour de cassation, Chambre civile 2, 26 mai 2011, n° 10-20.048.

¹²⁹⁷ « Règlement sur l'intelligence artificielle », 2024. Article 64.3.

litigieux peuvent permettre d'éclairer la probabilité pour le demandeur d'avoir subi une atteinte disproportionnée.

B- La nature des prérogatives accordées au juge

La concentration du législateur européen sur les mesures de production forcée. Les deux propositions de directives européennes contiennent deux actions spécifiques relatives à l'accès aux preuves par le juge. Premièrement, le législateur européen impose aux États membres de garantir la capacité des juges nationaux à pouvoir ordonner au défendeur la divulgation des éléments de preuve considérés comme pertinents pour le litige en question¹²⁹⁸. Pour rappel, la divulgation est définie juridiquement comme le « *fait de donner de la publicité à une donnée d'information ou de recherche non encore connue (ex. document, découverte, etc.) qui prend un caractère délictueux lorsque la révélation s'opère en violation d'un secret* »¹²⁹⁹. Secondement, la proposition de directive relative à la responsabilité en matière d'IA exige que « *les juridictions nationales, à la demande d'un demandeur, soient habilitées à ordonner des mesures spécifiques pour conserver les éléments de preuve* »¹³⁰⁰. Dans les deux cas, le législateur européen concentre son attention sur des actions portant sur des informations préexistantes en possession du destinataire de la mesure, qu'il s'agisse d'ordonner leur communication ou leur conservation.

L'importance des mesures d'instruction. Les directives européennes demeurent silencieuses sur la capacité du juge à ordonner la génération de nouvelles informations concernant le fonctionnement du système d'IA dans sa globalité et/ou la décision individuelle responsable du dommage subi par le demandeur. En comparaison, le règlement européen sur l'IA énonce explicitement la possibilité pour les autorités de surveillance du marché de réaliser des enquêtes auprès des fournisseurs de systèmes d'IA à haut risque, une prérogative déjà prévue par le

¹²⁹⁸ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 3. Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 28 septembre 2022. Article 8.

¹²⁹⁹ CORNU Gérard, Vocabulaire juridique, 2022.

¹³⁰⁰ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022.

règlement (UE) 2019/1020¹³⁰¹. Ensuite, les autorités de surveillance du marché se voient également accorder la possibilité d'accéder, sous conditions, à la documentation technique, aux jeux de données de conception¹³⁰² ainsi qu'au code-source du logiciel¹³⁰³. Enfin, le législateur rappelle que les autorités de surveillance sont compétentes pour « *recueillir des échantillons du produit, y compris sous une fausse identité, et de les soumettre à une inspection et à la rétro-ingénierie afin de détecter une non-conformité et d'obtenir des preuves* »¹³⁰⁴. En dépit de la redondance avec d'autres réglementations européennes, le législateur européen a choisi de rappeler avec clarté les possibilités pour les autorités de surveillance du marché de générer leurs propres éléments de preuve. À la lecture des dispositions proposées pour les dommages causés par des systèmes d'IA, le choix a été fait de ne pas insister sur ce type de prérogatives de génération proactive de preuves par le juge.

Pour autant, cela ne signifie pas que le juge est dépourvu de moyens d'accéder à de nouveaux éléments de preuves qui ne seraient pas déjà constitués par le fournisseur ou le déployeur du système d'IA. Au travers des mesures d'instructions, les juges français possèdent des prérogatives d'accès à la preuve bien plus variées que les actions mentionnées dans les deux propositions de directives européennes. Le code de procédure civile précise que les mesures d'instruction à disposition du juge doivent poursuivre soit un objectif d'établissement de preuves soit un objectif de conservation¹³⁰⁵. Pour parvenir à ces objectifs, les mesures peuvent prendre la forme d'une enquête mais aussi de la désignation d'un expert ou encore d'un huissier de justice. En d'autres termes, le juge est capable d'aller rechercher les éléments de preuve et n'est pas limité à la réception des éléments de preuve par celui qui les détient. Ainsi, le juge peut ordonner à un huissier de justice de rechercher des faits en rapport avec des faits de concurrence déloyale¹³⁰⁶. Il est toutefois indispensable que la mesure d'instruction prononcée soit précise et limitée aux faits litigieux¹³⁰⁷ sans quoi la mesure devient illicite. C'est ainsi

¹³⁰¹ « Les États membres confèrent à leurs autorités de surveillance du marché les pouvoirs de surveillance du marché, d'enquête et d'exécution nécessaires à l'application du présent règlement et à l'application de la législation d'harmonisation de l'Union. » - « Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019.

¹³⁰² « Règlement sur l'intelligence artificielle », 2024. Article 74.12.

¹³⁰³ *Ibid.* Article 74.13.

¹³⁰⁴ *Ibid.* Article 73.5. Il s'agit d'un renvoi à l'article 14 par. 4, point j) du règlement (UE) 2019/1020.

¹³⁰⁵ « S'il existe un motif légitime de conserver ou d'établir avant tout procès la preuve de faits dont pourrait dépendre la solution d'un litige [...] » - Article 145 du code de procédure civile. La question du motif légitime sera traitée ci-après.

¹³⁰⁶ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

¹³⁰⁷ *Ibid.*

qu'une mesure générale d'investigation portant sur l'ensemble de l'activité du défendeur a été considérée comme excédant les prévisions de l'article 145 du code de procédure civile¹³⁰⁸.

Par conséquent, il est envisageable que le juge français puisse ordonner à un expert, accompagné d'un informaticien si nécessaire, de réaliser une série de tests sur le système d'IA litigieux afin de révéler, par exemple, un risque plus important d'atteinte aux droits fondamentaux pour des situations similaires à celles du demandeur. Cette capacité d'investigation est un outil très important car elle permet au juge de se détacher partiellement de la collaboration du défendeur pour devenir lui aussi un acteur proactif dans la production de la preuve et non uniquement un collecteur passif d'informations existantes. L'importance de réaliser ses propres enquêtes et expertises a déjà été soulevée lors de l'étude des compétences des autorités de surveillance des marchés¹³⁰⁹. Premièrement, cela permet à leur commanditaire de ne pas devoir se fier aveuglément aux informations fournies par le défendeur, ou en dernier recours au tiers destinataire de la mesure de production de preuves. Deuxièmement, les mesures d'instructions offrent l'opportunité au juge d'investir la question de l'adéquation entre les choix des éléments de preuves réalisés par le fournisseur du système d'IA, les exigences de conception du règlement européen sur l'IA et les faits constitutifs de responsabilité prévus par les régimes de responsabilité *ex post*.

Sous-section 2 : L'évaluation incertaine de l'utilité probatoire par le juge

Annonce de plan. La réutilisation des preuves issus de la certification du système d'IA à haut risque dans le contexte d'une action en réparation du fait du logiciel dépend aussi du juge. En effet, si le législateur européen semble avoir choisi de mettre en avant les mesures de divulgation des éléments de preuves préexistants, c'est le juge qui a la charge de les ordonner. Le rôle du juge dans la réutilisation des preuves de conception, notamment les preuves techniques relatives au fonctionnement du logiciel, est intimement lié à la notion d'utilité probatoire (§1). Le juge sera alors amené à se prononcer indirectement sur la pertinence des éléments de preuves issus du processus de certification par le biais de son évaluation de l'utilité probatoire des mesures d'accès aux preuves (§2).

¹³⁰⁸ Cour de Cassation, Chambre civile 2, 7 janvier 1999, n° 97-10.831 COLLINGRIDGE David, *The Social Control of Technology*, Frances Pinter, 1980.

¹³⁰⁹ Voir Sous-section 1 : Le contrôle administratif de la .

§1. La théorie de l'utilité probatoire

La condition d'utilité du fait probatoire. *Frustra probatur quod probatum non relevant*¹³¹⁰. Cette maxime du 19^{ème} siècle, qui signifie « *il est vain de prouver les faits qui ne sont pas pertinents* », résume l'idée de pertinence ou d'utilité de la preuve¹³¹¹. En dépit de son apparente évidence, cet adage impose et offre au juge la tâche d'évaluer la pertinence de l'élément de preuve qui est présenté devant lui. Un fait probatoire ou preuve est pertinent dès lors qu'il permet d'établir des faits utiles à la résolution du procès¹³¹². L'évaluation de l'utilité d'un fait au sein d'un litige dépend de nombreux éléments qu'il conviendra de rappeler (caractéristiques intrinsèques, extrinsèques, etc.). Également, la pertinence de la preuve est évaluée différemment selon que la preuve est produite par la partie ou que la production de la preuve est demandée par une partie qui n'en dispose pas¹³¹³. Enfin, la sanction de la preuve d'un fait non pertinent diverge selon la manière dont elle est abordée par les juridictions. Ces trois éléments constituent autant de moyens pour le juge d'évaluer la pertinence des différentes preuves disponibles dans le cadre d'un préjudice causé par un système d'IA.

La place de la pertinence du fait à prouver. La notion de pertinence du fait probatoire est rare au sein des dispositions du code de procédure civile (CPC ci-après). Une rapide recherche dans ledit code permet de constater que le terme n'apparaît qu'à une seule reprise :

*« La partie qui demande une enquête doit préciser les faits dont elle entend rapporter la preuve. Il appartient au juge qui ordonne l'enquête de déterminer les faits pertinents à prouver. »*¹³¹⁴

La proposition de directive européenne relative à la responsabilité en matière d'IA et celle relative à la responsabilité du fait des produits défectueux mentionnent le terme respectivement quatre et trois fois¹³¹⁵. L'emploi du terme poursuit différentes finalités. Dans les deux propositions européennes, il sert à caractériser les éléments de preuves dont le juge peut

¹³¹⁰ ROGRON Joseph-Adrien, *Code de procédure civile expliqué par ses motifs et par des exemples.*, 1837. Cité dans LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.* P. 178.

¹³¹¹ Les termes « utilité » et « pertinence » seront ici utilisés comme des synonymes.

¹³¹² BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010. P. 314.

¹³¹³ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.*

¹³¹⁴ Article 222 du code de procédure civile.

¹³¹⁵ Il s'agit des utilisations de la notion de pertinence en relation avec les preuves dans le corps des textes. La proposition de directive européenne sur la responsabilité en matière d'IA fait référence à plusieurs reprises à ce concept dans ses considérants. Ces utilisations n'ont pas été prise en compte dans le comptage.

demander la divulgation¹³¹⁶. La directive européenne sur la responsabilité en matière y a recours pour qualifier les preuves que le demandeur doit avoir recherchées de façon proportionnée avant de demander au juge de prononcer une mesure de divulgation¹³¹⁷. Enfin, la proposition de directive de 2022 relative à la responsabilité du fait des produits défectueux utilise la notion de pertinence comme condition de l'octroi de la présomption juridictionnelle du lien de causalité¹³¹⁸. Dans ce contexte, les preuves pertinentes apparaissent ainsi comme les éléments auxquels le demandeur doit avoir accès pour étayer le bien-fondé de son initiative, et donc comme l'objectif poursuivi par le législateur au travers de ces dispositions.

Bien qu'à première vue, la condition de pertinence semble binaire (la preuve est pertinente ou ne l'est pas), il existe différents degrés dans l'utilité d'un élément de preuve. Pour présenter ces variations de la pertinence d'une preuve, il sera repris la classification réalisée par Aurélie Bergeaud entre l'élément pertinent et l'élément concluant¹³¹⁹. Le fait probatoire pertinent est défini comme le fait qui se rapporte au litige et qui peut influencer la solution du litige. Il se distingue de l'élément concluant qui exerce quant à lui une influence décisive sur la solution du litige. Il est important de préciser que cette distinction est purement doctrinale¹³²⁰. Cependant elle permet d'identifier les trois niveaux que pourraient se voir octroyer des éléments de preuves issus de la conception d'un système d'IA à haut risque : non pertinents, pertinents ou concluants.

La théorie de la pertinence du fait probatoire appliquée au contexte des systèmes d'IA.

Afin d'illustrer cette distinction et sa pertinence dans le contexte de l'IA, prenons l'exemple d'un préjudice causé par un système d'IA à haut risque. Le demandeur ou demandeur potentiel peut choisir de démontrer que le fournisseur de ce système a commis une faute au regard du

¹³¹⁶ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 3.1 COMMISSION EUROPÉENNE, Directive du Parlement européen et du Conseil concernant la responsabilité du fait des produits défectueux, 28 septembre 2022. Article 8.1.

¹³¹⁷ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 3.2.

¹³¹⁸ « Lorsqu'une juridiction nationale juge que le demandeur rencontre des difficultés excessives, en raison d'une certaine complexité technique ou scientifique, pour prouver la défectuosité du produit ou le lien de causalité entre sa défectuosité et le dommage, ou les deux, la défectuosité du produit ou le lien de causalité entre sa défectuosité et le dommage, ou les deux, sont présumés lorsque le demandeur a démontré, sur la base d'éléments de preuve suffisamment pertinents [...] » - COMMISSION EUROPÉENNE, op. cit. Article 9.4

¹³¹⁹ BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010. PP. 315-320.

¹³²⁰ Le critère de l'élément de preuve concluant n'est que peu employé dans la jurisprudence civile de la Cour de cassation et celle-ci ne semble pas chercher à le distinguer de la notion d'élément pertinent : LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

règlement européen sur l'IA¹³²¹. Que ce soit sur demande du juge ou à son initiative, le fournisseur communique au demandeur potentiel une liste d'éléments de preuves plus ou moins détaillés¹³²² destinés à démontrer qu'il est conforme à la réglementation européenne, en particulier aux normes européennes harmonisées. Les éléments de preuve en question sont certes pertinents mais ils seraient dangereux de les considérer comme concluants, cela pour deux raisons.

Premièrement, la conformité aux normes européennes harmonisées n'entraîne qu'une présomption simple de conformité aux exigences du règlement européen sur l'IA. L'existence de cette présomption simple renvoie à un déplacement de l'objet de la preuve. Dans ce cas, le fournisseur du système d'IA à haut risque prouve alors sa conformité aux normes européennes harmonisées, et non plus au règlement européen sur l'IA. Le fournisseur du système d'IA à haut risque peut donc apporter une preuve indirecte de sa conformité à la réglementation européenne. Le mécanisme est similaire à celui de la pertinence de l'élément de preuve, à la différence que le déplacement de l'objet de la preuve n'est pas réalisé par le plaideur mais par le législateur. Néanmoins, il s'agit bien d'une preuve indirecte de la conformité qui est rapportée par le fournisseur du système d'IA à haut risque. Par conséquent, cette preuve ne peut pas être considérée comme concluante mais au maximum comme pertinente.

Deuxièmement, le choix des éléments de preuves utilisés lors de la mise en conformité du système d'IA à haut risque est le résultat d'une interprétation des exigences par le fournisseur. Par un processus de formalisation, cette interprétation peut dissimuler une dénaturation de l'exigence en question, ou de manière générale, s'éloigner de l'interprétation faite par le juge. L'exemple de la divergence entre la parité statistique et l'égalité d'opportunité comme élément technique de preuve de l'égalité de traitement illustre parfaitement ce phénomène¹³²³. Ce risque de dénaturation au travers des éléments de preuve de la conformité pourrait être soulevé lors de l'évaluation de la pertinence de ces faits probatoires par le juge. En d'autres termes, il est envisageable que le juge considère que le fait prouvé par ces éléments, à savoir la situation d'une différence acceptable de traitement entre deux catégories de population (ou entre un individu et d'autres individus comparables) par le système d'IA, est trop éloigné de la conception d'une égalité de traitement au sens du droit à la non-discrimination pour constituer

¹³²¹ Sur le lien entre la faute du fournisseur du système d'IA à haut risque et la conformité avec le règlement européen sur l'IA, voir §1. La faute du fournisseur.

¹³²² Comprenant notamment ceux listés dans la section précédente (contenu des journaux d'enregistrement pour la décision de l'individu, contenu plus ou moins détaillé de la documentation technique, etc.).

¹³²³ Voir Sous-section 2 : La place concrète des preuves techniques au sein du règlement.

un fait pertinent dans le litige. Dans cette hypothèse, les éléments de preuve considérés comme pertinents au travers du règlement sur l'IA perdraient cette qualité dans le cadre d'un litige pour discrimination. Une telle décision de justice pourrait alors interroger le choix des éléments de preuve réalisé et validé par les fournisseurs et/ou par les organismes de normalisation. *A minima*, cela mettrait en avant les limites entre la conformité de fonctionnement d'un système d'IA et la conformité légale de ses résultats. Bien qu'il s'agisse d'un scénario prospectif, le critère de la pertinence de la preuve met en avant, d'une part, l'interaction des preuves de conception avec d'autres régimes juridiques (ici le droit à la non-discrimination) et, d'autre part, l'influence qu'ils peuvent exercer l'un sur l'autre.

Le prisme de l'utilité probatoire permet enfin de montrer la place que le juge pourrait avoir dans cette articulation. Il a été démontré que le juge détiendra un pouvoir très limité dans le contrôle des obligations du règlement européen sur l'IA et les normes techniques qui y seront associées. À cette étape, les acteurs en charge de la détermination de ces règles, et des éléments techniques de preuves afférents, sont principalement les organismes de normalisation, les autorités de la surveillance du marché, les organismes notifiés et la Commission européenne. Toutefois, l'articulation du règlement européen sur l'IA avec autres régimes juridiques appliqués à l'utilisation des systèmes d'IA, qu'il s'agisse des régimes de responsabilité ou de la protection des droits fondamentaux, sont une occasion pour le juge d'apprécier la portée des choix réalisés par les acteurs de la conformité des systèmes d'IA à haut risque.

§2. L'utilité du procédé probatoire

Différence entre l'offre de la preuve et la demande de preuve. La pertinence des éléments de preuves, plus précisément du fait qu'il cherche à prouver, est abordée différemment par les juridictions selon la manière dont les preuves en question sont apportées. En droit français, l'évaluation de la pertinence du fait que la partie cherche à attester n'est pas réalisée lorsque l'élément de preuve est fourni par ladite partie. En effet, les juges apprécient la pertinence uniquement lorsqu'une partie sollicite leur intervention pour obtenir des éléments de preuve¹³²⁴. En d'autres termes, la demande de preuve est soumise à l'examen de la pertinence tandis que l'offre de preuve n'y est pas sujette. Le critère de la pertinence devient alors une condition de

¹³²⁴ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 183.

la réalisation d'une mesure d'instruction¹³²⁵ ou de production forcée¹³²⁶ par le juge. Pour reprendre l'expression d'Aurélie Bergeaud, l'évaluation de la pertinence ne porte plus sur le fait objet de la preuve mais sur le moyen de preuve dont les juges apprécient l'intérêt probatoire¹³²⁷. Par conséquent, la pertinence est une notion essentielle de l'accès à la preuve par le demandeur.

La place des différentes utilisations du terme « pertinence » dans le code de procédure civile français ou dans les directives européennes sur la responsabilité, présentée précédemment, témoigne de l'utilisation exclusive de la notion pour des mécanismes de communication de la preuve au travers du juge. L'article 222 du CPC porte ainsi sur les mesures d'enquête que peut ordonner un juge. Également, la notion est employée indirectement à l'article 143 du CPC relatif aux décisions des mesures d'instruction : « *Les faits dont dépend la solution du litige peuvent, à la demande des parties ou d'office, être l'objet de toute mesure d'instruction légalement admissible.* »¹³²⁸. À contrario, si la mesure d'instruction demandée est considérée comme inutile, elle est refusée par les juridictions¹³²⁹.

L'évaluation de la pertinence du procédé probatoire est réalisée au regard des qualités intrinsèques dudit procédé de preuve mais surtout en fonction des éléments de preuve déjà existants au moment de la demande¹³³⁰. Dans un premier temps, en ce qui concerne les qualités propres du procédé probatoire, la jurisprudence offre plusieurs exemples de caractéristiques pouvant nuire à l'utilité de la mesure aux yeux des juges. Ainsi, une mesure d'instruction sera jugée inutile s'il est manifestement impossible qu'elle permette de recueillir les éléments de preuves pertinents souhaités par la partie. L'incapacité matérielle du procédé probatoire est ainsi caractérisée dans le cas d'une demande d'expertise visant à produire une comparaison entre l'état d'un immeuble avant et après un incendie¹³³¹. Le procédé probatoire ne peut donc pas influencer la solution du litige, il n'est par conséquent pas pertinent.

¹³²⁵ Cour de Cassation, Chambre commerciale, 28 mai 1974, n° 73-10.022.

¹³²⁶ Cour de Cassation, Chambre sociale, 7 mai 1998, n° 95-44.382.

¹³²⁷ BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010.

¹³²⁸ Article 143 du code de procédure civile.

¹³²⁹ « *La cour d'appel a fait ressortir l'inutilité de cette mesure d'instruction ; d'où il suit que [...] le deuxième moyen fondé [la méconnaissance des articles 9, 10 et 11 du CPC suite au refus de prescrire l'enquête sollicitée] n'est pas* ». Cour de Cassation, Chambre civile 1, 2 mai 1989, n° 86-17.804.

¹³³⁰ BERGEAUD Aurélie, *op. cit.*

¹³³¹ « *une comparaison entre l'état des immeubles avant et après l'incendie, permettant de déterminer les seuls désordres imputables à ce sinistre dont M. X... pouvait demander réparation, n'était pas possible, de sorte que la mesure était inutile* » - Cour de cassation, Chambre civile 2, 20 mars 2014, n° 13-14.985.

Dans un second temps, le procédé probatoire demandé au juge peut être qualifié d'inutile ou de non pertinent s'il ne peut plus influencer la solution du litige. C'est le cas lorsque le juge a déjà formé sa conviction et qu'aucun élément de preuve supplémentaire ne pourra venir le faire changer d'avis.¹³³² Précédemment, nous avons mentionné le fait que le législateur imposait au demandeur de rechercher les preuves pertinentes auprès du défendeur avant de recourir au juge. Ceci est une pratique logique et déjà reconnue en droit français¹³³³. L'ordre d'accès aux éléments de preuve peut donc avoir un effet sur l'accès par les parties à certaines mesures d'instruction ou de production forcée par le biais du juge. Pour reprendre la distinction doctrinale sur les degrés de pertinence, il est possible d'imaginer qu'en présence d'une preuve, ou d'un ensemble de preuves, considérée comme concluante pour la solution du litige, alors le juge refusera toute demande ultérieure de mesures d'accès à d'autres éléments de preuve.

L'influence des preuves de conception du système d'IA sur l'utilité de la demande de preuve par le demandeur. L'évaluation de l'utilité probatoire offre un prisme d'analyse complémentaire de l'articulation entre les preuves issues des obligations de conception du système d'IA à haut risque et les preuves pertinentes dans le cadre des régimes de responsabilité *ex post*. Dans le cadre d'un préjudice causé du fait d'un résultat d'un système d'IA, le demandeur lésé n'a aucune ou presque aucune preuve pertinente à sa disposition. Son accès aux preuves pertinentes dans le cadre du litige dépend exclusivement du défendeur et du juge.

Dès lors, l'acceptation d'une demande de preuve technique de conformité par le juge confirme la pertinence de l'élément de preuve recherché, et donc sa réutilisation dans le contexte de l'action en réparation. Prenons par exemple le cas d'un demandeur lésé par un système d'IA à haut risque qui souhaite démontrer que le fournisseur dudit logiciel a commis une faute. Dans notre illustration, la faute en question serait le manquement à l'article 15 du règlement sur l'IA concernant l'exactitude et la robustesse du logiciel. Le défendeur apporte alors des éléments de preuve démontrant que son système d'IA à haut risque respecte les normes européennes harmonisées, notamment sur l'obligation de l'article 15. Il peut fournir la preuve de sa certification, la présence de marquage CE, voire les résultats obtenus à l'issue des tests techniques réalisés pendant la conception. Persuadé de la non-conformité effective du système d'IA, le demandeur lésé peut demander au juge la réalisation de tests sur le logiciel afin de vérifier la véracité des éléments de preuves fournis par le défendeur. Également, le demandeur

¹³³² « Mais attendu que le juge n'est pas tenu d'ordonner une mesure d'instruction quand sa conviction est formée ; » - Cour de Cassation, Chambre civile 2, 12 mars 1970, n° 69-12.291.

¹³³³ Article 146 du code de procédure civile.

peut demander la réalisation de tests alternatifs au motif que la méthode d'évaluation de l'exactitude choisie par le fournisseur n'est pas adéquate et ne représente pas correctement la réalité du fonctionnement du logiciel. En d'autres termes, le demandeur désapprouve le choix de la métrique (ou indicateur) utilisé et considère qu'un choix différent mettrait en lumière la non-conformité du logiciel. Si le juge refuse cette demande du fait de l'absence d'utilité probatoire de ces mesures, il confirme, d'une part, la pertinence des éléments de preuves apportés par le fournisseur et, d'autre part, l'absence de pertinence des preuves demandées par l'individu lésé. À *contrario*, si le juge fait droit à la demande, il vient nuancer la pertinence des preuves de certification apportées par le fournisseur et reconnaît la pertinence des éléments de preuves souhaitées par le demandeur. Au fil des décisions, un ensemble d'éléments de preuves considérés comme pertinents par les juges pourrait alors être identifié.

Les effets sur le processus de certification des systèmes d'IA à haut risque. Le critère de l'utilité probatoire confère aux juges une double influence dans l'articulation entre les preuves de conception du système d'IA et les preuves pertinentes dans le cadre d'un dommage causé par ce même système d'IA. Dans un premier temps, son évaluation de la pertinence des preuves apportées lui permet de questionner les choix réalisés par les acteurs de la conformité, en particulier l'adéquation entre lesdites preuves et les exigences auxquelles elles sont associées. Dans un second temps, l'évaluation de l'intérêt probatoire d'une demande de preuve par une partie lui permet indirectement de mettre en lumière des éléments de preuves qu'il considère comme pertinents et qui auraient été négligés dans l'élaboration de la conformité légale des systèmes d'IA. Ainsi le juge retrouve une certaine place dans le droit de l'IA, et indirectement dans le droit des systèmes d'IA. Néanmoins, les discussions élaborées ici sont très complexes techniquement et la pertinence de chaque élément de preuve peut varier selon le contexte d'utilisation du système d'IA et la situation spécifique du demandeur. Par conséquent, une sensibilisation des juges aux enjeux et limites de la conformité élaborée par le règlement européen sur l'IA est nécessaire afin que les juridictions puissent faire un usage adéquat de ce pouvoir et ne pas se faire dépasser par la complexité technique des preuves en jeu.

CONCLUSION DE LA SECTION II :

En réglementant la question de l'accès aux informations détenues par le fournisseur sur son système d'IA à haut risque, le législateur européen confirme la possibilité d'une réutilisation de ces éléments en tant que preuves lors d'une action en réparation du fait du système d'IA. De plus, l'analyse des dispositions contenues dans la directive européenne relative à la

responsabilité du fait des produits défectueux et la proposition de directive en matière de responsabilité du fait de l'IA met en lumière que cette réutilisation est souhaitée par l'Union européenne. Les mécanismes juridiques mis en place dans ces textes témoignent d'une véritable incitation à employer des informations issues de la phase de conception, en particulier les éléments techniques de preuves de conception, comme preuves dans le contexte d'un préjudice causé lors de l'utilisation d'un système d'IA à haut risque. D'un côté, le demandeur se voit très fortement encouragé à demander au fournisseur du système d'IA, le défendeur, l'accès à ces preuves afin de constituer son dossier. De l'autre côté, les mesures accordées au juge pour répondre à l'asymétrie d'informations entre le demandeur et le défendeur sont limitées à l'accès et la conservation des informations déjà existantes, à savoir les éléments techniques de preuve de conception. Néanmoins, le juge peut prononcer, à son initiative ou sur demande d'une partie, des mesures visant à rechercher de nouvelles preuves alternatives à celles fournies par le fournisseur du système d'IA. Cette possibilité est toutefois envisageable uniquement si le juge n'est pas déjà convaincu par les preuves en sa possession et qu'il considère que les preuves alternatives demandées sont utiles pour rendre sa décision. Par cette évaluation de l'utilité probatoire, le juge devient donc un acteur majeur dans le phénomène de réutilisation des preuves de conception en droit de la responsabilité *ex post*.

CONCLUSION DU CHAPITRE 2 :

Documentation technique, explication de la décision individuelle, journaux d'enregistrement, le règlement européen sur l'IA impose au fournisseur d'un système d'IA à haut risque la constitution de nombreux procédés et éléments de preuves relatifs au fonctionnement du logiciel. Constituées, évaluées et détenues principalement par le fournisseur, ces preuves se révèlent particulièrement intéressantes dans le cadre d'une action en réparation pour un dommage causé par un système d'IA. Le législateur européen apparaît partisan de cette réutilisation dans le contexte de la responsabilité *ex post* en incitant à la fois le demandeur et le juge à s'adresser au fournisseur pour obtenir les éléments de preuves pertinents. Cependant, le juge peut prononcer des mesures d'accès aux preuves uniquement s'il considère ladite mesure et l'élément de preuve recherché comme pertinents. Au travers de cette évaluation de l'utilité probatoire, le juge apparaît alors comme l'acteur central autour duquel se cristallise la question de la réutilisation, ainsi que son ampleur, des éléments techniques de preuve de conception. En plus de prouver la conformité du système d'IA à haut risque, leur réutilisation dans le cadre de la responsabilité leur confère donc une seconde fonction, renforçant d'autant plus l'importance derrière le choix de ces éléments de preuves. En accordant l'accès à des éléments de preuves alternatifs au demandeur, le juge pourrait mettre en lumière les limites derrière le choix de preuves du fournisseur. *A contrario*, en se prononçant uniquement sur les preuves techniques de conception préconstituées, le juge valide indirectement l'interprétation de l'exigence de conformité réalisée par le fournisseur. À ce jour, il est trop tôt pour se prononcer sur l'ampleur de la réutilisation des preuves utilisées pour certifier le système d'IA à haut risque dans le contexte d'une action en réparation. Néanmoins, l'analyse des mécanismes derrière le phénomène de réutilisation permet de confirmer et accroître l'importance du choix des éléments techniques de preuve de conception au sein du règlement européen sur l'IA.

CONCLUSION DU TITRE 1 :

La mise en relation importante entre les exigences de responsabilité *ex ante* et *ex post* du fait de l'IA aboutit à donner aux preuves techniques de conception un rôle considérable dans le contexte d'un dommage causé par un système d'IA à haut risque. Tout d'abord, le respect du règlement européen sur l'IA étant intimement lié à la responsabilité civile du fournisseur d'un système d'IA, les obligations de conception du logiciel délimitent partiellement le périmètre de la responsabilité de cet acteur. Les preuves techniques de conception, en ce qu'elles contribuent à interpréter les obligations de fonctionnement des systèmes d'IA à haut risque, peuvent donc entraîner des conséquences sur la responsabilité *ex post* du fournisseur du logiciel. Ensuite, cette interaction entre les régimes de responsabilité confère aux preuves techniques de conception, choisies en amont, un intérêt particulier dans la recherche des faits constitutifs de responsabilité. Souhaitée par le législateur européen, leur réutilisation dans le contexte d'un dommage causé par un système d'IA s'avère extrêmement probable. Toutefois, si les preuves techniques de conception peuvent entraîner des conséquences majeures dans l'établissement de la responsabilité *ex post*, ces conséquences ne sont en rien certaines. L'adéquation entre les choix des éléments techniques de preuve pertinents pour le règlement sur l'IA et les besoins des régimes de responsabilité dépend de l'appréciation du juge. Ainsi, ce dernier peut décider si les preuves de conception choisies par le fournisseur sont suffisantes pour engager sa responsabilité lors d'un dommage causé par son produit.

En identifiant les conséquences possibles sur les régimes de responsabilité *ex post*, l'analyse menée confirme l'importance centrale des choix des preuves techniques pertinentes utilisées pour attester de la conformité du système d'IA à la législation européenne. Enfin, l'étude révèle que les actions en responsabilité du fait de l'IA constituent un contexte au sein duquel le juge retrouve un rôle majeur dans la détermination des règles substantielles constitutives de la conformité réglementaire des systèmes d'IA. Il aura ainsi la charge de déterminer les limites des effets de la conformité d'un produit aussi unique que le système d'IA, et donc des preuves qui y sont associées, sur la responsabilité de son fournisseur.

TITRE II : LA PREUVE PRODUITE PAR UN SYSTEME D'IA

Introduction. L'étude de la preuve produite par un système d'IA requiert un léger changement de perspective par rapport aux précédentes réflexions. Le terme de « preuve » ne renvoie plus aux éléments techniques utilisés lors de la certification d'un système d'IA à haut risque, conformément au règlement européen sur l'IA. La preuve désigne ici le résultat généré par un système d'IA. Il est ainsi une preuve en justice utilisée pour « *emporter la conviction du juge sur l'exactitude ou l'inexactitude d'une affirmation de fait dans une procédure juridictionnelle* »¹³³⁴. Dans notre contexte, le résultat est caractérisé par son procédé probatoire, à savoir le système d'IA. Dès lors, la notion de preuve produite par un système d'IA exclut les résultats dont l'origine est dissimulée ou omise. Afin d'étudier l'interaction entre le système d'IA et la manière dont son résultat est traité en tant que preuve judiciaire, il est indispensable que l'origine de ladite preuve soit connue des différents acteurs au procès, en particulier du juge. Par exemple, un hypertrucage (*deepfake* en anglais) produit à l'insu du juge n'entre pas dans le périmètre de l'étude. Par la mise en place d'un processus de certification des systèmes d'IA à haut risque, le règlement européen sur l'IA offre une garantie d'une certaine qualité de fonctionnement de ces logiciels. Fondée sur des éléments techniques de preuves, cette assurance de qualité peut alors avoir des conséquences sur le traitement du résultat par l'utilisateur du logiciel. L'identification des interactions entre la conception du système d'IA et la réception de ses résultats contribue à enrichir l'analyse du rôle des éléments techniques de preuves, choisis dans le cadre du règlement sur l'IA, lors de la phase de déploiement du logiciel. Pour cela, notre étude se concentrera sur l'articulation entre les caractéristiques techniques du système d'IA et le traitement judiciaire de son résultat au regard du droit de la preuve.

Annonce du plan. De façon traditionnelle, le résultat produit par un système d'IA doit être admis par le juge (**Chapitre 1**) avant de pouvoir avoir une quelconque influence sur la conviction du juge lors de la phase d'appréciation (**Chapitre 2**).

¹³³⁴ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

CHAPITRE 1 : L'ADMISSIBILITE DE LA PREUVE PRODUITE PAR UN SYSTEME D'IA

Introduction. En droit de la preuve, la notion d'admissibilité de la preuve se divise entre l'admissibilité restreinte d'une preuve, d'une part, et sa recevabilité, d'autre part. « *Hors les cas où la loi en dispose autrement, la preuve peut être apportée par tout moyen.* »¹³³⁵. Tout d'abord, le système de preuve en droit français est mixte, c'est-à-dire qu'il est composé d'un système de preuve légale, au sein duquel l'admissibilité est restreinte à certains modes de preuve explicitement énoncés par la loi, et d'un système de preuve libre. La qualification d'un procédé probatoire parmi les modes de preuves existants permet donc de délimiter son admissibilité restreinte au regard de ces systèmes de preuves. Ensuite, le second concept constitutif de l'admissibilité d'une preuve est contenue au sein de la notion de licéité de la preuve. La licéité d'une preuve détermine sa recevabilité, à savoir son examen ou son exclusion par le juge. Bien qu'intéressante, la réalisation d'une analyse exhaustive des conditions de recevabilité des résultats produits par un système d'IA est une possibilité qui doit être écartée. En effet, la licéité d'une preuve dépend de très nombreux critères, notamment la manière dont la preuve a été obtenue, qui ne dépendent pas de la nature ou du fonctionnement du procédé probatoire utilisé. Il convient donc de se limiter dans notre approche aux conséquences du règlement européen sur l'IA sur l'admissibilité des résultats produits par les systèmes d'IA à haut risque. À l'image du titre précédent, il s'agit alors d'étudier l'existence d'interactions entre la nouvelle réglementation sur les systèmes d'IA et les conditions d'admissibilité de leurs résultats.

Annnonce de plan. En tant que nouveau procédé probatoire, le résultat produit par un système d'IA doit donc faire l'objet d'une qualification juridique afin, d'une part, d'évaluer son admissibilité au sein du régime de preuve légale et, d'autre part, de lui associer un régime juridique existant (**Section I**). Condition essentielle de la recevabilité des éléments de preuve, la licéité des résultats produits par un système d'IA doit être aussi examinée afin d'identifier les interactions que la licéité pourrait entretenir avec le règlement européen sur l'IA (**Section II**).

¹³³⁵ Article 1358 du code civil.

Section I : La qualification de la preuve produite par un système d'IA

Annonce du plan. Opération essentielle de l'analyse juridique, la qualification juridique de la preuve produite par un système d'IA est une première étape indispensable à l'étude de son admissibilité. En tant que nouveau procédé probatoire, le système d'IA doit être inscrit dans une des catégories juridiques existantes en droit civil. La réalisation d'une catégorisation rigoureuse de ce mode de preuve permettra d'abord d'évaluer l'application des règles d'admissibilité restreinte de la preuve produite par un système d'IA au sein du système de preuve légale. Ensuite, le processus de qualification facilitera la réalisation d'analogies davantage pertinentes lors de l'analyse. La preuve produite par un système d'IA se voit écartée de la plupart des catégories de modes de preuves (**Sous-section 1**) pour être finalement qualifiée de preuve indiciaire (**Sous-section 2**).

Sous-section 1 : Les catégories de modes de preuves écartées

Annonce de plan. Le résultat produit par un système d'IA ne peut pas se voir assimiler à une preuve par déclaration (**§1**). De même, il semble difficile pour un résultat de systèmes d'IA d'être rattaché à la catégorie de l'écrit mais cela n'en demeure pas moins envisageable dans certaines situations (**§2**).

§1. La qualification comme preuve par déclaration logiquement exclue

Annonce de plan. L'expression « preuve par déclaration » est empruntée ici à Etienne VERGES, Géraldine VIAL et Olivier LECLERC dans leur ouvrage intitulé *Droit de la preuve*¹³³⁶. Cette expression regroupe toutes les catégories de modes de preuves produites par déclaration, que celles-ci soient produites par les parties (l'aveu et le serment), par des tiers (le témoignage) ou par des professionnels spécialisés. Après un bref rappel des principaux modes

¹³³⁶ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. Il s'agit de l'intitulé du titre II de la partie 3 de l'ouvrage.

de preuves visés **(A)**, il sera démontré que la preuve produite par le système d'IA ne peut en aucun cas prétendre à ce groupement de modes de preuves **(B)**.

A- La définition des principaux modes de preuves par déclaration

Le témoignage. Le témoignage est l'un des trois modes de preuves par déclaration contenus dans le code civil. Pour autant, l'unique article le concernant (article 1381 du code civil) n'en fournit aucune définition à proprement parler, indiquant seulement qu'il s'agit d'une « *déclaration faite par un tiers* »¹³³⁷ puis renvoyant aux dispositions du code de procédure civile en ce qui concerne sa valeur probante. Les articles 199 et suivants du code de procédure civile auxquels le texte renvoie n'offrent pas non plus de définition du témoignage. Par conséquent, la définition doit être cherchée du côté de la doctrine qui définit le témoignage comme « *la preuve résultant de déclarations de personnes qui relatent ce qu'elles ont vu ou entendu, soit par hasard soit sur l'invitation de l'une des parties* »¹³³⁸.

L'aveu. Le code civil définit l'aveu comme « *la déclaration par laquelle une personne reconnaît un fait de nature à produire contre elle des conséquences juridiques* »¹³³⁹. Selon qu'il soit ou non prononcé devant le juge du fond compétent lors de l'instance en cours, il peut être qualifié d'aveu judiciaire¹³⁴⁰ ou d'aveu extrajudiciaire¹³⁴¹. Contrairement au témoignage, l'aveu est une déclaration produite par une partie au procès et non par un tiers.

Le serment. Dernier mode de preuve nommé par le code civil, le serment est « *la déclaration, réalisée solennellement par un plaideur devant le juge, afin d'affirmer un fait qui est favorable à sa prétention* »¹³⁴². Dans son contenu, le serment se distingue de l'aveu en ce qu'il profite à la situation de son auteur tandis que l'aveu est nécessairement contre lui.

¹³³⁷ « *La valeur probante des déclarations faites par un tiers dans les conditions du code de procédure civile est laissée à l'appréciation du juge.* » - Article 1381 du code civil

¹³³⁸ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.*

¹³³⁹ Article 1382 du code civil.

¹³⁴⁰ « *L'aveu judiciaire est la déclaration que fait en justice la partie ou son représentant spécialement mandaté.* » - Article 1383-2 du code civil.

¹³⁴¹ Non défini par le code civil, l'aveu extrajudiciaire désigne simplement l'aveu qui ne satisfait pas les conditions de l'aveu judiciaire.

¹³⁴² LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 585.

B- L'absence rédhibitoire de personnalité juridique du système d'IA

Des déclarations produites par une personne. Les régimes probatoires du témoignage, de l'aveu ou encore du serment sont bien plus riches que les brèves descriptions réalisées ci-dessous. De même, ces trois modes de preuve ne constituent pas la totalité des modes de preuves par déclaration. Elles servent ici à mettre en lumière une condition de validité commune à ce type de mode de preuves : leur auteur doit être une personne physique. Selon le mode de preuve choisi, l'auteur doit satisfaire plusieurs autres conditions mais la première d'entre elles est de bénéficier de la personnalité juridique.

L'impossible qualification du résultat de systèmes d'IA comme preuve par déclaration.

L'absence de personnalité juridique est donc rédhibitoire pour bénéficier de cette qualification. À la lecture de cette première condition, il apparaît clairement une impossibilité totale à ce qu'un résultat de systèmes d'IA puisse être qualifié de preuve par déclaration. Par conséquent, il n'est pas nécessaire de s'interroger sur les autres conditions de qualification de ces modes de preuves, toutes étant liées à celle de la personnalité juridique de l'auteur de la déclaration.

§2. La catégorie de l'écrit électronique difficilement envisageable

Annonce de plan. La preuve par écrit, quant à elle, est définie par le code civil comme « *une suite de lettres, de caractères, de chiffres ou de tous autres signes ou symboles dotés d'une signification intelligible, quel que soit leur support.* »¹³⁴³. Parmi la pluralité de systèmes d'IA développés à ce jour, il existe une catégorie de logiciels spécifiquement conçus dans le but de générer, avec divers niveaux d'autonomie, du contenu tel que du texte (ChaptGPT en figure de proue), des images (Midjourney, DALL-E entre autres) ou d'autres médias (audios ou vidéos) en réponse à des instructions données en entrée par l'utilisateur¹³⁴⁴. Dès lors, les résultats produits par ces systèmes d'IA génératifs peuvent se rapprocher d'une preuve littérale au sens du droit civil. Au regard de la nature des technologies d'IA, l'écrit généré sera nécessairement un écrit électronique. Il est alors possible de distinguer deux types de preuves par écrit en

¹³⁴³ Article 1365 du code civil, modifié par l'article 4 de l'« Ordonnance n° 2016-131 du 10 février 2016 portant réforme du droit des contrats, du régime général et de la preuve des obligations », 2016.

¹³⁴⁴ « *Foundation models specifically intended to be used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, Or, video ("generative AI")* » - « Règlement sur l'intelligence artificielle », 2024. Version du compromis du parlement européen de mai 2023. Article 28b.5a.

fonction de la finalité du système d'IA. Tout d'abord, les systèmes d'IA génératifs peuvent générer un résultat s'apparentant à un écrit **(A)** Ils sont aussi capables de reproduire un écrit existant sous forme électronique, le résultat peut alors être rapproché de la copie électronique **(B)**.

A- L'écrit produit par un système d'IA

Typologie des preuves littérales. Dépendamment de ses caractéristiques, la preuve littérale peut revêtir différentes formes. Dans le système de preuve légale, le législateur français envisage trois catégories principales de preuves par écrit : l'acte authentique, l'acte sous signature privée et le commencement de preuve par écrit. Pour être qualifié dans l'une de ces trois catégories, l'écrit électronique que produirait un système d'IA génératif doit satisfaire les mêmes conditions que l'écrit papier mais également offrir des garanties d'imputabilité et d'intégrité¹³⁴⁵.

L'impossibilité de produire un écrit authentique. « *L'acte authentique est celui qui a été reçu, avec les solennités requises, par un officier public ayant compétence et qualité pour instrumenter.* »¹³⁴⁶. L'acte authentique se caractérise donc par des conditions d'authenticité, relatives à l'auteur de l'acte, d'une part, et les conditions relatives à la rédaction de l'acte (« *les solennités requises* »), de l'autre. Pour être authentique, l'acte doit tout d'abord avoir été rédigé par un officier public. Instantanément, l'hypothèse où un système d'IA produirait un acte authentique devient impossible. En effet, le système d'IA ne dispose pas de la qualité juridique d'une personne et encore moins celle d'officier public. Cette impossibilité est d'autant plus importante que le rôle de l'officier public ne se limite pas à la rédaction de l'acte. Il doit également¹³⁴⁷ approuver toute rature réalisée sur l'écrit¹³⁴⁸ ainsi qu'apposer sa signature. L'acte

¹³⁴⁵ « *L'écrit électronique a la même force probante que l'écrit sur support papier, sous réserve que puisse être dûment identifiée la personne dont il émane et qu'il soit établi et conservé dans des conditions de nature à en garantir l'intégrité.* » - Article 1366 du code civil.

¹³⁴⁶ Article 1369 du code civil.

¹³⁴⁷ Des conditions d'authenticité liées à la rédaction de l'écrit s'ajoutent selon la catégorie d'officier public en présence (par exemple, les conditions spécifiques à l'huissier sont énoncées dans la *Loi n° 92-644 du 13 juillet 1992 modifiant la loi n° 91-650 du 9 juillet 1991 portant réforme des procédures civiles d'exécution et l'ordonnance n° 45-2592 du 2 novembre 1945 relative au statut des huissiers et comportant diverses dispositions relatives aux procédures civiles d'exécution*). Le système d'IA ne pouvant être un officier public, ces conditions ne peuvent a fortiori pas être satisfaites.

¹³⁴⁸ « *Attendu d'autre part que si la signature de l'huissier au bas de l'acte suffit à authentifier celui-ci, cet effet ne saurait concerner les ratures non approuvées de l'exploit ;* » - Cour de Cassation, Chambre civile 2, 20 juin 1968.

doit ensuite être évidemment signé par chacune des parties. Le manquement de ces conditions entraînant la nullité absolue de l'acte¹³⁴⁹, il est impossible qu'un résultat de système d'IA puisse être qualifié d'acte authentique.

L'écrit sous signature privée. Contrairement à l'écrit authentique, l'écrit sous signature privée ne connaît pas de restrictions en ce qui concerne son procédé d'écriture. L'écrit sous signature privée peut être rédigé par les parties mais aussi par un tiers¹³⁵⁰, au crayon¹³⁵¹ ou même à partir de formules pré-imprimées¹³⁵². Par conséquent, cette absence de condition est un obstacle de moins pour retenir cette qualification pour le résultat d'un système d'IA.

Toutefois, tous les écrits sous signature privée satisfont une condition indispensable¹³⁵³ à leur validité, à savoir la signature de l'écrit par les parties¹³⁵⁴. L'absence de personnalité juridique des systèmes d'IA ne permet pas de les considérer comme signataires de l'acte. Par conséquent, le résultat généré par le système d'IA se révèle imparfait et ne peut obtenir la qualification d'écrit sous signature privée¹³⁵⁵.

La catégorie de commencement de preuve par écrit. Selon le code civil, le commencement de preuve par écrit est défini comme suivant :

« Constitue un commencement de preuve par écrit tout écrit qui, émanant de celui qui conteste un acte ou de celui qu'il représente, rend vraisemblable ce qui est allégué.

Peuvent être considérés par le juge comme équivalant à un commencement de preuve par écrit les déclarations faites par une partie lors de sa comparution personnelle, son refus de répondre ou son absence à la comparution.

¹³⁴⁹ Sur l'absence de signature de l'une des parties : « Qu'en statuant ainsi, alors que l'acte de vente passé le 5 juillet 1975 n'était pas signé des époux X...-D..., parties à l'acte comme acquéreurs indivis, de sorte que cet acte, comme celui qui le rectifie, étant nul, de nullité absolue, la cour d'appel a violé les textes susvisés ; » - Cour de cassation, Chambre civile 1, 12 juillet 2007, n° 06-10.362.

¹³⁵⁰ FERRAND Frédérique, « Le juge et la preuve écrite », *Répertoire de procédure civile*, Dalloz, 2023.

¹³⁵¹ Cour de Cassation, Chambre commerciale, 8 octobre 1996, n° 94-17.96.

¹³⁵² Cour de Cassation, Chambre civile 1, 21 février 2006, n° 04-13.51.

¹³⁵³ Certains écrits sous signature privée se voient imposée des conditions de validité spécifique supplémentaire à celle de la signature des parties. Ainsi, les contrats synallagmatiques doivent être faits « en autant d'originaux qu'il y a de parties ayant un intérêt distinct, à moins que les parties ne soient convenues de remettre à un tiers l'unique exemplaire dressé » (article 1375 alinéa 1 du code civil).

¹³⁵⁴ « en dehors des exceptions prévues par la loi, l'acte sous seing privé n'est soumis à aucune autre condition de forme que la signature de ceux qui s'obligent » - Cour de Cassation, Chambre sociale, 14 mars 2000, n° 97-45.899
Cour de cassation, Chambre civile 3, 1 mars 2023, n° 21-25.497.

¹³⁵⁵ Cour de Cassation, Chambre civile 1, 2 juillet 1996, n° 94-14.943.

La mention d'un écrit authentique ou sous signature privée sur un registre public vaut commencement de preuve par écrit. »¹³⁵⁶

La formulation de la catégorie de « commencement de preuve par écrit » se révèle particulièrement maladroite par rapport à ce qu'elle désigne réellement. Ainsi le commencement de preuve par écrit n'est pas en réalité un écrit au sens de la preuve, c'est-à-dire qu'il n'est ni un écrit authentique ni un écrit sous signature privée¹³⁵⁷. Ainsi l'élément de preuve n'a pas besoin d'être signé pour être qualifié de commencement de preuve par écrit¹³⁵⁸. Le commencement de preuve par écrit est davantage un mécanisme juridique¹³⁵⁹ actionnable par divers autres modes de preuves¹³⁶⁰ qu'un mode de preuve à proprement parlé.

Néanmoins, la qualification de commencement de preuve par écrit nécessite la satisfaction de deux conditions. La première condition est une condition de forme, à savoir que le commencement de preuve par écrit doit émaner « *de celui qui conteste un acte ou de celui qu'il représente* »¹³⁶¹. La seconde condition vise le contenu du mode de preuve qui doit « *rendre vraisemblable le fait allégué* »¹³⁶². L'analyse se concentrera sur la condition d'imputabilité car la condition relative au contenu ne présente aucun lien avec le procédé probatoire du commencement de preuve par écrit.

L'imputabilité du résultat. Dans le cadre de la preuve produite par un système d'IA, la problématique repose sur l'imputabilité de ladite preuve à son auteur. Cette imputabilité peut être matérielle et/ou intellectuelle¹³⁶³. Par exemple, la signature d'un écrit entraîne l'imputabilité matérielle de cet écrit au signataire. Lorsque la personne a rédigé elle-même l'élément de preuve, l'imputabilité est à la fois matérielle et intellectuelle. L'imputabilité matérielle d'un élément de preuve ne pose alors pas de réelles difficultés¹³⁶⁴.

En revanche, qu'en est-il de l'élément de preuve qui n'aurait été ni rédigé ni signé par la personne à qui on l'oppose ? Quid du cas où l'action de cette personne s'est limitée à l'entrée d'instructions dans un système d'IA générateur d'actes ? Est-il possible de considérer que

¹³⁵⁶ Article 1362 du code civil.

¹³⁵⁷ LARDEUX Gwendoline, « Preuve : modes de preuve », *Répertoire de droit civil*, 2023.

¹³⁵⁸ Cour de Cassation, Chambre civile 1, 27 janvier 1971, n° 69-13.273.

¹³⁵⁹ Pour l'origine de l'expression « mécanisme juridique », voir LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

¹³⁶⁰ L'article 1362 intègre dans cette notion certaines déclarations des parties, un refus de répondre ou encore une mention d'un acte dans un registre public.

¹³⁶¹ Article 1362 du code civil.

¹³⁶² Article 1362 du code civil.

¹³⁶³ LARDEUX Gwendoline, « Preuve : modes de preuve », *Répertoire de droit civil*, 2023.

¹³⁶⁴ *Ibid.*

l'écrit rédigé automatiquement exprime la volonté de l'utilisateur du logiciel ? À moins que la personne n'ait approuvé l'élément de preuve *a posteriori*¹³⁶⁵, l'imputabilité matérielle ne peut pas être caractérisée dans cette situation. Seul l'imputabilité intellectuelle peut être envisagée, c'est-à-dire que le texte est bien représentatif de la volonté de la personne contre laquelle il est invoqué, sans pour autant qu'elle l'ait rédigé. Cette mise en situation pose une double difficulté. Dans un premier temps, la jurisprudence relative à l'imputabilité du commencement de preuve par écrit est ancienne et rare, par conséquent peu significative¹³⁶⁶. Dans un second temps, l'action de rédaction n'est pas déléguée à une personne tierce mais à une machine qui dispose d'un degré important d'autonomie et d'imprévisibilité¹³⁶⁷. Ainsi, là où il aurait été envisageable d'entendre la déclaration détaillée du tiers qui se serait vu dicter le texte, le logiciel d'IA se révèle plus avare en informations sur la participation de son utilisateur. En effet, le dépoyeur du système d'IA n'a pas dicté mot à mot le texte à rédiger mais a seulement entré des instructions dans le système. Alors, est-il possible de déduire l'imputabilité intellectuelle d'un élément de preuve généré par un système d'IA du seul fait des instructions qu'il a entré ? L'action de soumission d'une requête à un système d'IA génératif peut-elle permettre de considérer que le résultat généré émane du dépoyeur ? À notre sens, le caractère imprévisible du fonctionnement de systèmes d'IA, en particulier les technologies utilisées pour les systèmes d'IA génératifs, ne permet pas de répondre par l'affirmative. Par conséquent, le seul texte généré par un système d'IA ne pourra pas satisfaire la condition d'imputabilité sans l'ajout d'autres éléments qui caractérisaient l'imputabilité du dépoyeur¹³⁶⁸. Ainsi, la requête de du dépoyeur, si elle est suffisamment claire et détaillée, pourrait permettre d'envisager la satisfaction de la condition d'imputabilité. Cependant, dans cette situation, il apparaît alors préférable pour la partie d'apporter la requête en tant que commencement de preuve par écrit car son imputabilité au dépoyeur est davantage caractérisée.

B- La copie électronique produite par un système d'IA

Exemple illustratif. L'analyse d'image est une des nombreuses possibilités offertes par les technologies d'IA génératives. Depuis plusieurs années déjà, des systèmes d'IA sont capables

¹³⁶⁵ Cour de Cassation, Chambre civile 1, 24 novembre 1969.

¹³⁶⁶ LARDEUX Gwendoline, *op. cit.*

¹³⁶⁷ Sur la différence entre l'automatisation et la délégation du processus de production du résultat, voir notamment SARANGI Saswat et SHARMA Pankaj, *Artificial Intelligence*, Routledge India, 2018.

¹³⁶⁸ Il peut s'agir par exemple d'une signature ou d'une reconnaissance expresse de l'utilisateur du système d'IA que l'acte généré est cohérent avec sa volonté.

d'analyser une image et d'en produire un texte¹³⁶⁹, qu'il s'agisse d'une description détaillée plus ou moins fidèle¹³⁷⁰ ou de mots-clés à ajouter en description sur les réseaux sociaux¹³⁷¹. En ce qui nous concerne, cette fonctionnalité peut se révéler utile pour lire les documents rédigés à la main par des personnes et en produire des versions numérisées. Le fichier texte produit ressemble à n'importe quel document rédigé directement sur un ordinateur. À la différence d'une image produite par un scan, cette action permettrait au lecteur de réaliser des recherches ciblées de mots mais aussi d'effectuer de l'analyse de données, le format texte étant bien plus adapté à cela qu'une image d'un texte écrit à la main.

La copie produite par un système d'IA. La copie est une reproduction littérale d'un écrit, dit écrit original, sur papier ou support électronique¹³⁷². La fonction d'une copie est de permettre de multiplier les exemplaires d'un écrit sans pour autant qu'il en perde son utilité, à savoir sa fonction d'instrument de preuve. Par conséquent, la copie n'a d'intérêt que si elle conserve la force probante de l'original. Pour parvenir à cet objectif de force probante identique, la copie, dans notre cas électronique, doit satisfaire depuis 2016¹³⁷³ à la seule condition de fiabilité lors de sa création¹³⁷⁴. Cette condition s'applique à la fois pour les copies d'actes authentiques et pour les copies d'écrit sous signature privée. À défaut de donner une définition de la notion de fiabilité, le législateur français confie l'évaluation de la fiabilité à l'appréciation du juge et met en place deux mécanismes de présomptions de fiabilité : une présomption simple pour la copie de l'écrit sous signature privée¹³⁷⁵ et une présomption irréfragable pour la copie de l'écrit authentique¹³⁷⁶.

Les copies électroniques d'écrit authentique bénéficiant d'une présomption irréfragable de fiabilité. La fiabilité d'une telle copie est présumée dans deux situations. Premièrement, la copie authentique d'un acte authentique bénéficie elle aussi d'une présomption de fiabilité. Au-delà de l'utilisation d'un système de numérisation garantissant une reproduction à l'identique, l'officier public doit y apposer sa signature qualifiée, son sceau, la date de la copie et une

¹³⁶⁹ Il s'agit d'une technologie appelée OCR pour *Optical character recognition* (reconnaissance optique de caractères en français).

¹³⁷⁰ Voir notamment *Gradio's InkyMM* (<https://mm-frontend-ees4nzlylh8.octoai.cloud/>)

¹³⁷¹ Par exemple : <https://www.captionit.ai/>

¹³⁷² CORNU Gérard, *Vocabulaire juridique*, 2022.

¹³⁷³ « Ordonnance n° 2016-131 du 10 février 2016 portant réforme du droit des contrats, du régime général et de la preuve des obligations », 2016, p. 201. L'ordonnance a simplifié le régime juridique de la copie d'un écrit qui se divisait alors selon qu'il s'agisse d'un écrit authentique ou d'un écrit sous signature privée. De plus, une distinction s'opérait encore selon que l'écrit original existe ou ait disparu)

¹³⁷⁴ Article 1379 du code civil.

¹³⁷⁵ Article 1379 al. 2

¹³⁷⁶ Article 1379 al. 1

mention de la conformité de la copie électronique par rapport à l'original¹³⁷⁷. Deuxièmement, les copies exécutoires (aussi appelée « grosse ») sont des « *copies authentiques qui se terminent par la même formule que les jugements des tribunaux [...]* »¹³⁷⁸. Inscrite par un officier public, la formule exécutoire est un élément essentiel pour fournir la force exécutoire à tous les actes susceptibles d'exécution forcée (arrêts, jugements, mandats de justice, etc.). Cette présomption de fiabilité est considérée comme irréfragable car elle contraint la personne s'opposant à sa fiabilité à contester la copie au travers d'une procédure d'inscription en faux, pareillement à l'original¹³⁷⁹. À nouveau, l'obligation que l'auteur de la copie soit un officier public limite législativement mais aussi pratiquement le recours à des systèmes d'IA pour ce genre de tâche. Au-delà de la présomption cependant, il est toujours possible d'apporter la preuve de la fiabilité d'une copie électronique d'un écrit afin de bénéficier d'une force probante identique à l'écrit original.

Fiabilité et présomption simple de fiabilité de la copie. La condition de fiabilité est commune à la copie de l'écrit authentique et à la copie de l'acte sous signature privée. La copie électronique d'un écrit est fiable à deux conditions. Premièrement, il doit s'agir d'une reproduction identique sur la forme et le contenu de l'acte. Deuxièmement, l'intégrité de la copie est garantie dans le temps selon des procédés énoncés par décret en Conseil d'Etat¹³⁸⁰.

L'intégrité de la copie électronique produite par un système d'IA dans le temps. En ce qui concerne la condition d'intégrité de la copie dans le temps, l'utilisation d'un système d'IA pour générer la copie d'un écrit ne pose pas de problématiques d'altération particulières par rapport à d'autres technologies de copies électroniques (scan, capture d'écran, etc.). Ainsi, les conditions de conservation, l'accès aux appareils de copie¹³⁸¹ ainsi que les potentielles solutions technologiques applicables¹³⁸² (horodatage, cachet électronique ou signature électronique¹³⁸³)

¹³⁷⁷ « Décret n°71-941 du 26 novembre 1971 relatif aux actes établis par les notaires », 71-941, 1971. Article 37.

¹³⁷⁸ *Ibid.* Article 33 al. 2. La formule exécutoire est énoncée à l'article 1 du « Décret n°47-1047 du 12 juin 1947 relatif à la formule exécutoire. », 47-1047, 1947.

¹³⁷⁹ Article 1371 du code civil.

¹³⁸⁰ Article 1379 du code civil.

¹³⁸¹ « Décret n° 2016-1673 du 5 décembre 2016 relatif à la fiabilité des copies et pris pour l'application de l'article 1379 du code civil », 2016-1673, 2016. Articles 2 à 6.

¹³⁸² « *L'intégrité de la copie résultant d'un procédé de reproduction par voie électronique est attestée par une empreinte électronique qui garantit que toute modification ultérieure de la copie à laquelle elle est attachée est détectable.* » - *Ibid.* Article 3 al. 1.

¹³⁸³ « Règlement (UE) n° 910/2014 du Parlement européen et du Conseil du 23 juillet 2014 sur l'identification électronique et les services de confiance pour les transactions électroniques au sein du marché intérieur et abrogeant la directive 1999/93/CE », OJ L, 2014. Le recours à ces techniques à un niveau dit « qualifié » offre une présomption d'intégrité à la copie électronique.

le sont tout autant pour le résultat produit par un système d'IA que pour un autre procédé de copie électronique.

La double exigence de fidélité de la copie électronique. La condition de reproduction identique est évaluée sous le prisme de la fidélité du contenu et de la forme¹³⁸⁴. Sur le premier critère, le recours à un système d'IA pour réaliser une copie identique apparaît problématique car le procédé se fonde sur des méthodes statistiques qui contiennent de façon intrinsèque un taux d'erreur. Dès lors, même un système d'IA extrêmement précis ne pourrait garantir une fidélité totale concernant le contenu de la copie. La difficulté est d'autant plus grande que dans le cas d'un écrit manuscrit, la qualité d'écriture de l'auteur ou encore la variété des symboles utilisés (équations, alphabets différents, etc.) peuvent affecter l'exactitude du système¹³⁸⁵. La capacité d'un système d'IA à garantir la fidélité du contenu d'un écrit original est encore indéterminée. Sur ce point, ce sera au juge de trancher si cet aspect de la fidélité est satisfait ou non comme il l'a déjà fait pour la photocopie¹³⁸⁶ ou encore le fax¹³⁸⁷.

Ensuite, la copie électronique réalisée par un système d'IA doit être identique sur la forme. Pour reprendre les exemples précédents, à savoir la photocopie, le fax ou encore le scanner, ils reproduisent identiquement la forme de l'original. En effet, un scan d'une missive ressemblera à ladite lettre. Une photocopie d'un écrit rédigé manuellement conservera le style manuscrit de ses auteurs ainsi que leur signature inscrite en bas du document. À l'inverse, la copie électronique générée par un système d'IA transforme la forme manuscrite de l'original en un document texte parfaitement banal. En ce qui concerne la reproduction de la signature, les systèmes d'IA actuels pourraient avoir la capacité de l'identifier et de la reproduire parfaitement. Outre le contenu qui peut avoir fait l'objet d'erreur de reproduction, la forme sera amenée à changer¹³⁸⁸. Une suite d'échange épistolaires ne sera qu'un enchaînement de textes les uns à la suite des autres sur un fichier informatique. Ainsi, il semble difficile d'imaginer que le juge qualifiera cette transformation formelle comme fidèle. Par conséquent, il nous semble peu probable qu'un résultat généré par un système d'IA puisse être qualifié de copie électronique au sens du code civil.

¹³⁸⁴ Article 1379 du code civil.

¹³⁸⁵ « Handwriting Recognition in 2023 » sur *Research - Multiplie AI*.

¹³⁸⁶ Cour de Cassation, Chambre civile 1, 9 mai 1996, n° 94-13.310. Cour de Cassation, Chambre civile 1, 25 juin 1996, n° 94-11.745.

¹³⁸⁷ LAUTRETTE Laurence, « Telex, télécopie, télégramme : valeur juridique et force probante », *Les Petites Affiches*, 1996.

¹³⁸⁸ L'entraînement d'un système d'IA pour réaliser une simple photocopie ne présente aucun intérêt pratique.

Sous-section 2 : La qualification du résultat d'un système d'IA comme preuve indiciaire

Annonce de plan. L'aveu, le serment, le témoignage et l'écrit étant désormais écartés, le dernier mode de preuve civile nommé à analyser est celui de la présomption judiciaire. Considérée comme une catégorie de mode de preuve intrinsèquement ouverte, la présomption judiciaire est une qualification traditionnelle pour les nouveaux éléments de preuves produits par des avancées scientifiques¹³⁸⁹. Cependant, ce n'est pas l'expression de « présomption judiciaire » qui a été choisie pour le présent paragraphe mais celle de « preuve indiciaire ». Ce choix différent de celui du législateur français impose dans un premier temps de définir le terme de « preuve indiciaire » et de présenter ces différences par rapport aux notions proches (§1). Il est ensuite important de mettre en avant les critères de qualification de la preuve indiciaire dans sa fonction probatoire (§2). Enfin, il s'agira de voir dans quelle mesure le résultat produit par un système d'IA s'intègre dans cette catégorie de preuve indiciaire (§3).

§1. Le refus de l'expression « présomption judiciaire »

Une présomption non légale. La présomption judiciaire est définie par l'unique article 1382 du code civil comme :

*« Les présomptions qui ne sont pas établies par la loi, sont laissées à l'appréciation du juge, qui ne doit les admettre que si elles sont graves, précises et concordantes, et dans les cas seulement où la loi admet la preuve par tout moyen. »*¹³⁹⁰

Les présomptions judiciaires sont donc définies dans un premier temps en opposition avec les preuves établies par la loi¹³⁹¹. Appelées « présomptions légales » ou « présomptions de droit », ces dernières sont désignées comme des dispenses de preuve¹³⁹². En réalité, il ne s'agit pas d'une dispense de la charge de la preuve pour la partie en question mais d'un déplacement de

¹³⁸⁹ « Les innovations technologiques et scientifiques ont également conduit à l'extension du domaine de la preuve indiciaire. Le caractère intrinsèquement ouvert de ce mode de preuve a permis d'admettre, à titre d'indice pouvant constituer une présomption, tous les procédés de captation d'images et de sons [...] » - BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010. P. 180.

¹³⁹⁰ Article 1382 du code civil.

¹³⁹¹ « La présomption que la loi attache à certains actes ou à certains faits en les tenant pour certains dispense celui au profit duquel elle existe d'en rapporter la preuve. » - Article 1354 du code civil.

¹³⁹² *Ibid.*

l'objet de la preuve. Ce déplacement de l'objet de la preuve s'impose au juge qui ne peut s'opposer à l'admissibilité du fait présumé si la partie apporte la preuve conforme du fait établi (l'objet déplacé de la preuve).

L'absence de pertinence du critère de l'auteur de la présomption. Ainsi, les présomptions judiciaires sont les présomptions établies non pas par le législateur mais par les juges eux-mêmes. Selon le code civil, l'auteur de la présomption est le critère déterminant pour catégoriser et qualifier les présomptions juridiques existantes. Paradoxalement, cela ne signifie pas que le mécanisme de la présomption est l'apanage du législateur. En 2012, la Cour de cassation publie dans son rapport annuel consacré à la preuve que les présomptions peuvent émaner aussi bien de la loi que de la jurisprudence¹³⁹³. En pratique, cette affirmation s'illustre à de très nombreuses reprises dans les décisions de la Cour de cassation. Ainsi en 2014, la Haute juridiction énonce que, dans le cadre d'un contrat de dépôt, la détérioration d'un véhicule confié à un garagiste implique une présomption de faute dudit gardien et une présomption de lien de causalité entre cette faute et ladite détérioration¹³⁹⁴. Plus récemment, les juges de la Cour de cassation indiquent que « *l'atteinte portée par un chirurgien à un organe ou un tissu que son intervention n'impliquait pas, est fautive [...]* »¹³⁹⁵, posant alors une présomption simple de faute¹³⁹⁶. Bien qu'elles soient d'origine prétorienne, ces présomptions entraînent elles aussi un déplacement de l'objet de la preuve et une réduction du pouvoir d'appréciation du juge. Par conséquent, la distinction entre les présomptions légales et les présomptions judiciaires en raison de leur auteur n'est pas pertinente pour caractériser le mode de preuve qu'est la présomption judiciaire.

Le choix du terme « présomption de fait ». Qu'elles soient issues du législateur ou des juges, les présomptions présentées précédemment présentent des points communs caractéristiques des présomptions de droit. Celles-ci sont définies comme :

*« Un lien général et abstrait entre les deux termes du raisonnement probatoire. La preuve du fait connu entraîne automatiquement celle du fait inconnu et le juge est lié par cette relation de droit entre les deux faits. »*¹³⁹⁷

¹³⁹³ COUR DE CASSATION, *La preuve dans la jurisprudence de la Cour de cassation*, la Documentation française, 2013.

¹³⁹⁴ Cour de cassation, Chambre civile 1, 5 février 2014, n° 12-23.467.

¹³⁹⁵ Cour de cassation, Chambre civile 1, 26 février 2020, n° 19-13.423 19-14.240.

¹³⁹⁶ La présomption de faute peut être renversée dès lors que le chirurgien rapporte la preuve « *d'une anomalie rendant l'atteinte inévitable ou de la survenance d'un risque inhérent à cette intervention qui, ne pouvant être maîtrisé, relève de l'aléa thérapeutique.* » - *Ibid.*

¹³⁹⁷ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. PP. 255-256.

Au sens du droit civil, la description de la présomption judiciaire ne correspond pas à cette définition de la présomption de droit. Le législateur y indique que ces présomptions sont élaborées par le juge à condition qu'elles soient « *graves, précises et concordantes, et dans les cas où la loi admet la preuve par tout moyen* »¹³⁹⁸. Au travers de ces conditions, le législateur vise le raisonnement probatoire des juges sur les éléments de preuve qui leurs sont fournis par les parties. Premièrement, la présomption étant élaborée par le juge lui-même pour un cas d'espèce précis, ce dernier ne peut pas être considéré comme lié par sa propre présomption de fait. Deuxièmement, la présomption de fait n'entraîne pas de modification de la charge de la preuve selon la CJUE¹³⁹⁹. En matière de responsabilité du fait des produits défectueux dans le domaine médical, les juges de la Cour de cassation ont présumé le lien de causalité entre le vaccin contre l'hépatite B et la sclérose en plaques des patients à partir de certains indices¹⁴⁰⁰ et ont refusé cette présomption dans d'autres cas¹⁴⁰¹. La présomption de fait est donc un mécanisme probatoire que le juge applique au travers de son raisonnement¹⁴⁰² de façon concrète et différente à chacune des affaires qui lui sont présentées.

Par son article 1382, le code civil désigne en réalité une présomption de fait, à savoir le regroupement d'indices et l'élaboration d'inférence par le juge pour déterminer la véracité du fait à prouver à partir de ces indices¹⁴⁰³. L'effet de cette disposition s'opère sur ce que le droit de la *common law* appelle le « standard de preuve », à savoir le niveau de certitude que doit avoir le juge afin de faire droit à une demande. Le juge ne peut donc pas se fonder sur n'importe quel élément de preuve, souvent appelé preuve indiciaire ou indice, pour présumer le fait pertinent objet de la preuve.

¹³⁹⁸ Article 1382 du code civil.

¹³⁹⁹ Sur le cas de la responsabilité du fait des vaccins contre l'hépatite B, la CJUE a considéré que la présomption de fait réalisé par les juges lors de leur raisonnement probatoire n'aboutit pas « à méconnaître la charge de la preuve institué par ledit article [...] » - Cour de justice, 21 juin 2017, n° C-621/15.

¹⁴⁰⁰ Du fait de l'absence de consensus scientifique sur le sujet.

¹⁴⁰¹ Pour un cas d'espèce dans lequel les juges ont présumé le lien de causalité du fait de la temporalité entre la vaccination et la maladie ainsi que l'absence de prédisposition du patient, voir Cour de cassation, Chambre civile 1, 9 juillet 2009, n° 08-11.073. Pour un cas de refus de cette présomption de fait sur le lien causal, voir Cour de cassation, Chambre civile 1, 22 mai 2008, n° 06-18.848.

¹⁴⁰² De ce fait, le contrôle des présomptions de fait échappe à la compétence de la Cour de cassation.

¹⁴⁰³ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 256.

§2. Les caractéristiques matérielles de la présomption de fait

La présomption de fait. Au-delà des distinctions procédurales et des discussions notionnelles entre les différents types de présomptions, la présomption doit être étudiée sur un plan matériel. La question n'est alors plus de déterminer le mécanisme juridique, la place de la présomption au sein du droit ou encore ses impacts sur les différents piliers de la preuve¹⁴⁰⁴. Il s'agit désormais d'étudier les éléments qui composent la notion de présomption. Afin de réaliser cette analyse, nous partons de la définition de la présomption fournie par Ljupcho GROZDANOVSKI, à savoir :

*« La présomption est une preuve indirecte, dont l'objet est une probabilité qui, dans un état normal des choses, apparaît comme le substitut plausible à un fait dont la preuve directe est recherchée, mais est indisponible ou difficile à produire. »*¹⁴⁰⁵

Cette définition met en exergue trois caractéristiques principales de la présomption que sont le fait inconnu dont la preuve est recherchée mais indisponible, un fait connu dont la preuve est apportée et enfin un phénomène de substitution fondée sur la probabilité permettant de déduire du fait connu le fait inconnu (d'où la dénomination de preuve indirecte). L'exercice d'inférence¹⁴⁰⁶ présent dans la présomption de fait se rapproche du raisonnement probatoire du juge.

L'existence d'un fait inconnu à établir. Le fait inconnu désigne un élément dont l'existence fait l'objet de doute car la preuve directe est manquante. Cette carence probatoire justifie, dans un premier temps, le recours au mécanisme de raisonnement probatoire de la présomption. Le fait est considéré inconnu s'il est matériellement impossible de le prouver, mais également s'il représente une charge de preuve déraisonnablement lourde. Par exemple, la proposition de directive européenne sur la responsabilité en matière d'IA¹⁴⁰⁷ pose une présomption de lien de causalité entre la faute du défendeur et le préjudice causé par le résultat du système d'IA à

¹⁴⁰⁴ À savoir l'objet de la preuve, la charge de la preuve, les modes de preuves et l'administration de la preuve. Voir not. LARDEUX Gwendoline, « Preuves : règles de preuves », *Répertoire de droit civil*, 2018.

¹⁴⁰⁵ GROZDANOVSKI Ljupcho, « Le probable, le plausible et le vrai. Contribution à la théorie générale de la présomption en droit », *Revue interdisciplinaire d'études juridiques*, 84, Université Saint-Louis - Bruxelles, 2020.

¹⁴⁰⁶ « Opération logique par laquelle on admet une proposition en vertu de sa liaison avec d'autres propositions déjà tenues pour vraies. » - « Inférence - Définitions, synonymes, conjugaison, exemples | Dico en ligne Le Robert ».

¹⁴⁰⁷ Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022.

condition que le demandeur ne puisse pas « *raisonnablement accéder à une expertise et à des éléments de preuves suffisants*¹⁴⁰⁸ » ou s'il « *est excessivement difficile pour le demandeur de prouver le lien de causalité*¹⁴⁰⁹ ». Ces deux formulations désignent toutes les deux la situation de carence probatoire du fait à prouver, ici le lien de causalité, et donc du recours à la présomption de droit afin d'éviter que le demandeur soit écrasé sous sa charge probatoire.

Le second argument pour l'utilisation d'une présomption est la nécessité à établir ce fait inconnu. Pour caractériser cette nécessité, le fait doit être pertinent pour le litige en présence. Il s'agit de la théorie de la pertinence du fait à prouver présentée précédemment¹⁴¹⁰. L'analyse de la pertinence des preuves est laissée à l'appréciation souveraine des juges et sera étudiée dans le cadre du prochain chapitre sur la force probante. Toutefois, c'est bien la pertinence du fait inconnu qui justifie sa nécessité à être prouvé. Un fait inconnu mais inutile pour l'issue du litige ne fera pas l'objet de présomptions, que l'on considère les présomptions de droit ou les présomptions de fait.

L'existence d'un ou plusieurs faits certains établis. Les faits certains sont, par nature, dénués de tout doute. Ils servent d'indices à celui qui raisonne pour présumer le fait inconnu¹⁴¹¹. Jean Domat disait ainsi que « *la certitude ou l'incertitude des présomptions [...] dépend de la certitude ou l'incertitude des faits dont on tire les présomptions, et de la justesse des conséquences qu'on tire de ces faits pour la preuve de ceux dont il s'agit.* »¹⁴¹². La qualification de preuve indiciaire pour un élément de preuve signifie donc qu'il fait partie de ce « *réseau de vérités* », pour reprendre l'image de Domat, sur lequel se fonde la présomption du juge. Pour illustrer la notion de faits établis, reprenons le raisonnement du juge concernant le lien de causalité entre le vaccin de l'hépatite B et la sclérose en plaque du demandeur. Du fait de l'absence de consensus scientifique, la preuve du lien de causalité, à savoir le fait à prouver, était impossible à apporter. Il devient alors le fait inconnu. Dans l'arrêt n° 08-11.073 du 9 juillet 2009, la temporalité entre la vaccination et la maladie ainsi que l'absence de prédisposition du

¹⁴⁰⁸ *Ibid.* Article 4.4 relatif aux systèmes d'IA à haut risque.

¹⁴⁰⁹ *Ibid.* Article 4.5 relatif aux systèmes d'IA qui ne sont pas à haut risque.

¹⁴¹⁰ Voir la sous-section 2 de la section 2 du chapitre précédent dédiée à la théorie de l'utilité probatoire. Pour rappel, les preuves offertes par les parties ne font pas l'objet d'une analyse de leur pertinence mais seulement les preuves demandées au juge, au travers d'une analyse de l'utilité probatoire des mesures d'accès aux dites preuves. Le parallèle est donc fait avec la notion théorique et non avec son application en droit civil français.

¹⁴¹¹ CANSACCHI Giorgio, *Le prezunzioni nel diritto internazionale : contributo allo studio della prova nel processo internazionale*, Naples, Eugenio Jovene, 1939, p. 11 cité par GROZDANOVSKI Ljupcho, « Le probable, le plausible et le vrai. Contribution à la théorie générale de la présomption en droit », *Revue interdisciplinaire d'études juridiques*, 84, Université Saint-Louis - Bruxelles, 2020. P. 43.

¹⁴¹² DOMAT Jean, *Œuvres complètes*, Paris, Firmin Didot, 1828, P. 166, cité par *Ibid.* P. 49.

patient sont les indices établis sur lequel s'est appuyé le juge pour déduire l'existence d'un lien de causalité¹⁴¹³, le fait inconnu.

§3. Le résultat d'un système d'IA comme indice au service d'une présomption de fait

Annonce de plan. La qualification du résultat produit par le système d'IA comme une preuve indiciaire au service du raisonnement du juge apparaît comme l'option la plus logique. Toutefois, le résultat du système d'IA peut également prétendre au statut de fait inconnu présumé. Cette double possibilité dépend du rôle joué par le résultat dans le litige : soit en tant que simple élément existant **(A)**, soit en tant qu'information signifiante¹⁴¹⁴ **(B)**.

A- La qualification du résultat en tant qu'élément existant

Le simple élément produit par un système d'IA. Pour rappel, le règlement européen sur l'IA définit le système d'IA comme suivant :

« Un système automatisé conçu pour fonctionner à différents niveaux d'autonomie, qui peut faire preuve d'une capacité d'adaptation après son déploiement et qui, pour des objectifs explicites ou implicites, déduit, à partir des données d'entrée qu'il reçoit, la manière de générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels ; »¹⁴¹⁵

Dans cette définition, le législateur européen donne plusieurs exemples de catégories de résultats pouvant être produits par des systèmes d'IA. L'expression « élément existant » n'appartient ni ne désigne aucune de ces catégories de résultats. Cette proposition d'expression désigne le résultat produit par un système d'IA pris comme un simple fait existant établi ayant entraîné des conséquences, avérées ou potentielles, sur son environnement. Cette catégorie peut donc aussi bien désigner un contenu, une prédiction, une décision ou une recommandation. Le

¹⁴¹³ Cour de cassation, Chambre civile 1, 9 juillet 2009, n° 08-11.073.

¹⁴¹⁴ Cette dichotomie se rapproche de la taxonomie plus générale élaborée par Andrea Roth entre les preuves indépendantes de la qualité de leur machine et les preuves dépendantes de cette qualité : ROTH Andrea L., « Machine Testimony », *The Yale Law Journal*, 126, 2017.

¹⁴¹⁵ « Règlement sur l'intelligence artificielle », 2024. Article 3.1. Emphase ajoutée par l'auteur.

critère de qualification repose sur l'emploi du résultat comme élément de preuve et non sur sa nature. Par conséquent, cette qualification est contextuelle, un même résultat pouvant être employé de différentes manières selon le litige.

Illustration de la notion. En ce qui concerne les images, prenons l'exemple d'une situation où une personne, généralement une femme, est victime d'hypertrucages à caractère pornographique sur Internet. La publication de telles photos sans le consentement de la personne constitue un délit sanctionné pénalement jusqu'à un an d'emprisonnement et 15.000 euros d'amende¹⁴¹⁶. Ce type de montage peut également être constitutif de harcèlement moral au sens des articles 222-33-2 et suivants du code pénal¹⁴¹⁷. Dans ce contexte, les hypertrucages sont des résultats produits par un système d'IA et qualifiés comme des preuves établies caractérisant l'un des éléments matériels de ces infractions.

En ce qui concerne les décisions ainsi que les prédictions, prenons l'exemple d'un résultat utilisé comme décision ou comme aide à une décision ayant abouti à un préjudice envers un individu. La directive européenne sur la responsabilité du fait de l'IA proposée par la Commission européenne énonce qu'une présomption de lien de causalité sous trois conditions. L'une de ces conditions est la preuve par le demandeur « *que le résultat du système d'IA ou l'incapacité de celui-ci à produire un résultat est à l'origine du dommage.* ». L'existence d'un résultat particulièrement négatif pour le demandeur est alors un indice pouvant permettre au juge de présumer que l'existence d'un tel résultat a pu exercer une influence sur la survenance du dommage subi. À nouveau, c'est la seule existence du résultat du système d'IA qui constitue une preuve. Sa signification (la solvabilité financière du demandeur, le potentiel du candidat à l'entretien d'embauche, etc.), et donc sa fiabilité, n'est pas pertinente dans l'intérêt du résultat du système d'IA utilisé en tant que preuve.

La qualification comme preuve indiciaire. Dans ce contexte d'utilisation probatoire du résultat d'un système d'IA, l'existence du résultat en tant qu'élément existant représente l'intérêt central de la preuve par un système d'IA. Dès lors que la preuve que le système d'IA a bien produit un tel résultat est établi, son existence devient certaine, indépendamment de la fiabilité et l'absence de biais du système d'IA l'ayant généré. La preuve produite ne dépend alors pas de la qualité du système d'IA l'ayant produite¹⁴¹⁸. Par conséquent, en tant que fait

¹⁴¹⁶ Article 229-1 du code pénal.

¹⁴¹⁷ La publication de photo montage est une forme d'harcèlement parmi d'autres. « Le cyberharcèlement », sur *Prévention internet et réseaux sociaux*.

¹⁴¹⁸ ROTH Andrea L., « Machine Testimony », *The Yale Law Journal*, 126, 2017.

certain, il correspond à la définition de la preuve indiciaire que le juge, pénal ou civil, peut utiliser souverainement pour forger son intime conviction, de façon similaire à tout autre élément de preuve¹⁴¹⁹.

B- La qualification du résultat en tant qu'information signifiante

Le résultat d'un système d'IA en tant que probabilité. Le résultat d'un système d'IA peut également être employé en tant qu'élément de preuve pour la connaissance qu'il apporte sur le litige. En d'autres termes, ce n'est pas tant l'existence du résultat qui sert d'élément de preuve mais sa signification. Le résultat est alors pris pour ce qu'il est censé représenter. Pour illustrer, nous utiliserons l'exemple d'un système d'IA qui serait utilisé pour évaluer la performance actuelle et future d'un salarié¹⁴²⁰. La production d'une mauvaise note de performance par le système d'IA est une réalité qui, prise en tant que telle, correspond à un indice pour le juge afin de procéder à son raisonnement probatoire sur le dommage qu'aurait subi le salarié évalué. Il faut différencier cette preuve de la preuve de la mauvaise performance du salarié déduite et/ou prédite par le système d'IA. D'un côté, l'existence d'un score associé au salarié par le système d'IA est un fait certain qui, à titre d'exemple, a pu entraîner des effets préjudiciables pour le salarié. De l'autre côté, la performance du salarié, bonne ou mauvaise, ne peut pas être affirmée avec certitude sur la base du résultat produit par le système d'IA. En d'autres termes, l'information signifiée par le résultat du système d'IA, la performance du salarié, n'est pas un fait certain. La confiance dans l'information signifiée par le résultat du système d'IA dépend alors de l'exactitude, et plus généralement de la qualité de fonctionnement, du système d'IA. Par conséquent, l'existence du score produit par le système d'IA et l'existence d'une mauvaise performance du salarié sont deux faits différents qui doivent être distingués.

La qualification délicate du résultat probabiliste comme preuve indiciaire. Les systèmes d'IA utilisés aujourd'hui développent une logique de raisonnement à partir de données passées. Ils appliquent ensuite ce fonctionnement à des données qui leur sont fournies, appelées données

¹⁴¹⁹ À l'exception des modes de preuves bénéficiant d'une certaine force probante imposée au juge par le législateur.

¹⁴²⁰ Il s'agit d'un usage des systèmes d'IA dans le secteur de la gestion de la main-d'œuvre est considéré à haut risque par le règlement européen sur l'IA : « systèmes d'IA destinés à être utilisés pour prendre des décisions influant sur les conditions de relations professionnelles, la promotion ou le licenciement dans le cadre de relations professionnelles contractuelles, pour attribuer des tâches sur la base du comportement individuel, de traits de personnalité ou de caractéristiques personnelles ou pour suivre et évaluer les performances et le comportement de personnes dans le cadre de telles relations. » - Annexe III.4(b).

d'entrée. Dans notre exemple, le score de performance du salarié représente en réalité le fait que son comportement, décrit au travers des données d'entrées, s'approche de façon plus ou moins importante du comportement d'autres salariés considérés comme des mauvais salariés. Au-delà de savoir si la définition de la performance intégrée dans le raisonnement du système d'IA est pertinente, la caractéristique importante est que la mauvaise performance du salarié est induite de plusieurs éléments (données passées et informations sur le salarié en question) qui rendent suffisamment probable le fait qu'il corresponde à un mauvais salarié. Ce raisonnement ressemble alors grandement à la description d'une présomption. Ainsi, la mauvaise performance, actuelle et/ou future du salarié, est présumée algorithmiquement à partir de faits considérés comme établis (les données d'entrées). Pris dans son aspect signifiant, le résultat du système d'IA s'apparente donc davantage à un fait présumé qu'à un fait certain. Dès lors, il s'éloigne significativement de la définition de l'indice établi servant idéalement de base à une présomption de fait pour le juge. Malgré cela, il n'est pas envisageable d'avancer que le résultat d'un système d'IA, pris sous son aspect probabiliste signifiant, ne mérite pas la qualification de preuve indiciaire au service d'une présomption de fait. Ainsi, l'évaluation des salariés par leur supérieur lors d'un entretien ont été déjà utilisés comme éléments de preuve par le salarié lui-même pour contester son licenciement¹⁴²¹. Or, l'évaluation réalisée par un humain compétent ne représente pas non plus la performance du salarié de façon certaine. Par conséquent, la distinction entre le résultat généré par un système d'IA pris comme un élément existant et le résultat généré par un système d'IA pris comme une information signifiante n'a pas de conséquences sur sa qualification juridique. Dans les deux cas, le résultat produit par un système d'IA est une preuve indiciaire.

CONCLUSION DE LA SECTION I :

Employés au service d'un objectif probatoire, les résultats produits par un système d'IA deviennent des éléments de preuve. À l'image d'autres procédés technologiques de preuves avant lui, le système d'IA entre, par élimination, dans la catégorie des preuves par présomption judiciaire. Ainsi, les résultats produits par un système d'IA peuvent être assimilés à des indices sur lesquels le juge peut s'appuyer pour présumer un fait inconnu, en l'espèce l'objet de la preuve. Si cette qualification s'explique par la nature du système d'IA, elle ne présente aucun lien avec la qualité de fonctionnement du logiciel utilisé comme procédé probatoire. Par

¹⁴²¹ Par exemple, sur l'utilisation de son évaluation élogieuse par le salarié pour contester son licenciement intervenu moins de deux mois après le rendu de l'évaluation : Cour de cassation, Chambre sociale, 22 mars 2011, n° 09-68.693.

conséquent, la qualification du résultat produit par un système d'IA ne s'articule aucunement avec les exigences du règlement européen relatives au fonctionnement dudit logiciel. De plus, la qualification du résultat d'un système d'IA dans la catégorie des preuves par présomption judiciaire entraîne son exclusion du système de preuve légale. Conformément à la loi, le recours aux preuves par présomption judiciaire est limité au système de preuve libre¹⁴²², excluant donc, en principe, l'usage des systèmes d'IA pour la preuve d'un acte juridique¹⁴²³.

Appliqué au sein du système de preuve libre, le principe de liberté de la preuve ne signifie pas que tous les éléments de preuves collectés par les parties sont recevables. Afin d'être pris en compte par le juge, le résultat produit par un système d'IA doit tout de même satisfaire les conditions du principe de licéité imposé par le droit de la preuve.

Section II : La licéité de la preuve produite par un système d'IA

Annonce de plan. Malgré son nom, le régime de preuve libre ne signifie pas que toutes les preuves offertes ou demandées par les parties dans ce contexte sont automatiquement admises. Toutes les preuves doivent respecter un principe général en droit de la preuve qui est le principe de licéité de la preuve. Par conséquent, le recours aux preuves par un système d'IA, dans le système de preuve libre, ne garantit pas leur prise en compte par le juge. De façon similaire au travail de qualification réalisé précédemment, il s'agit d'identifier les interactions possibles entre la qualité de fonctionnement du système d'IA et la licéité de son résultat. Au préalable, il convient de définir le principe complexe et relativement désordonné législativement¹⁴²⁴ qu'est la licéité de la preuve (**Sous-section 1**). Ensuite, il sera réalisé une analyse de l'articulation entre le fonctionnement du système d'IA et ce principe de licéité, en particulier avec certains principes généraux du droit de la preuve (**Sous-section 2**).

¹⁴²² « Les présomptions qui ne sont pas établies par la loi, sont laissées à l'appréciation du juge, qui ne doit les admettre que si elles sont graves, précises et concordantes, et dans les cas seulement où la loi admet la preuve par tout moyen. » - Article 1382 du code civil.

¹⁴²³ Il existe plusieurs exceptions à l'obligation de prouver un acte juridique par écrit. Ces exceptions peuvent se justifier en raison de l'existence d'une convention de preuve entre les parties (Article 1356 du code civil) ou parce que la preuve de l'acte est apportée par un tiers (Cour de Cassation, Chambre civile 1, 10 mai 2005, n° 02-11.759.). Également, la preuve d'un écrit en matière prud'homale est libre (Cour de Cassation, Chambre sociale, 27 mars 2001.) de la même manière que la preuve de l'exécution d'une obligation (Cour de Cassation, Chambre civile 1, 15 juillet 1975, n° 74-12.965.). Cependant, il s'agit d'exceptions au principe. Par conséquent, le recours à des résultats de systèmes d'IA dans ces situations ne sera pas abordé au sein de notre analyse.

¹⁴²⁴ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

Sous-section 1 : Le principe de licéité de la preuve.

Annnonce de plan. En dehors de l'admissibilité restreinte de certains modes de preuve dans le système de preuve légale, les règles de recevabilité des preuves en droit français sont éparpillées entre des décisions de justice et des règles spécifiques à certains modes de preuves. Le principe de licéité de la preuve est le prisme au travers duquel sont étudiées les conséquences du fonctionnement d'un système d'IA, en particulier au regard du règlement européen sur l'IA, sur la recevabilité de son résultat comme élément de preuve. Par conséquent, il est indispensable de définir ce principe complexe afin de mettre en lumière les composantes qui seront utilisées lors de la présente étude (§1). Le recours au principe de licéité étant dépendant du contexte du litige, son périmètre d'application doit aussi être précisé afin de clarifier les limites de l'analyse réalisée (§2).

§1. Un principe constitué de deux blocs

La licéité des preuves spéciales. Le principe de licéité de la preuve trouve son fondement dans l'article 9 du code de procédure civile qui impose à « *chaque partie de prouver conformément à la loi les faits nécessaires au succès de sa prétention* »¹⁴²⁵. Le principe s'est toutefois relativement détaché de son fondement législatif pour s'élargir aux règles de droit. Désormais, la licéité de la preuve peut être définie comme le principe imposant aux parties « *que chaque élément de preuve soit produit ou recherché conformément aux règles de droit* »¹⁴²⁶. Ce passage de la conformité à la loi à la conformité au droit explique l'existence de deux blocs de licéité. Le premier bloc correspond aux situations où la preuve a été produite ou recherchée en violation d'une règle législative de preuve spéciale à un mode de preuve spécifique, aussi appelée « preuve nommée » par le pénaliste Albert Maron¹⁴²⁷.

A titre d'exemple, le code de procédure pénale autorise les officiers et agents de police judiciaires, lors d'une enquête préliminaire, à adresser des réquisitions à des personnes qualifiées afin qu'elles réalisent des examens techniques et scientifiques après avoir obtenu

¹⁴²⁵ Article 9 du code de procédure civile. Emphase ajoutée par l'auteur.

¹⁴²⁶ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.* P. 373.

¹⁴²⁷ Les preuves nommées sont les preuves faisant l'objet d'un régime probatoire spécifique énoncé par la loi. Albert Maron, « Mais où sont les neiges d'antan ? » (2007) 91 Dr Pen (Droit pénal) et MARON Albert, « Dr. Pén. », *Droit pénal*, 69, 2018., cités dans VERGES Etienne, « Loyauté et licéité, deux apports majeurs à la théorie de la preuve », *Recueil Dalloz*, 2014.

l'autorisation du procureur de la République¹⁴²⁸. Sans cette autorisation, les résultats desdits examens sont empreints d'illicéité et sont alors déclarés nuls par les juges¹⁴²⁹. Les réquisitions sont donc un mode de preuve réglementé par des règles spécifiques du code de procédure pénale qui précisent les modalités de production, de collecte ainsi que la sanction encourue en cas de manquement à ce régime probatoire spécial. La même logique s'applique aussi aux témoignages ou encore à la géolocalisation¹⁴³⁰.

L'absence de régimes probatoires spécifiquement applicables à la preuve par systèmes d'IA. Le recours aux systèmes d'IA à des fins probatoires est une finalité récente mais dont l'utilisation est déjà mise en place¹⁴³¹. À l'étranger, les technologies d'IA sont utilisées pour extraire et analyser de très grandes quantités de données stockées au sein d'appareils électroniques¹⁴³² ou encore analyser et comparer des échantillons d'ADN¹⁴³³. En France, les technologies d'IA sont utilisées particulièrement par le biais de la reconnaissance faciale. Le recours à la reconnaissance faciale est encadré par des règles sur les traitements automatisés de données à caractère personnel dans le cadre d'activités répressives prévues par la loi n°2018-493 relative à la protection des données personnelles¹⁴³⁴, transposant de la directive européenne « Police-Justice »¹⁴³⁵. Ce cadre juridique limite la reconnaissance faciale à la comparaison avec le fichier de traitement des antécédents judiciaires (TAJ) seulement en cas de nécessité absolue au regard de la finalité du traitement¹⁴³⁶ et s'il existe un doute sur l'identité de la personne¹⁴³⁷.

¹⁴²⁸ Article 77-1 du code de procédure pénale.

¹⁴²⁹ Cour de Cassation, Chambre criminelle, 14 octobre 2003, n° 03-84.539 Cour de cassation, Chambre criminelle, 18 juin 2019, n° 19-80.105 DELAGE Pierre-Jerôme, « Vers un principe de fiabilité de la preuve pénale ? », *RSC*, 2019.

¹⁴³⁰ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

¹⁴³¹ LELIEUR Juliette, « L'intelligence artificielle, une nouvelle technologie probatoire en émergence », *AJ Pénal*, 2022.

¹⁴³² Il s'agit du logiciel Hansken utilisé aux Pays-Bas notamment par les équipes d'enquêteurs en matière fiscale : <https://www.hansken.nl/an-introduction-to-hansken>

¹⁴³³ C'est le cas des États-Unis et du Canada par le biais du logiciel STRmix : <https://strmix.com/>

¹⁴³⁴ « LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles (1) », 2018-493, 2018. « Ordonnance n° 2018-1125 du 12 décembre 2018 prise en application de l'article 32 de la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles et portant modification de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel », 2018.

¹⁴³⁵ « Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil », OJ L, 2016.

¹⁴³⁶ « Peuvent être enregistrées dans le présent traitement [TAJ] les catégories de données à caractère personnel et informations suivantes : [...] - photographie comportant des caractéristiques techniques permettant de recourir à un dispositif de reconnaissance faciale (photographie du visage de face) ; » - Article R. 40-26 du code de procédure pénale.

¹⁴³⁷ Conseil d'État, 26 avril 2022, n° 442364.

Pour autant, il n'existe pas encore de dispositions probatoires encadrant l'utilisation spécifique de la reconnaissance faciale en droit français.

Parmi les usages de systèmes d'IA qualifiés comme « à risque élevé » dans le règlement sur l'IA, le législateur européen intègre l'utilisation d'un système d'IA par les autorités répressives, notamment pour des finalités liées à la collecte et au traitement d'éléments de preuves. La fonction probatoire des systèmes d'IA est ainsi envisagée dans les considérants du règlement européen :

« Compte tenu de la nature des activités en question et des risques y afférents, ces systèmes d'IA à haut risque devraient comprendre en particulier les systèmes d'IA destinés à être utilisés par les autorités répressives pour réaliser des évaluations individuelles des risques, pour servir de polygraphes ou d'outils similaires ou pour analyser l'état émotionnel de personnes physiques, pour détecter les hypertrucages, pour évaluer la fiabilité des preuves dans les procédures pénales, pour prédire la survenance ou la répétition d'une infraction pénale réelle ou potentielle sur la base du profilage de personnes physiques, ou pour évaluer les traits de personnalité, les caractéristiques ou les antécédents délictuels de personnes physiques ou de groupes à des fins de profilage dans le cadre d'activités de détection, d'enquête ou de poursuite relatives à des infractions pénales, ainsi que d'analyse de la criminalité des personnes physiques »¹⁴³⁸

S'il met en place une certification obligatoire pour les systèmes d'IA à haut risque, les obligations contenues dans le règlement européen sur l'IA ne doivent pas être confondues avec des exigences relatives à la recevabilité des preuves produites par un système d'IA¹⁴³⁹. Pour autant, cela ne signifie pas que le règlement sur l'IA ne pourra pas inspirer les législateurs nationaux qui souhaiteraient établir des règles spécifiques à la recevabilité des résultats produits par un système d'IA. Ainsi, la loi n° 2023-380 relative aux jeux Olympiques et Paralympiques de 2024 indique que le traitement algorithmique, qui sera réalisée sur les images enregistrées par des systèmes de vidéoprotection, devra respecter certaines exigences, notamment :

« Lorsque le traitement algorithmique employé repose sur un apprentissage, des garanties sont apportées afin que les données d'apprentissage, de validation et de test choisies soient pertinentes, adéquates et représentatives. Leur traitement doit être loyal et éthique, reposer sur des critères objectifs et

¹⁴³⁸ « Règlement sur l'intelligence artificielle », 2024. Considérant 38.

¹⁴³⁹ La notion de légalité de la preuve doit être clairement distinguée de celle de licéité. Voir le paragraphe suivant.

permettre d'identifier et de prévenir l'occurrence de biais et d'erreurs. Ces données font l'objet de mesures de sécurisation appropriées ; »¹⁴⁴⁰

A cela s'ajoute une obligation de mise en place d'un système d'enregistrement des événements¹⁴⁴¹, de mesures de contrôle humain¹⁴⁴² ainsi que la fourniture d'une documentation technique si le traitement algorithmique des images est réalisé par un tiers¹⁴⁴³. Les garanties imposées par le texte français présentent des points communs avec les obligations contenues dans le règlement européen sur l'IA. À ce jour, le traitement algorithmique des enregistrements issus de systèmes de vidéoprotection est une mesure temporaire menée à titre expérimental dans le seul cadre de certaines manifestations sportives, récréatives ou culturelles¹⁴⁴⁴. Du fait de ce statut expérimental, la loi ne prévoit aucune disposition relative à l'utilisation probatoire de ce traitement algorithmique des enregistrements de vidéosurveillance. Par conséquent, les systèmes d'IA ne font, à ce jour, l'objet d'aucun régime de preuve spécifique clair. Par conséquent, l'articulation entre le fonctionnement d'un système d'IA et la licéité de ses résultats en tant que preuve ne sera pas abordé au sein de ce premier bloc de licéité.

L'analyse de la licéité au regard des principes généraux du droit de la preuve. Le second bloc composant le principe de licéité de la preuve est celui du respect des principes généraux de la preuve. Ces principes s'appliquent à toute preuve, que la preuve fasse déjà l'objet d'un régime probatoire spécial ou non. Dans le cas d'une preuve nommée, sa recherche et sa production doivent donc respecter à la fois son régime probatoire spécial et les principes généraux du droit. Les règles spéciales ne dérogent pas aux principes généraux mais s'y superposent. Ainsi, la sonorisation de cellules de garde à vue conformément aux règles du code

¹⁴⁴⁰ « LOI n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions (1) », 2023-380, 2023. Article 10. VI. 1°

¹⁴⁴¹ « 2° Le traitement comporte un enregistrement automatique des signalements des événements prédéterminés détectés permettant d'assurer la traçabilité de son fonctionnement ; » - Ibid. Article 10. VI.

¹⁴⁴² « 3° Le traitement permet des mesures de contrôle humain et un système de gestion des risques permettant de prévenir et de corriger la survenue de biais éventuels ou de mauvaises utilisations ; » - Ibid. Article 10. VI.

¹⁴⁴³ « Lorsque le traitement est développé ou fourni par un tiers, celui-ci fournit une documentation technique complète [...] » - Ibid. Article 10.VI.5°.

¹⁴⁴⁴ « A titre expérimental et jusqu'au 31 mars 2025, à la seule fin d'assurer la sécurité de manifestations sportives, récréatives ou culturelles qui, par l'ampleur de leur fréquentation ou par leurs circonstances, sont particulièrement exposées à des risques d'actes de terrorisme ou d'atteintes graves à la sécurité des personnes [...] » - Ibid. Article 10. I.

de procédure pénale¹⁴⁴⁵ ne garantit en aucun cas la licéité, et donc la recevabilité, des enregistrements dès lors que ces derniers ont été obtenus de manière illicite¹⁴⁴⁶.

Les principes généraux du droit de la preuve sont au nombre de cinq : le droit au respect de la vie privée, la loyauté de la preuve, le droit au respect de la dignité, les droits de la défense et le respect des secrets protégés juridiquement¹⁴⁴⁷. Imposée à toutes les preuves, la conformité aux principes généraux du droit de la preuve doit donc être interrogée dans le cas d'une preuve produite par un système d'IA. Cependant, il n'est pas question d'analyser de façon exhaustive le respect par une preuve générée par un système d'IA vis-à-vis des principes généraux du droit de la preuve. Premièrement, certains de ces principes présentent des variations selon le contexte d'application (une preuve fournie par l'employeur ou l'employé, une preuve collectée par les forces de l'ordre, etc.). En raison de l'approche transversale choisie ici, ces particularités dans l'application des principes généraux du droit de la preuve ne seront pas traitées. Secondement, le respect de certains des principes généraux du droit de la preuve dépend de la manière dont les preuves ont été obtenues et non les caractéristiques et/ou le fonctionnement du procédé probatoire. Or, la présente étude a pour objectif d'étudier l'articulation entre le fonctionnement du système d'IA, en particulier éléments techniques de preuve établis dans le cadre du règlement européen sur l'IA, et la recevabilité du résultat généré par un système d'IA pris en tant que preuve. Par conséquent, seuls les principes généraux du droit de la preuve qui présentent un lien avec les caractéristiques et/ou le fonctionnement du procédé probatoire seront analysés.

§2. Le champ d'application du principe

La relation avec le système de preuve légale. Dans la précédente section a été étudiée l'hypothèse de l'utilisation de la preuve par système d'IA au sein du système de preuve légale. La faible probabilité qu'un résultat de systèmes d'IA puisse être utilisé comme preuve d'un acte juridique a conduit la présente étude à se concentrer sur la licéité comme critère de recevabilité

¹⁴⁴⁵ Articles 706-92 et suivants du code de procédure pénale.

¹⁴⁴⁶ « [...] la conjugaison des mesures de garde à vue, du placement de MM. Y... et X... dans des cellules contiguës et de la sonorisation des locaux participait d'un stratagème constituant un procédé déloyal de recherche des preuves » - Cour de cassation, Chambre criminelle, 7 janvier 2014, n° 13-85.246. Voir également : VERGES Etienne, « Loyauté et licéité, deux apports majeurs à la théorie de la preuve », *Recueil Dalloz*, 2014.

¹⁴⁴⁷ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. PP.380 et suivantes.

de la preuve par système d'IA. Pour autant, le contrôle de la licéité des preuves ne s'arrête pas aux portes du système de preuve légale. En effet, un écrit sous signature privée est une preuve qui satisfait les conditions d'admissibilité restreinte que pose le système de preuve légale. Pour autant, si cet écrit viole les règles du droit de la preuve, il deviendra illicite et donc irrecevable au même titre que toute preuve apportée dans un système de preuve libre. C'est le cas d'un écrit qui n'aurait pas été communiqué à la partie adverse conformément au principe du contradictoire. L'interaction entre ces deux concepts peut être résumée par la formule suivante : *« une preuve légale peut se révéler illicite, et donc écartée des débats, de même qu'une preuve licite peut être déclarée inadmissible, car ne figurant pas dans la liste des preuves légales »*¹⁴⁴⁸. Par conséquent, si la preuve produite par un système d'IA parvient à l'avenir à obtenir une place parmi les preuves légales, elle n'échappera pas pour autant au contrôle de sa licéité.

La relation avec les concepts de « légalité de la preuve ». Un autre risque de confusion du principe de licéité de la preuve est possible avec les concepts de « légalité de la preuve ». Au préalable, il faut préciser que la « légalité de la preuve » n'est pas un synonyme de la « preuve légale » au sens du système de preuve légale. Le premier renvoie à la conformité par rapport à la loi de l'élément de preuve tandis que le second désigne la possibilité pour une preuve d'être admise dans le cadre du système de preuve légale.

Dans un premier temps, la légalité de l'élément de preuve n'affecte pas nécessairement sa licéité. D'une part, l'illégalité d'un élément de preuve renvoie à son manquement à une loi générale et non à une règle de preuve¹⁴⁴⁹. Par conséquent, une preuve illégale peut parfaitement être recevable. Ainsi, le fichier contenant des données personnelles non déclaré à la CNIL¹⁴⁵⁰ est licite et recevable malgré sa violation du droit des données personnelles¹⁴⁵¹. Dans un second temps, la « légalité de la preuve » est un concept particulièrement présent en droit pénal. Dans ce contexte, il désigne le fait que la preuve ait été recueillie conformément à des règles de preuve spécifiques explicitement mentionnées dans la loi. Comme expliqué précédemment, ces règles de preuves spéciales (composant le premier bloc de licéité) ne dérogent pas aux principes généraux du droit de la preuve (constituant le second bloc de licéité) mais s'y superposent. C'est

¹⁴⁴⁸ *Ibid.* P. 375.

¹⁴⁴⁹ *Ibid.* P. 375

¹⁴⁵⁰ Bien que rendu en 2020, la décision désigne ici la réglementation antérieure à l'entrée en vigueur du règlement européen 2016/679 qui met fin au régime de déclaration préalable des traitements de données personnelles.

¹⁴⁵¹ Cour de cassation, Chambre sociale, 25 novembre 2020, n° 17-19.523. En l'espèce, la Cour de cassation parle de « preuve illicite », ce qui n'aide pas à clarifier la distinction discutée ici et amenant certains auteurs à parler d'« admissibilité de la preuve illicite » - BARBIER Hugo, « L'admissibilité de la preuve illicite », *Revue trimestrielle de droit civil*, 2021.

ainsi que la Cour de cassation peut qualifier le recours à différentes techniques procédurales, toutes exécutées en respect des règles probatoires spécifiques, comme constituant un stratagème qui viole le principe de loyauté de la preuve¹⁴⁵². Par conséquent, le choix de la présente thèse de ne pas traiter du premier bloc de licéité¹⁴⁵³ pour se concentrer, de façon plus transversale, sur la relation entre la preuve produite par un système d'IA et les principes généraux du droit de la preuve (second bloc) pose certes une limite quant à l'exhaustivité de la présente analyse mais n'affecte en aucun cas ni sa pertinence, ni son obsolescence¹⁴⁵⁴.

Sous-section 2 : L'articulation avec les principes généraux du droit de la preuve

La sélection des principes généraux du droit pertinents. Provenant de sources juridiques différentes, les principes généraux du droit de la preuve regroupent un volume très important de règles. Bien qu'essentiels à la bonne administration de la justice, les principes généraux du droit de la preuve ne sont pas tous pertinents à aborder au sein de la présente recherche. Dès lors, le choix a été fait de se limiter aux principes qui présentent aujourd'hui le plus d'interactions avec le fonctionnement du système d'IA, et donc avec le règlement européen sur l'IA.

Premièrement, le principe du droit au respect de la dignité est issu de l'article 3 de la Conv. EDH : « *Nul ne peut être soumis à la torture ni à des peines ou traitements inhumains ou dégradants.* »¹⁴⁵⁵. Par ce principe, la Cour européenne des droits de l'Homme considère, de façon constante, comme irrecevables les preuves collectées grâce à des actes commis en

¹⁴⁵² Cour de cassation, Chambre criminelle, 7 janvier 2014, n° 13-85.246.

¹⁴⁵³ Le premier bloc de licéité a été écarté en raison de l'absence de règles de preuves spécifiques aux preuves produites par un système d'IA. Pour plus de détails, voir §1. Un principe constitué de deux blocs.

¹⁴⁵⁴ En d'autres termes, l'apparition dans les prochaines années de dispositions législatives probatoires spécifiques à la preuve par système d'IA n'entachera pas l'analyse réalisée ci-après.

¹⁴⁵⁵ CONSEIL DE L'EUROPE, « Convention européenne des droits de l'Homme », 1950.

violation dudit article (des menaces¹⁴⁵⁶ ou de la torture¹⁴⁵⁷ à l'encontre de l'accusé ou d'un tiers¹⁴⁵⁸). À ce jour, il est difficilement envisageable que le recours à un système d'IA puisse être qualifié comme un acte de torture au sens de la convention européenne des droits de l'Homme. Dans l'hypothèse contraire, une telle qualification sera justifiée par l'utilisation réalisée du système d'IA ayant produit la preuve et non sur son fonctionnement. Par conséquent, le principe du droit au respect de la dignité ne présente pas de lien avec le fonctionnement du système d'IA, et donc avec le règlement européen sur l'IA, et sera donc écarté du périmètre de la présente recherche.

Le deuxième principe général est le principe, exclusivement jurisprudentiel¹⁴⁵⁹, de loyauté de la preuve. Les preuves qualifiées de « déloyales » sont, d'une part, celles obtenues par le biais d'un stratagème et, d'autre part, celles recueillies à l'insu de la personne à l'encontre de laquelle elles sont opposées¹⁴⁶⁰. Tout d'abord, la notion de stratagème est défini comme « *un montage, une mise en scène, une opération clandestine* »¹⁴⁶¹. Par définition, le stratagème désigne une manière d'obtention d'un élément de preuve qui rend ce dernier déloyal. Par exemple, la provocation à la faute en droit civil¹⁴⁶², à l'infraction en droit pénal¹⁴⁶³ ou encore la dissimulation de son identité¹⁴⁶⁴ en vue de l'obtention de preuves sont des stratagèmes qui empreignent les preuves d'illicéité. Ensuite, la preuve peut être qualifiée de déloyale si elle a été obtenue alors que l'individu contre qui elles sont produites n'en était pas conscient. Il apparaît alors clairement que le critère est la connaissance de la récolte des preuves, généralement par enregistrement. Par conséquent, au sein de ces deux catégories de preuves déloyales, la manière dont est obtenue la preuve est le critère central. La qualité de fonctionnement du procédé probatoire n'affecte pas la loyauté de la preuve produite. Ainsi, le

¹⁴⁵⁶ « lorsque l'emploi d'une méthode d'interrogatoire prohibée par l'article 3 a eu des conséquences défavorables pour un requérant dans la procédure pénale dirigée contre lui, une réparation adéquate et suffisante implique, outre les exigences susmentionnées, des mesures de restitution se rapportant à l'incidence que cette méthode d'interrogatoire prohibée continue d'avoir sur le procès, parmi lesquelles figure en particulier le rejet des éléments de preuve que la violation de l'article 3 a permis de recueillir. » - Cour européenne des droits de l'Homme, 1 juin 2010, n° 22978/05. Par. 128. Emphase ajoutée par l'auteur.

¹⁴⁵⁷ « des éléments à charge [...] rassemblés au moyen d'actes de violence ou de brutalité ou d'autres formes de traitements pouvant être qualifiés de torture – ne doivent jamais, quelle qu'en soit la valeur probante, être invoqués pour prouver la culpabilité de la victime. » - *Jalloh c Allemagne*, 2006 Cour européenne des droits de l'Homme. Par. 105. ; *Nechiporuk et Yonkalo c Ukraine*, 2011 Cour européenne des droits de l'Homme.

¹⁴⁵⁸ Cour européenne des droits de l'Homme, 5 octobre 2017, n° 39014/12.

¹⁴⁵⁹ LARDEUX Gwendoline, « Preuves : règles de preuves », *Répertoire de droit civil*, 2018.

¹⁴⁶⁰ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 294.

¹⁴⁶¹ Cour de cassation, Chambre commerciale, 10 novembre 2021, n° 20-14.670.

¹⁴⁶² Cour de Cassation, Chambre sociale, 16 janvier 1991, n° 89-41.052.

¹⁴⁶³ Cour de Cassation, Chambre criminelle, 5 mai 1999, n° 97-83.117.

¹⁴⁶⁴ Cour de cassation, Chambre sociale, 18 mars 2008, n° 06-45.093.

principe de loyauté de la preuve ne semble pas particulièrement s'articuler avec le fonctionnement d'un système d'IA à des fins probatoires, et ne présente donc aucun lien avec le règlement européen sur l'IA. Dès lors, ce principe général du droit de la preuve est également écarté du périmètre de recherche.

Enfin, le troisième principe général à examiner est le respect des secrets protégés juridiquement. Qu'il s'agisse du secret des affaires¹⁴⁶⁵, du secret médical¹⁴⁶⁶, du secret bancaire¹⁴⁶⁷ ou encore du secret des correspondance entre l'avocat et son client¹⁴⁶⁸, tous sont regroupés au sein de ce principe général du droit de la preuve. Au travers de ces exemples, il apparaît clairement que l'existence d'un secret protégé juridiquement est dépendante du contexte. Par exemple, le secret médical désigne l'ensemble des informations dont un professionnel intervenant dans le système de santé a connaissance concernant un patient¹⁴⁶⁹. Dans ce contexte, le fait qu'une preuve soit produite par un système d'IA ou par un humain n'affecte pas spécifiquement sa protection par le secret médical. Sans qu'il soit nécessaire d'entrer plus en détails dans les règles applicables à chacun de ces secrets, il ne semble pas exister de lien spécifique entre le respect des secrets protégés juridiquement et la qualité du fonctionnement du système d'IA qui aurait produit ces preuves. Il ne s'agit pas de démontrer qu'il est impossible pour une preuve produite par un système d'IA d'être couverte par un secret protégé juridiquement. L'objectif est de mettre en lumière que cette protection ne dépend ni du procédé ayant produit la preuve en question ni de la qualité de fonctionnement de ce procédé probatoire. Par conséquent, le principe du respect des secrets protégés juridiquement ne semble pas s'articuler avec le règlement européen sur l'IA dans le cadre d'une preuve produite par un système d'IA. Il convient donc de l'exclure également de l'analyse de l'articulation entre les éléments techniques de preuve utilisés pour certifier un système d'IA à haut risque, au sens du règlement européen sur l'IA, et la recevabilité du résultat produit par ledit système.

¹⁴⁶⁵ « si le secret des affaires ne fait pas en lui-même obstacle au prononcé d'une mesure d'instruction préventive ainsi qu'à la production forcée de pièces, c'est à la condition qu'elles soient nécessaires, donc qu'elles constituent le seul moyen de préserver les droits du demandeur » - Cour de cassation, Chambre civile 1, 22 juin 2017, n° 15-27.845. « Décret n° 2018-1126 du 11 décembre 2018 relatif à la protection du secret des affaires », 2018-1126, 2018. Article 1. Codifié à l'article R. 153-1 du code du commerce.

¹⁴⁶⁶ Pour une application du secret médical en droit du travail : Cour de Cassation, Chambre sociale, 9 octobre 2007, n° 06-42.959. En droit pénal, voir : Cour de cassation, Chambre criminelle, 24 avril 2007, n° 06-88.051.

¹⁴⁶⁷ Cour de cassation, Chambre commerciale, 4 juillet 2018, n° 17-10.158.

¹⁴⁶⁸ Cour de cassation, Chambre commerciale, 14 septembre 2010, n° 09-16.347.

¹⁴⁶⁹ « l'ensemble des informations concernant la personne venues à la connaissance du professionnel, de tout membre du personnel de ces établissements, services ou organismes et de toute autre personne en relation, de par ses activités, avec ces établissements ou organismes. » - Article L 1110-4 du code de la santé publique.

Annonce de plan. Une fois écartés les principes du droit à la dignité, de loyauté et de respect des secrets protégés juridiquement, seuls deux principes fondamentaux restent donc à analyser. Grâce à des illustrations concrètes, il sera étudié l'articulation du fonctionnement d'un système d'IA avec le droit au respect de la vie privée, d'une part (§1), et avec les droits de la défense, d'autre part (§2).

§1. Le droit au respect de la vie privée

Annonce de plan. En matière probatoire, le droit au respect de la vie privée se fonde sur l'article 9 du code civil¹⁴⁷⁰, l'article préliminaire du code de procédure pénale¹⁴⁷¹, l'article 7 de la Charte des droits fondamentaux de l'Union européenne¹⁴⁷² et sur l'article 8 de la Convention européenne des droits de l'Homme¹⁴⁷³. Autant en matière pénale¹⁴⁷⁴ que civile¹⁴⁷⁵, le droit au respect a la vie privée est venu s'opposer au droit à la preuve des parties. Ces deux droits possèdent une valeur normative identique¹⁴⁷⁶. Il ne semble alors pas possible d'en privilégier l'un sur l'autre par principe¹⁴⁷⁷. Cependant, la Convention européenne prévoit la possibilité de déroger au droit au respect de la vie privée contrairement au procès équitable (dont le droit à la preuve est issu), amenant certains auteurs à considérer que le premier prime sur le second¹⁴⁷⁸. Malgré cette possibilité de domination du droit à la preuve sur le droit à la vie privée, les juges ont décidé d'articuler de façon plus souple les deux droits. L'interaction entre ces deux principes a donc pris la forme d'une conciliation par la jurisprudence et des critères d'articulation ont été élaborés (A). Notoirement connu pour leurs risques envers le droit au

¹⁴⁷⁰ « Chacun a droit au respect de sa vie privée. [...] » - Article 9 du code civil

¹⁴⁷¹ « Au cours de la procédure pénale, les mesures portant atteinte à la vie privée d'une personne ne peuvent être prises, sur décision ou sous le contrôle effectif de l'autorité judiciaire, que si elles sont, au regard des circonstances de l'espèce, nécessaires à la manifestation de la vérité et proportionnées à la gravité de l'infraction. » - Article préliminaire du code de procédure pénale.

¹⁴⁷² « Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de ses communications. » - « Charte des droits fondamentaux de l'Union européenne », OJ C, 2012.

¹⁴⁷³ « Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance. » - CONSEIL DE L'EUROPE, « Convention européenne des droits de l'Homme », 1950.

¹⁴⁷⁴ Cour européenne des droits de l'Homme, 14 avril 1990, n° 11105/84.

¹⁴⁷⁵ Cour de Cassation, Chambre sociale, 2 octobre 2001, n° 99-42.942.

¹⁴⁷⁶ Le droit à la preuve se fonde sur l'article 6 de la Convention européenne des droits de l'Homme. Pour plus d'information sur le droit à la preuve, voir not. BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010.

¹⁴⁷⁷ LOISEAU Grégoire, « Droits de la personnalité », 2017.

¹⁴⁷⁸ LEPAGE Agathe, « Droits de la personnalité - Droit de la personnalité et droit à la preuve », *Répertoire de droit civil*, 2023. SAINT-PAU Jean-Christophe, *Droits de la personnalité*, LexisNexis, 2013.

respect de la vie privée, le recours aux technologies d'IA à des fins probatoires impose de s'interroger sur la conformité des preuves produites au principe du respect de la vie privée **(B)**.

A- La conciliation entre le droit à la preuve et le droit au respect de la vie privée

Une atteinte nécessaire et proportionnée au droit au respect de la vie privée en matière civile. La logique de conciliation entre le droit au respect de la vie privée et les besoins probatoires est évoquée pour la première fois en 1987 concernant une mesure de production forcée prononcée par un juge envers un tiers :

« Attendu, ensuite, que le juge civil a le pouvoir d'ordonner à un tiers de produire tout document qu'il estime utile à la manifestation de la vérité ; que ce pouvoir n'est limité que par l'existence d'un motif légitime tenant soit au respect de la vie privée, sauf si la mesure s'avère nécessaire à la protection des droits et libertés d'autrui, soit au secret professionnel¹⁴⁷⁹ ; »

La première chambre civile énonce alors un critère de nécessité qui permet de justifier l'atteinte au droit d'un tiers. Vingt ans plus tard, la Cour de cassation pose à nouveau un principe de conciliation entre les deux droits dans un contexte de production de preuve par une partie :

« Toute atteinte à la vie privée n'est pas interdite, et qu'une telle atteinte peut être justifiée par l'exigence de la protection d'autres intérêts, dont celle des droits de la défense, si elle reste proportionnée au regard des intérêts antinomiques en présence¹⁴⁸⁰ ; »

On retrouve alors une condition de proportionnalité de l'atteinte au droit au respect de la vie privée, à laquelle vient s'ajouter un an après la condition de nécessité¹⁴⁸¹ déjà énoncé en matière de demande de preuve. Par conséquent, qu'il s'agisse d'une production de preuve ou d'une

¹⁴⁷⁹ Cour de Cassation, Chambre civile 1, 21 juillet 1987, n° 85-16.436.

¹⁴⁸⁰ Cour de cassation, Chambre commerciale, 15 mai 2007, n° 06-10.606.

¹⁴⁸¹ Cour de cassation, Chambre civile 1, 16 octobre 2008, n° 07-15.778. On retrouve aussi la formulation « indispensable » dans la jurisprudence, qui renvoie à la condition de nécessité : Cour de cassation, Chambre civile 1, 25 février 2016, n° 15-12.403. Au niveau européen, voir Cour européenne des droits de l'Homme, 7 août 2021, n° 27516-14. Par. 48.

demande de preuve¹⁴⁸², l'atteinte au droit au respect de la vie privée est possible uniquement si elle était nécessaire et proportionnée¹⁴⁸³.

Le contrôle de la nécessité et de la proportionnalité de l'atteinte en matière pénale.

L'évaluation de l'atteinte au droit au respect de la vie privée suit les mêmes conditions de nécessité et de proportionnalité de l'atteinte qu'en droit civil. Toutefois, cet examen n'est pas appliqué de la même manière selon que le procédé probatoire dispose d'un régime probatoire spécifique (preuve nommée) ou s'il s'agit d'un mode de preuve dépourvu de cadre juridique particulier (preuve innommée)¹⁴⁸⁴. À l'exception de l'usage de la reconnaissance faciale sur les données enregistrées au sein du fichier TAJ¹⁴⁸⁵, le recours à un système d'IA comme mode de preuve ne possède pas de régime probatoire dédié en droit commun. Pour l'utilisation de la reconnaissance faciale dans le cadre du TAJ¹⁴⁸⁶, le contrôle de l'atteinte au droit au respect de la vie privée sera réalisé au cas par cas par le juge¹⁴⁸⁷ comme c'est le cas en cas de sonorisation d'une cellule¹⁴⁸⁸.

L'utilisation des technologies à des fins probatoires par les autorités publiques dépassera le seul recours au TAJ. Les autorités répressives comptent bien utiliser à l'avenir les systèmes d'IA dans leurs activités comme en témoigne le règlement européen sur l'IA. Tout d'abord, l'annexe III du texte listant les systèmes d'IA à haut risque contient plusieurs utilisations relatives aux activités répressives¹⁴⁸⁹. Une fois le certificat européen obtenu, ces systèmes d'IA seront utilisables dans le respect des réglementations européennes et nationales. Ensuite, l'interdiction du recours à système d'IA « *d'identification biométrique à distance en temps réel dans des espaces accessibles au public à des fins répressives* »¹⁴⁹⁰, prévu par le règlement sur l'IA, est

¹⁴⁸² Il est important de préciser que le contrôle des conditions de l'atteinte au droit au respect de la vie privée diffère selon que la preuve soit produite par une partie ou demandée par le juge, ce-dernier ayant réalisé un examen de l'utilité de la mesure probatoire *a priori*. Pour plus de détails sur cette différence, voir LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P.383-385.

¹⁴⁸³ Cour de cassation, Chambre civile 1, 10 septembre 2014, n° 13-22.612 Cour de cassation, Chambre sociale, 9 novembre 2016, n° 15-10.203 Cour de cassation, Chambre sociale, 30 septembre 2020, n° 19-12.058.

¹⁴⁸⁴ Sur la distinction entre preuve nommée et preuve innommée, voir LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.*

¹⁴⁸⁵ Article R40-26 du code de procédure pénale.

¹⁴⁸⁶ Article 230-6 et suivants du code de procédure pénale.

¹⁴⁸⁷ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.* PP. 393-394.

¹⁴⁸⁸ Cour de cassation, Chambre criminelle, 17 mars 2015, n° 14-88.351.

¹⁴⁸⁹ « Règlement sur l'intelligence artificielle », 2024. Annexe III.6.

¹⁴⁹⁰ *Ibid.* Article 5.h)

accompagnée de trois exceptions¹⁴⁹¹. De plus, de nombreux projets de recherche sont actuellement en cours pour concevoir des systèmes d'IA destinés aux activités des autorités répressives au sein de l'Union européenne¹⁴⁹². Enfin, l'utilisation de traitements algorithmiques prévue dans le cadre de l'expérimentation autorisée par la loi de 2023 relative aux Jeux olympiques et paralympiques¹⁴⁹³ pourrait être généralisée à l'avenir comme un procédé probatoire de droit commun.

En outre, les futurs usages probatoires de l'IA n'ont pas besoin d'attendre un cadre juridique spécifique pour être appliqués. Bien que le recours à des nouveaux modes de preuve sans régime spécifique a été sanctionné par la CEDH¹⁴⁹⁴, la Cour de cassation accepte généralement ces derniers sur le fondement de dispositions plus générales comme le pouvoir de réquisition du Procureur de la République¹⁴⁹⁵. La Haute juridiction examine alors la nécessité et la proportionnalité de l'atteinte en observant à la fois le mode de preuve utilisé¹⁴⁹⁶ et les garanties de protection mises en place¹⁴⁹⁷.

¹⁴⁹¹ « i) la recherche ciblée de victimes spécifiques d'enlèvement, de la traite ou de l'exploitation sexuelle d'êtres humains, ainsi que la recherche de personnes disparues ; » ; « la prévention d'une menace spécifique, substantielle et imminente pour la vie ou la sécurité physique de personnes physiques ou d'une menace réelle et actuelle ou réelle et prévisible d'attaque terroriste ; » ; « la localisation ou l'identification d'une personne soupçonnée d'avoir commis une infraction pénale, aux fins de mener une enquête pénale, d'engager des poursuites ou d'exécuter une sanction pénale pour des infractions visées à l'annexe II et punissables dans l'État membre concerné d'une peine ou d'une mesure de sûreté privatives de liberté d'une durée maximale d'au moins quatre ans. » - Ibid.

¹⁴⁹² Pour un projet en cours, voir le projet STARLIGHT « *Enhancing the EU's strategic autonomy in the field of artificial intelligence (AI) for law enforcement agencies (LEAs)* » (<https://www.starlight-h2020.eu/>). Voir également les offres de projets H2020 *Fighting Crime and Terrorism 2024* (<https://www.horizon-europe.gouv.fr/fighting-crime-and-terrorism-2024-34192>)

¹⁴⁹³ « LOI n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions (1) », 2023-380, 2023.

¹⁴⁹⁴ Cour européenne des droits de l'Homme, 31 mai 2005, n° 59842. Cour européenne des droits de l'Homme, 8 février 2018, n° C-630/16.

¹⁴⁹⁵ Articles 77-1-1 et 77-1-2 du code de procédure pénale.

¹⁴⁹⁶ Le mode de preuve qui ne porte qu'une atteinte minime au droit au respect de la vie privée selon les juges est licite. Par exemple, policiers effectuant des constatations de la présence d'un véhicule dans un parking souterrain ne porte qu'une atteinte minime au droit à la vie privée : Cour de cassation, Chambre criminelle, 23 janvier 2013, n° 12-85.059.

¹⁴⁹⁷ En matière de vidéosurveillance, la Cour de cassation exige que l'autorisation octroyé par le magistrat précise le périmètre et la durée de la vidéosurveillance : LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

B- La licéité de la preuve par un système d'IA au regard de l'atteinte au droit au respect de la vie privée

Typologie des atteintes au droit à la vie privée causées par la preuve produite par un système d'IA. Le système d'IA comme mode de preuve ne peut pas être résumé à la seule reconnaissance faciale. Au-delà de minimiser le nombre d'applications possibles, cela réduit le problème de l'atteinte au respect du droit à la vie privée seulement face aux systèmes d'IA dont la finalité est la preuve. Or, l'objectif est de réfléchir à la licéité du résultat produit par un système d'IA, que celui-ci soit pensé dès le départ comme un instrument probatoire ou non. Dès lors, la notion de preuve par un système d'IA désigne une très grande diversité de résultats et d'utilisations possibles des technologies d'IA. Lister l'ensemble des cas possibles serait une démarche peu pertinente. Ainsi la preuve par un système d'IA peut représenter une atteinte au droit au respect à la vie privée de deux manières différentes : l'atteinte causée par le recours au système d'IA lors de la recherche de la preuve, d'une part, et la production d'un résultat portant atteinte au droit à la vie privée, d'autre part. Loin d'être exclusifs, ces deux types d'atteintes peuvent coexister dans une même preuve produite par un système d'IA.

L'atteinte au droit à la vie privée du fait de l'utilisation d'un système d'IA. La première cause d'atteinte au droit au respect à la vie privée est celle produite par le recours à un système d'IA. L'un des avantages des systèmes d'IA est leur capacité à analyser des quantités massives d'informations à grande vitesse. Cette capacité permet de réaliser des tâches malgré des contraintes de temps ou de ressources humaines. Par exemple, lors d'une enquête pour cyberattaque, les policiers peuvent obtenir les informations contenues dans les appareils électroniques de la victime. Ces informations représentent une quantité impressionnante de données que des êtres humains sont incapables d'analyser dans un temps raisonnable. Ainsi, pour des raisons pratiques, le recours à un système d'IA permettrait d'analyser les informations susceptibles d'identifier la date et l'origine de la cyberattaque en quelques heures, là où il faudrait des mois entiers pour une équipe humaine. Dans cet exemple, le recours à un système d'IA implique cependant un accroissement des données traitées en pratique. Si l'analyse a lieu pendant une garde à vue sur les appareils électroniques de la personne, l'atteinte à la vie privée apparaît clairement. En droit administratif, les autorités de contrôle peuvent, elles aussi, recourir

à des outils d'IA afin de faire face à la quantité de données à analyser¹⁴⁹⁸. L'augmentation quantitative de données personnelles traitées par le système d'IA implique donc une plus grande atteinte à la vie privée. Du point de vue du droit des données personnelles, l'utilisation de l'IA a un impact clair et entraîne, notamment, le besoin de réaliser une analyse d'impact relative à la vie privée¹⁴⁹⁹ conformément au RGPD¹⁵⁰⁰.

Même si le respect du droit des données personnelles n'implique pas nécessairement l'illicéité de la preuve produite¹⁵⁰¹, les deux notions ne sont pas complètement décorréées. En droit pénal par exemple, le responsable du fichier TAJ a été sanctionné à plusieurs reprises par la CEDH pour ses manquements au RGPD¹⁵⁰². En matière civile, le constat est le même. Si l'utilisation de la vidéosurveillance sur le lieu de travail peut être considérée comme une atteinte nécessaire et proportionnée¹⁵⁰³, le recours à la reconnaissance faciale ou de tout autre analyse de l'image par un système d'IA peut être considéré comme une atteinte disproportionnée. Par conséquent, la preuve produite par un système d'IA peut être illicite car le procédé probatoire qui a permis de la produire porte une atteinte non nécessaire ou disproportionnée au droit au respect de la vie privée de la personne contre laquelle elle est produite.

L'atteinte au droit à la vie privée causée par un résultat produit par un système d'IA. La seconde cause d'atteinte au droit au respect à la vie privée est celle produite par l'information que révèle le résultat. Par exemple, c'est le cas du résultat d'un profilage réalisé par un système d'IA sur une des parties ou un tiers au litige. Le profilage est entendu ici au sens du règlement (UE) 2016/679 (RGPD ci-après), à savoir :

« Toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données à caractère personnel pour évaluer certains aspects personnels relatifs à une personne physique, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la

¹⁴⁹⁸ En matière commerciale par exemple, l'autorité de contrôle prudentiel et de résolution (ACPR) a recours à un système d'IA, « LUCIA », lors de certaines de ses missions de contrôle. Voir Jérôme Lasserre-Capdeville, « Autorité de contrôle prudentiel et de résolution : supervision bancaire - Pouvoirs de l'ACPR » (2023) (Répertoire de droit commercial) .

¹⁴⁹⁹ Voir B— Le principe de vie privée.

¹⁵⁰⁰ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 35. « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Article 62.

¹⁵⁰¹ Cour de cassation, Chambre sociale, 25 novembre 2020, n° 17-19.523.

¹⁵⁰² Sur la durée de conservation, voir *MK c/ France*, 2013 Cour européenne des droits de l'Homme ainsi que Cour européenne des droits de l'Homme, 22 juin 2017, n° 8806/12. Sur les droits des personnes concernées : Cour européenne des droits de l'Homme, 18 septembre 2014, n° 21010/10.

¹⁵⁰³ Cour de cassation, Chambre sociale, 2 février 2011, n° 10-14.263.

*situation économique, la santé, les préférences personnelles, les intérêts, la fiabilité, le comportement, la localisation ou les déplacements de cette personne physique ; »*¹⁵⁰⁴

Les conséquences du profilage sont limitées par le RGPD et la loi de 1978 relative à l'informatique, aux fichiers et aux libertés¹⁵⁰⁵ (LIL ci-après) qui interdisent toute décision critique¹⁵⁰⁶ prise sur le seul fondement du profil établi à l'aide des données personnelles de l'individu¹⁵⁰⁷. Trois exceptions¹⁵⁰⁸ sont prévues à cette interdiction : le consentement explicite de la personne concernée, la nécessité contractuelle et la prise de décision administrative individuelle¹⁵⁰⁹. Prises ensemble, ces exceptions représentent une très grande quantité et variété de situations dans lesquelles le profilage d'un individu est autorisé par la réglementation relative aux données personnelles. Ce sont autant de preuves relatives à un aspect personnel qui pourraient être produites par le détenteur du résultat (le déployeur du système d'IA et/ou l'individu évalué) ou demandées par le juge. Toutefois, la réutilisation de ces résultats de systèmes d'IA comme preuve semble rencontrer deux obstacles.

Le premier est un obstacle législatif. L'article 47 de la LIL interdit également de fonder une décision de justice sur une évaluation automatisée de certains aspects de la personnalité affectée par la décision de justice :

« Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement

¹⁵⁰⁴ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 4.4)

¹⁵⁰⁵ « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Modifié par « Ordonnance n° 2018-1125 du 12 décembre 2018 prise en application de l'article 32 de la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles et portant modification de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel », 2018. Article 1.

¹⁵⁰⁶ À nouveau le terme « critique » est employé ici pour désigner l'existence d'une condition de gravité de la décision dans la disposition. Dans cette situation, cela désigne la « *décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative* » - Article 22 du RGPD et article 47 de la LIL.

¹⁵⁰⁷ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 22.1. « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Article 47.

¹⁵⁰⁸ Deux exceptions sont imposées par le règlement européen tandis que la troisième exception concernant les décisions administratives individuelles a été ajoutée à l'initiative du législateur français.

¹⁵⁰⁹ À l'exception des décisions relatives à des recours administratifs. Voir « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Article 47 alinéa 5.

*automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne »*¹⁵¹⁰

Contrairement à l'alinéa 2 mentionné précédemment, l'alinéa 1 de l'article 47 ne contient pas la formulation « *sur le seul fondement* » lorsqu'il se réfère à la décision prise. Par conséquent, la présence ou non d'un intervenant humain dans le raisonnement de la décision de justice n'est pas une condition à cette interdiction. Toutefois, l'article précise que l'évaluation doit porter sur la personne faisant l'objet de la décision de justice. L'usage d'un profilage produit par un système d'IA concernant un tiers au litige ou la victime dans un procès pénal ne semble donc pas interdite. L'effet de cette interdiction sur le droit à la preuve est aujourd'hui encore flou, en l'absence de décisions de justice portant sur le sujet. Pour autant, il est intéressant de se questionner sur l'étendue des conséquences de cet alinéa sur le droit à la preuve. Interprétée strictement, la disposition entraîne l'interdiction de confier à un système automatisé (le plus souvent un système d'IA au sens du règlement européen sur l'IA) l'évaluation du comportement de l'individu lorsque ce comportement est déterminant pour la décision de justice. Par exemple, un score d'agressivité donné par un système d'IA sur un individu ne pourrait pas être utilisé par le juge pour qualifier l'existence de violences conjugales. Néanmoins, une interprétation plus large de la disposition conduit à exclure tout profilage produit par un système d'IA offert comme élément de preuve fourni dans le dossier probatoire. De la même manière, cela signifie que le juge ne peut pas demander la production aux parties ou à des tiers un profilage qu'ils auraient réalisé sur un ou plusieurs aspects de la personnalité des parties. La disposition relative à la vie privée se rapproche alors d'une règle d'admissibilité de la preuve spécifique aux preuves par un système d'IA, plus particulièrement le profilage par un système d'IA.

L'évaluation de l'atteinte au droit à la vie privée par le résultat d'un système d'IA est le second obstacle à sa réutilisation en tant que preuve judiciaire¹⁵¹¹. Le résultat d'un système d'IA, en particulier les évaluations de comportements et de personnalité, peut représenter une information intime sur la personne. Le profil d'un client dépendant qu'un site de paris en ligne pourrait dresser d'un internaute habitué ou le profil de « femme enceinte » qu'un supermarché peut déduire des achats d'une consommatrice sont des informations qui portent une atteinte considérable à la vie privée des personnes concernées. Tout d'abord, la nécessité de l'atteinte

¹⁵¹⁰ « *Aucune décision de justice impliquant une appréciation sur le comportement d'une personne ne peut avoir pour fondement un traitement automatisé de données à caractère personnel destiné à évaluer certains aspects de la personnalité de cette personne.* » - « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Article 47 alinéa 1.

¹⁵¹¹ Cet obstacle est étudié indépendamment du premier relatif à l'article 47 alinéa 1 de la LIL.

peut être discutée. Dans de nombreuses situations, les données d'entrées du système d'IA ayant produit la preuve peuvent être elles-mêmes des éléments de preuve pertinents tout en étant moins attentatoires au droit à la vie privée. Le juge et/ou l'expert judiciaire pourra alors être chargé de produire une analyse manuelle de ces données si cela s'avère pertinent. Le caractère « nécessaire » ou « indispensable » de l'atteinte que représente le résultat du système d'IA est alors absent. Ensuite la preuve par un système d'IA sera rarement produite seule au dossier. Au regard du caractère probabiliste du résultat produit par un système d'IA, il est probable que la partie soumette d'autres preuves en vue d'étayer son dossier. Si le juge est déjà convaincu par les autres preuves, le résultat devient alors un élément de preuve superfétatoire. Pour être nécessaire, il faudrait que la preuve produite par le système d'IA soit déterminante pour le litige et qu'elle ne puisse pas être produite de façon raisonnable par un expert judiciaire ou par le raisonnement du juge. Quant à la proportionnalité de l'atteinte, elle dépend de la finalité poursuivie ou de l'infraction en cause, il est donc très difficile de prédire, sans disposition ou jurisprudence spécifique, des aspects précis sur la manière dont la proportionnalité des preuves produites par un système d'IA sera évaluée au cas par cas par les juges.

§2. Le respect des droits de la défense

Annnonce de plan. Les droits de la défense trouvent leur fondement dans l'article 6 de la Conv. EDH relatif au droit à un procès équitable¹⁵¹². Le lien étroit entre ces deux notions est rappelé par la jurisprudence de la CEDH qui affirme que l'évaluation du caractère équitable d'une procédure passe généralement par la vérification du respect des droits de la défense¹⁵¹³. Les droits de la défense apparaissent alors comme un mécanisme d'exclusion des preuves¹⁵¹⁴. La CEDH évalue le respect à ce principe en se fondant sur plusieurs critères.

¹⁵¹² « Toute personne a droit à ce que sa cause soit entendue équitablement, publiquement et dans un délai raisonnable, par un tribunal indépendant et impartial, établi par la loi, qui décidera, soit des contestations sur ses droits et obligations de caractère civil, soit du bien-fondé de toute accusation en matière pénale dirigée contre elle » - CONSEIL DE L'EUROPE, « Convention européenne des droits de l'Homme », 1950. Article 6.

¹⁵¹³ « Pour déterminer si la procédure dans son ensemble a été équitable, il faut aussi se demander si les droits de la défense ont été respectés. » - Cour européenne des droits de l'Homme, 10 mars 2009, n° 4378/02. Cour européenne des droits de l'Homme, 5 novembre 2002, n° 48539/99.

¹⁵¹⁴ Jusqu'à présent, les droits de la défense n'avaient été abordé qu'au travers du droit à la preuve, mécanisme juridique favorisant la recherche de la vérité par les parties. Ici, c'est l'aspect inhibiteur de la recherche de la vérité qui sera étudié au regard des preuves par un système d'IA. Pour plus d'informations sur l'ambivalence des droits de la défense en matière probatoire, voir LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. PP. 408-409.

Dans un premier temps, elle vérifie que le « *requérant s'est vu offrir la possibilité de remettre en question l'authenticité de l'élément de preuve et de s'opposer à son utilisation.* »¹⁵¹⁵. Aujourd'hui acquis et intégré dans toutes les procédures en droit français¹⁵¹⁶, l'obligation de communication des preuves au requérant, composante du principe du contradictoire, semble entrer en opposition avec l'opacité importante des systèmes d'IA **(A)**. Dans un second temps, les juges européens examinent « *la qualité de l'élément de preuve, y compris le point de savoir si les circonstances dans lesquelles il a été recueilli font douter de sa fiabilité ou de son exactitude.* »¹⁵¹⁷. Entre les hypertrucages et les résultats probabilistes incertains, le critère de la qualité de la preuve devient un angle d'analyse essentiel du respect des droits de la défense par la preuve produite par un système d'IA **(B)**.

A- L'opacité des systèmes d'IA face au principe du contradictoire

Présentation du principe du contradictoire. En droit européen, le principe du contradictoire est défini par la CEDH comme la garantie de « *la faculté pour les parties à un procès, pénal ou civil, de prendre connaissance de toute pièce ou observation présentée au juge, même par un magistrat indépendant, en vue d'influencer sa décision, et de la discuter* »¹⁵¹⁸. Au sein du droit français, le principe est codifié dans les dispositions liminaires du code de procédure civile comme un des principes directeurs du procès¹⁵¹⁹. Tout d'abord, le principe implique que les parties doivent être informées de l'existence de l'instance¹⁵²⁰. Aussi bien en droit civil qu'en droit pénal¹⁵²¹, il impose que les éléments de preuves utilisés soient communiqués entre les parties¹⁵²². Enfin le principe du contradictoire s'impose au juge qui doit le faire respecter, en dépit du silence des parties, sous peine de violer lui-même ledit principe¹⁵²³. Par conséquent, il

¹⁵¹⁵ Cour européenne des droits de l'Homme, 10 mars 2009, n° 4378/02. Par. 90.

¹⁵¹⁶ Code de procédure civile, articles 14 à 17. Il s'agit d'un principe fondamental du procès civil : FERRAND Frédérique, « Preuve », *Répertoire de procédure civile*, Dalloz, 2024. En matière pénale, il est inscrit à l'article 427 alinéa 2 du code de procédure pénale.

¹⁵¹⁷ Cour européenne des droits de l'Homme, 10 mars 2009, n° 4378/02. Par. 90.

¹⁵¹⁸ Cour européenne des droits de l'Homme, 20 février 1996, n° 19075/91.

¹⁵¹⁹ Articles 14 à 17 du code de procédure civile.

¹⁵²⁰ Article 14 du code de procédure civile.

¹⁵²¹ En droit pénal, le ministère public est dans l'obligation à la fois les preuves à charge mais aussi les preuves à décharge en sa possession : « *De surcroît, l'article 6 § 1 exige [...] que les autorités de poursuite communiquent à la défense toutes les preuves pertinentes en leur possession, à charge comme à décharge* » - Cour européenne des droits de l'Homme, 16 février 2000, n° 27052/95. Par. 51.

¹⁵²² Article 15 du code de procédure civile et article 427 alinéa 2 du code de procédure pénale.

¹⁵²³ Article 16 du code de procédure civile et article 427 alinéa 2 du code de procédure pénale.

doit rejeter toute preuve n'ayant pas été discutée contradictoirement, y compris les preuves communiquées trop tardivement pour pouvoir être débattues¹⁵²⁴.

Néanmoins, le droit à la divulgation des preuves pertinentes n'est pas absolu¹⁵²⁵. Il doit être mis en balances avec des intérêts concurrents comme la sécurité nationale, la sauvegarde d'un intérêt public important ou la protection des témoins¹⁵²⁶. L'atteinte aux droits de la défense doit être justifiée par son caractère absolument nécessaire¹⁵²⁷ et doit être accompagnée de compensations procédurales¹⁵²⁸.

Le principe du contradictoire face à l'opacité des systèmes d'IA. Pour rappel, l'opacité des technologies d'IA est constituée de trois couches différentes : le secret, la complexité technique et l'opacité par conception¹⁵²⁹. Chacune de ces couches constitue un défi pour le respect du principe du contradictoire.

Premièrement, le secret est un obstacle important dans le cadre d'une preuve produite par un système d'IA. La communication du système d'IA ayant produit l'élément de preuve peut porter atteinte à des secrets protégés par la loi. En particulier, les informations relatives au fonctionnement d'un système d'IA peuvent être protégés par le secret des affaires, consacré législativement en droit européen depuis la directive (UE) 2016/943¹⁵³⁰ et transposé en droit français par la loi 2018-670 relative à la protection du droit des affaires¹⁵³¹. Dans le contexte de l'IA, la tension entre l'accès aux preuves pertinentes et le secret des affaires a été illustré par l'arrêt *State v. Loomis* rendue aux États-Unis en 2017 où la Cour suprême a refusé l'accès au code-source de COMPAS, un système algorithmique litigieux dans l'affaire, en raison de son caractère propriétaire¹⁵³². En droit français, l'atteinte au secret des affaires est autorisée

¹⁵²⁴ Civ. 2e, 9 mai 1983, cité dans FERRAND Frédérique, « Preuve », *Répertoire de procédure civile*, Dalloz, 2024.

¹⁵²⁵ Cour européenne des droits de l'Homme, 16 février 2000, n° 27052/95. Par. 52.

¹⁵²⁶ *Ibid.* Sur la mise en balance avec la protection de témoins contre des représailles, voir Cour européenne des droits de l'Homme, 26 mars 1996, n° 20524/92. Par. 70.

¹⁵²⁷ Cour européenne des droits de l'Homme, 16 février 2000, n° 27052/95.

¹⁵²⁸ Cour européenne des droits de l'Homme, 26 mars 1996, n° 20524/92. Par. 72. Cour européenne des droits de l'Homme, 20 novembre 1989, n° 62171/00. Par. 43.

¹⁵²⁹ BURRELL Jenna, « How the machine 'thinks' », *Big Data & Society*, 3, SAGE Publications Ltd, 2016. Pour plus de détails, voir A– Le principe de transparence.

¹⁵³⁰ « Directive (UE) 2016/943 du Parlement européen et du Conseil du 8 juin 2016 sur la protection des savoir-faire et des informations commerciales non divulgués (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

¹⁵³¹ « LOI n° 2018-670 du 30 juillet 2018 relative à la protection du secret des affaires (1) », 2018-670, 2018.

¹⁵³² *State v Loomis*, 2017 Cour suprême des États-Unis. Pour d'autres exemples, voir WEXLER Rebecca, « Life, Liberty, and Trade Secrets », 70, 2018.

lorsqu'elle est nécessaire et proportionnée¹⁵³³. Cette mise en équilibre avec le droit à la preuve se manifeste au travers de la procédure de mise sous séquestre¹⁵³⁴. Pendant cette étape provisoire, le juge évalue la nécessité de chacune des preuves dont l'accès est demandé¹⁵³⁵ (au travers d'une demande de production forcée ou une mesure d'instruction), puis nomme individuellement les personnes pouvant y avoir accès¹⁵³⁶. La mise sous séquestre fait l'objet d'une jurisprudence abondante sur les critères d'évaluation du juge, dont la prise en compte des propositions du bénéficiaire du secret¹⁵³⁷ et la limitation aux faits litigieux et aux personnes concernées¹⁵³⁸. Bien qu'applicable aux systèmes d'IA, le secret des affaires est un obstacle traditionnel pour le respect des droits de la défense et ne représente en rien un défi particulier dans le cadre d'un système d'IA comme procédé de preuve. Ainsi, les dispositions concernant la divulgation des éléments de preuves pertinents, énoncées par les propositions de directives européennes applicables au régime de responsabilité du fait de l'IA¹⁵³⁹, ne représentent pas des changements importants au regard du respect du principe du contradictoire. De plus, cette couche d'opacité ne présente pas de lien avec la fiabilité du système d'IA et, par conséquent, ne s'articule pas avec les caractéristiques techniques imposées par le règlement européen sur l'IA.

Deuxièmement, l'opacité des systèmes d'IA s'explique par la complexité technique des technologies utilisées. Dans une logique similaire à celle appliquée pour le secret des affaires, il convient d'écarter la complexité technique de l'analyse. D'une part, la difficulté pour la

¹⁵³³ « Si le secret des affaires ne fait pas en lui-même obstacle au prononcé d'une mesure d'instruction préventive ainsi qu'à la production forcée de pièces, c'est à la condition qu'elles soient nécessaires, donc qu'elles constituent le seul moyen de préserver les droits du demandeur » ; « Qu'en se déterminant ainsi, sans rechercher, comme il le lui incombait, si cette mesure d'instruction, confiée à un tiers soumis au secret professionnel, n'était pas proportionnée au droit des sociétés Allianz d'établir la preuve d'actes de concurrence interdite ou déloyale attribués à l'agent général et à la préservation des secrets d'affaires des sociétés GAN, la cour d'appel n'a pas donné de base légale à sa décision ; » - Cour de cassation, Chambre civile 1, 22 juin 2017, n° 15-27.845.

¹⁵³⁴ « Décret n° 2018-1126 du 11 décembre 2018 relatif à la protection du secret des affaires », 2018-1126, 2018. Codifié à l'article R. 15-1 du code du commerce.

¹⁵³⁵ Cet examen est réalisé à la lumière des preuves en question et du mémoire remis par la partie bénéficiaire du secret des affaires. Article R. 153-3 du code de commerce.

¹⁵³⁶ Articles R. 153-3 à R. 153-6 du code de commerce.

¹⁵³⁷ Cour de cassation, Chambre commerciale, 5 juin 2019, n° 17-22.192.

¹⁵³⁸ Cour de Cassation, Chambre civile 2, 8 février 2006, n° 05-14.198.

¹⁵³⁹ « Les juridictions nationales limitent la divulgation d'éléments de preuve à ce qui est nécessaire et proportionné pour étayer une action ou une action potentielle en réparation, et la conservation de ces éléments de preuve à ce qui est nécessaire et proportionné pour étayer une telle action en réparation. » - Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 28 septembre 2022. Article 3.4. « Les États membres veillent à ce que les juridictions nationales limitent la divulgation des éléments de preuve à ce qui est nécessaire et proportionné [...] » - Proposition de directive du Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 28 septembre 2022. Article 8.2.

personne non experte à comprendre un élément de preuve ne pose aucun problème à ce que celle-ci soit divulguée¹⁵⁴⁰. Le juge ou la partie qui reçoit l'élément de preuve technique complexe peut recourir à un expert, qui viendra alors lui expliquer les implications de ladite preuve si nécessaire. D'autre part, la complexité technique, à savoir la difficulté de compréhension de l'information pour une personne non initiée, n'est pas particulièrement liée avec les caractéristiques techniques des systèmes d'IA.

La transparence du procédé de preuve et le principe du contradictoire. La troisième et dernière couche d'opacité que peut présenter un système d'IA est son opacité intrinsèque, aussi appelée l'opacité par la conception. Spécifique aux technologies d'IA, l'opacité intrinsèque peut rendre extrêmement difficile l'accès à des informations relatives au logiciel, et impossible la compréhension de son fonctionnement. En tant que procédé de preuve, le manque de transparence du système d'IA peut constituer une atteinte importante à la capacité de la partie à discuter et contester son résultat comme élément de preuve. Le principe du contradictoire peut exiger la connaissance très précise des caractéristiques techniques du système d'IA en question. Les variables sélectionnées, les paramètres de la prise de décision et leurs poids sont des choix du fournisseur (et/ou du déployeur le cas échéant) qui peuvent entraîner des conséquences déterminantes pour la licéité du résultat produit.

L'importance de la transparence du procédé de preuve pour garantir le respect du principe du contradictoire a été mis en lumière par la Cour de justice de l'UE lors d'un arrêt de 2019 opposant la Commission européenne à UPS et Fedex¹⁵⁴¹. La Cour de justice a annulé la sanction de la Commission européenne parce que celle-ci n'avait pas communiqué le modèle économétrique qu'elle avait utilisé à l'autre partie¹⁵⁴². Plus précisément, la Commission avait modifié son modèle entre deux étapes de l'analyse économétrique et n'a communiqué que le modèle avant la modification. Ladite modification a été qualifiée comme non négligeable par le tribunal de l'Union européenne qui a alors caractérisé l'atteinte au droit de la défense du fait de la non-communication de la version finale du modèle économétrique par la Commission européenne¹⁵⁴³. Or, cette modification consistait en un changement de nature d'une unique variable, la Commission ayant transformé une variable discrète en une variable continue. Par

¹⁵⁴⁰ Par exemple, la Commission européenne doit communiquer à la partie adverse le modèle économétrique utilisé durant l'exercice de son contrôle de concentration. Cour de justice de l'Union européenne, 16 janvier 2019, n° C-265/17.

¹⁵⁴¹ *Ibid.*

¹⁵⁴² *Ibid.*

¹⁵⁴³ Tribunal de l'Union européenne, 7 mars 2017, n° T-194/13.

analogie, il apparaît clairement que le niveau de transparence d'un système d'IA est un facteur déterminant pour le respect du principe du contradictoire, et donc pour la licéité de son résultat. L'utilisation d'un élément de preuve produit par un système d'IA dont le fonctionnement opaque ne permet pas la communication d'informations pertinentes sur son fonctionnement à la partie adverse pourrait alors constituer une atteinte significative auxdits droits dont la nécessité absolue sera difficilement justifiable.

La transparence du fonctionnement du système d'IA. Les droits de la défense imposent donc un niveau de transparence élevé dans les techniques probatoires utilisées, opérant ainsi un lien avec les obligations de transparence contenues dans le règlement sur l'IA. L'articulation entre le règlement européen et le principe du contradictoire se situe au niveau des informations disponibles en raison des exigences de la législation sur l'IA. La documentation technique, le code-source ou encore des informations relatives aux jeux de données de conception, d'entrées ou au déploiement du système d'IA à haut risque sont des renseignements pertinents dont la communication pourrait s'avérer nécessaire au regard du principe du contradictoire. Toutefois, il est possible que d'autres informations doivent être communiquées à la partie adverse concernant le système d'IA ayant produit l'élément de preuve. Dans le cas où la transparence du fonctionnement du logiciel ne serait pas suffisante pour connaître les informations nécessaires demandées, l'opacité intrinsèque de certains systèmes d'IA à haut risque pourrait alors devenir un obstacle à l'utilisation, prévue ou non, de leurs résultats comme éléments de preuve. Pour les systèmes d'IA à haut risque spécifiquement destinés à une fonction probatoire, les exigences du règlement sur l'IA en lien avec la transparence devront donc être interprétés en prenant en compte le principe du contradictoire. Enfin, ce besoin de transparence pourrait également être un obstacle insurmontable pour la licéité de résultats produits par des systèmes d'IA autres qu'à haut risque, au sens du règlement européen sur l'IA, comme éléments de preuve¹⁵⁴⁴.

B- La contestation de la fiabilité des preuves par un système d'IA

L'authenticité de la preuve par un système d'IA : le cas des hypertrucages. L'objectif premier de l'obligation de communication des preuves entre les parties est de permettre à

¹⁵⁴⁴ Afin de relativiser cette conclusion, il faut admettre que l'emploi de ce type de résultat comme preuve importante dans un litige risque d'être bien plus rare que pour les systèmes d'IA à haut risque.

chacune d'entre elles de les discuter, notamment d'en contester l'authenticité et de s'opposer à leur utilisation¹⁵⁴⁵. La question de l'authenticité des preuves par un système d'IA amène naturellement notre analyse sur le problème des hypertrucages (ou *deep fakes* en anglais). En effet, l'hypertrucage est défini par le règlement européen sur l'IA comme :

*« une image ou un contenu audio ou vidéo généré ou manipulé par l'IA, présentant une ressemblance avec des personnes, des objets, des lieux, des entités ou événements existants et pouvant être perçu à tort par une personne comme authentiques ou véridiques »*¹⁵⁴⁶

Les systèmes d'IA capables de générer du contenu (vidéo, audio ou image) représentent un danger pour les droits fondamentaux¹⁵⁴⁷, allant de la propagation de fausses informations au service d'une opération de manipulation massive¹⁵⁴⁸ au montage humiliant publié sur Internet¹⁵⁴⁹. Du point de vue de la recherche de la vérité, ce nouveau type de trucage inquiète car il est plus difficile à détecter que les trucages ordinaires. De plus, l'apparition des hypertrucages s'inscrit dans un mouvement préoccupant de « *démocratisation de la fraude* »¹⁵⁵⁰ dans la preuve, permis notamment par l'intégrité intrinsèquement faible des preuves électroniques par rapport aux preuves matérielles¹⁵⁵¹.

Au niveau européen, un encadrement légal des hypertrucages émerge avec le règlement européen sur l'IA qui oblige spécifiquement les déployeurs de systèmes d'IA capables de produire des hypertrucages d'indiquer que le contenu a été généré ou manipulé artificiellement¹⁵⁵². Le texte européen précise que cette information doit être fournie de « *manière claire et reconnaissable au plus tard au moment de la première interaction ou de la première exposition* »¹⁵⁵³. En droit français, la mention obligatoire du caractère truqué du contenu publié était déjà imposée par l'article 226-8 du code pénal qui sanctionne d'un an

¹⁵⁴⁵ Cour européenne des droits de l'Homme, 10 mars 2009, n° 4378/02.

¹⁵⁴⁶ « Règlement sur l'intelligence artificielle », 2024. Article 3.60)

¹⁵⁴⁷ « Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux », Commission nationale consultative des droits de l'Homme, 2022.

¹⁵⁴⁸ PARLEMENT EUROPÉEN, « Les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle », 2020. Par. 17.

¹⁵⁴⁹ Fabien Soyeux, « Le porno deepfake explose, et c'est un vrai problème », (5 November 2022), online: *CNET France* <<https://www.cnetfrance.fr/news/le-porno-deepfake-explose-et-c-est-un-vrai-probleme-39949326.htm>>.

¹⁵⁵⁰ Mustapha MEKKI lors d'une conférence à Montréal le 25 septembre 2019 : *Loi + preuve dans 10, 20, 30 ans...*, 2020 : <https://www.youtube.com/watch?v=9CoweIO-RkE>

¹⁵⁵¹ Sur le sujet de manière générale, voir GAUTRAIS Vincent et GINGRAS Patrick, « La preuve des documents technologiques », *Les Cahiers de propriété intellectuelle*, 22, Les cahiers de propriété intellectuelle Inc., 2017 VIAL Géraldine, VERGÈS Étienne et GAUTRAIS Vincent, « Preuves scientifiques et technologiques », *Cahiers Droit, Sciences & Technologies*, PUP, 2019.

¹⁵⁵² « Règlement sur l'intelligence artificielle », 2024. Article 52.3.

¹⁵⁵³ *Ibid.* Article 52.5.

d'emprisonnement et 15.000€ d'amende tout montage dissimulé d'une personne sans son consentement. Au travers de la récente loi visant à sécuriser et réguler l'espace numérique, le législateur français prévoit de sanctionner spécifiquement les montages à caractère sexuel¹⁵⁵⁴. De plus, l'usage d'un traitement algorithmique pour réaliser cette infraction devient une circonstance aggravante entraînant une sanction pouvant aller jusqu'à trois ans d'emprisonnement et 75.000€ d'amende¹⁵⁵⁵.

Du point de vue du droit de la preuve, l'apparition des trucages produits par des systèmes d'IA ne provoque pas de bouleversement particulier. En dehors de la difficulté importante à les détecter, les hypertrucages demeurent sur le principe des éléments de preuves trompeurs comme cela existe depuis toujours. La différence principale que peut entraîner la multiplication des hypertrucages est un recours beaucoup plus fréquent à une expertise judiciaire¹⁵⁵⁶ pour attester de l'authenticité des contenus apportés comme preuve¹⁵⁵⁷.

La qualité de la preuve produite par un système d'IA. Le second élément pris en compte par la CEDH pour vérifier que les droits de la défense ont été respectés est la qualité de la preuve¹⁵⁵⁸. Les juges européens précisent ainsi que plus la fiabilité de la preuve est faible, plus il existe un besoin de la corroborer avec d'autres éléments de preuve pour respecter le droit à un procès équitable sans pour autant que cela soit une nécessité¹⁵⁵⁹. Cette logique explique le rejet par la Cour de cassation des auditions sous hypnose comme élément de preuve¹⁵⁶⁰. De la même manière, l'autorisation du procureur de la République pour réaliser des examens techniques est considérée par la chambre criminelle comme une obligation garantissant « la

¹⁵⁵⁴ « LOI n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique (1) », 2024-449, 2024. Article 21. Codifié à l'article 226-8-1 du code pénal.

¹⁵⁵⁵ *Ibid.* Voir également GROFFE-CHARIER Julie, « Vers un encadrement légal des deepfakes », *Dalloz actualités*, 2023.

¹⁵⁵⁶ Mustapha MEKKI : note 305.

¹⁵⁵⁷ Par exemple, le tribunal judiciaire de Paris a soumis le contenu d'une vidéo au département « Signal Image et Parole » de l'Institut de recherche criminelle de la Gendarmerie nationale qui a écarté la possibilité d'un hypertrucage soulevé par le défendeur. Tribunal judiciaire de Paris (17e ch.), 15 septembre 2022, n° 20304000328, Rachel K. et a. c/ Dieudonné M. LÉGIPRESSE, « Dieudonné condamné pour injure à caractère racial : les propos lui sont imputables, bien qu'il invoque avoir été victime d'un « deepfake » », 2022.

¹⁵⁵⁸ Cour européenne des droits de l'Homme, 10 mars 2009, n° 4378/02. Par. 90.

¹⁵⁵⁹ « Si un problème d'équité ne se pose pas nécessairement lorsque la preuve obtenue n'est pas corroborée par d'autres éléments, il faut noter que lorsqu'elle est très solide et ne prête à aucun doute, le besoin d'autres éléments à l'appui devient moindre » - Cour européenne des droits de l'Homme, 5 novembre 2002, n° 48539/99. Par. 43. Cour européenne des droits de l'Homme, 12 mai 2000, n° 35394/97. Par. 35 à 37.

¹⁵⁶⁰ Cour de Cassation, Chambre criminelle, 12 décembre 2000, n° 00-83.852. Cour de Cassation, Chambre criminelle, 28 novembre 2001, n° 01-86.467.

fiabilité de la recherche et de l'administration de la preuve »¹⁵⁶¹ et son manquement est une circonstance créant un doute important sur leur fiabilité ou leur exactitude.

Du fait de sa nature statistique, l'exactitude du résultat produit par un système d'IA n'est pas et ne sera jamais certaine. Cependant, toutes les techniques de preuves présentent une partie de doute¹⁵⁶² et c'est l'agrégation d'éléments de preuves qui, au travers notamment du mécanisme des présomptions de faits, parvient à l'éliminer. D'autres modes de preuves techniques produisant des résultats statistiques sont acceptés en justice¹⁵⁶³ et même si la fiabilité d'une preuve par un système d'IA peut être relative, elle ne peut pas être comparée avec un processus probatoire comme l'hypnose. Par conséquent, il n'est pas envisageable de considérer que le résultat produit par un système d'IA soit, par sa seule nature probabiliste, une atteinte par principe aux droits de la défense. Toutefois, il n'est pas exclu que le niveau d'exactitude particulièrement faible (inférieur ou égale à 50% par exemple) d'un système d'IA dans un contexte donné soit considéré comme un critère mettant en doute la fiabilité de la recherche ou de la collecte d'éléments de preuve par ledit système d'IA.

Sans entrer dans le périmètre de la force probante, il est important de discuter des circonstances qui pourraient créer un doute significatif sur l'exactitude d'une preuve par un système d'IA qui, apportée seule, pourrait représenter un manquement aux droits de la défense. Il convient de distinguer entre, d'un côté, l'exactitude du système d'IA et du résultat produit et, de l'autre côté, des circonstances dans lesquelles le résultat a été généré. Si le second élément peut être analysé au regard des droits de la défense, le premier relève davantage de l'appréciation de sa force probante par le juge.

Il s'agit alors d'explorer prospectivement les circonstances qui pourraient amener à porter atteinte à l'exercice des droits de la défense. Sans prétendre à l'exhaustivité de ces situations, un cas nous semble particulièrement problématique : la production d'un résultat d'un système d'IA sans pouvoir apporter les données d'entrée utilisées. Cette absence peut s'expliquer par la perte desdites données, que celle-ci soit accidentelle (problème technique par exemple) ou

¹⁵⁶¹ Cour de cassation, Chambre criminelle, 18 juin 2019, n° 19-80.105. Sur l'idée d'un principe de fiabilité en droit pénal, voir DELAGE Pierre-Jerôme, « Vers un principe de fiabilité de la preuve pénale ? », RSC, 2019.

¹⁵⁶² « Aucune technique de preuve ne permet d'atteindre une certitude absolue. Toutes intègrent, à un degré plus ou moins élevé, une part de probabilité » - LLORENS-FRAYSSE Françoise, *La présomption de faute dans le contentieux administratif de la responsabilité* [microfiche], Librairie générale de droit et de jurisprudence, 1985. Cité dans GROZDANOVSKI Ljupcho, « Le probable, le plausible et le vrai. Contribution à la théorie générale de la présomption en droit », *Revue interdisciplinaire d'études juridiques*, 84, Université Saint-Louis - Bruxelles, 2020. P.44.

¹⁵⁶³ L'expertise biologique de paternité, la comparaison d'empreintes digitales ou autres données biométriques par exemple.

volontaire. La perte volontaire peut résulter d'une intention malveillante ou bien d'une non-conservation des données d'entrée du fait de leur inutilité une fois le résultat produit¹⁵⁶⁴. La non-communication des données d'entrée crée un doute considérable sur la fiabilité du résultat produit. L'utilisation de données erronées ou lacunaires affecte nécessairement la qualité du résultat produit. Pour reprendre à nouveau la maxime de Domat :

*« La certitude ou l'incertitude des présomptions [...] dépend de la certitude ou l'incertitude des faits dont on tire les présomptions, et de la justesse des conséquences qu'on tire de ces faits pour la preuve de ceux dont il s'agit. »*¹⁵⁶⁵

En suivant cette logique, l'absence des faits fondant la présomption statistique qu'est le résultat du système d'IA empêche clairement la partie adverse de discuter et contester équitablement de son exactitude et de sa fiabilité.

CONCLUSION DE LA SECTION II :

Sans pour autant être absentes, les interactions entre le fonctionnement d'un système d'IA et la recevabilité de ces résultats en tant que preuves se révèlent minimes. Tout d'abord, la distinction entre la légalité du procédé de preuve et la licéité de l'élément de preuve produit réduit grandement les conséquences que pourra avoir le règlement européen sur l'IA concernant l'évaluation de la recevabilité de ces résultats par le juge. En l'absence de futures règles de preuves (ou décisions de justice) spécifiques à la recevabilité de la preuve par IA, il n'est pas possible d'assurer que la preuve produite par un système d'IA à haut risque non certifié (ou certifié à tort) soit exclue des débats par le juge. Ce dernier évaluera la licéité de la preuve de façon usuelle au regard du respect des principes fondamentaux du droit de la preuve. Le respect de ces principes dépend peu du fonctionnement du procédé probatoire en lui-même mais davantage de la manière dont les acteurs l'ont utilisé. Par conséquent, l'articulation avec le règlement européen sur l'IA s'en trouve grandement réduit. Cependant, l'opacité et la qualité de fonctionnement des systèmes d'IA entrent en tension avec les exigences des droits de la défense et certaines dispositions spécifiques du droit à la vie privée semblent limiter l'utilisation de certains résultats de systèmes d'IA comme élément de preuve. De façon globale, le règlement européen sur l'IA, et plus généralement les caractéristiques de fonctionnement du système d'IA,

¹⁵⁶⁴ Cependant, les droits d'accès aux données personnelles, octroyés par différents textes européens et nationaux (Articles 15 du RGPD et L. 311-1 et suivants du CRPA entre autres) entraînent une obligation de conservation desdites données d'entrées.

¹⁵⁶⁵ J. DOMAT, *Œuvres complètes*, Paris, Firmin Didot, 1828, P. 166, cité par GROZDANOVSKI Ljupcho, *op. cit.* P. 49.

ne devraient pas avoir de conséquences significatives sur la licéité de ses résultats utilisés comme élément de preuve.

CONCLUSION DU CHAPITRE 1 :

L'évaluation de l'admissibilité des preuves produites par un système d'IA dévoile l'existence d'interactions entre les conditions d'admissibilité en droit de la preuve et le fonctionnement d'un système d'IA. Sans surprise, le résultat d'un système d'IA, à haut risque ou non, se voit qualifier de preuve indiciaire pouvant être utilisée par le juge pour la réalisation d'une présomption judiciaire. Toutefois, cette qualification n'est pas liée à la logique de fonctionnement du système d'IA en question, ni à sa qualité ni à l'obtention d'une quelconque certification européenne. Par conséquent, sa catégorisation parmi les preuves indiciaires, et donc son exclusion du système de preuve légale, ne présente aucun lien direct avec le règlement européen sur l'IA ou avec les normes techniques qui y sont associées. En l'absence de dispositions spécifiques à l'admissibilité des preuves produites par un système d'IA, la séparation entre la légalité de l'élément de preuve et sa licéité empêche tout lien de cause à effet entre le respect du nouveau texte européen sur l'IA et la recevabilité de ses résultats. L'articulation entre le recours à un système d'IA comme procédé probatoire et les principes généraux du droit de la preuve démontre l'importance accrue de l'utilisation du système d'IA comme procédé probatoire. En ce qui concerne la conception du système d'IA, les liens affichés sont plus minces et indirects. Ainsi, une opacité trop importante ou un niveau d'exactitude trop faible sont des indicateurs pouvant entraîner une atteinte aux droits de la défense. Dès lors, le fournisseur d'un système d'IA à haut risque pour une finalité probatoire devra prendre en compte ces exigences issues du droit de la preuve, en plus des normes techniques qui pourraient être élaborées à l'avenir. Certains éléments contenus dans la documentation technique deviennent alors pertinents dans l'évaluation de la licéité de la preuve par le juge. Par conséquent, les règles d'admissibilité du droit de la preuve français pourront entraîner des conséquences sur les exigences réglementaires (et/ou les normes techniques) relatives à la conception des systèmes d'IA à haut risque. À l'heure actuelle, le règlement européen sur l'IA, plus particulièrement ses exigences substantielles relatives au fonctionnement des systèmes d'IA à haut risque, n'emporte donc que des conséquences minimales sur l'admissibilité des résultats de systèmes d'IA comme éléments de preuve judiciaire.

CHAPITRE 2 : L'APPRECIATION DES PREUVES PRODUITES PAR UN SYSTEME D'IA

Introduction. Le processus d'appréciation des preuves désigne « *l'ensemble des opérations intellectuelles consistant pour les juges du fond à appréhender les faits litigieux afin d'en constater l'existence et en peser la portée, la gravité, la valeur, les caractères ;* »¹⁵⁶⁶. En d'autres termes, il s'agit de l'étape où le juge du fond s'interroge sur le caractère convaincant de l'élément de preuve au regard de la thèse défendue par la partie qui l'apporte au dossier¹⁵⁶⁷. L'appréciation d'un résultat généré par un système d'IA implique que la nature de l'élément de preuve soit connue du juge. Ainsi, le problème de l'utilisation d'hypertrucages comme preuves judiciaires est exclu du champ d'étude dès lors que le juge n'a pas conscience qu'il se prononce sur un élément de preuve qui n'est pas d'origine humaine. En ignorant l'origine de l'élément de preuve, le juge ne peut, en outre, envisager de s'intéresser à la qualité du système d'IA ayant produit la preuve, écartant immédiatement toute articulation avec le règlement européen sur l'IA.

L'appréciation d'un élément de preuve peut nécessiter pour le juge l'examen de la crédibilité du procédé probatoire. Plus une technique probatoire est fiable, plus les résultats qu'elle génère sont convaincants. Par conséquent, les critères d'appréciation du résultat pourraient nécessiter des informations comme celles élaborées dans le contexte du règlement européen sur l'IA. Au travers d'une approche prospective et à l'aide d'exemples analogues tirés d'autres technologies probatoires, il sera donc étudié comment les preuves techniques utilisées pour certifier un système d'IA à haut risque peuvent être réemployées pour évaluer la crédibilité des résultats générés en tant qu'éléments de preuve.

Annonce de plan. L'appréciation des preuves, en particulier des preuves indiciaires¹⁵⁶⁸, est une obligation qui incombe personnellement aux juges¹⁵⁶⁹. Cependant, si le juge demeure le protagoniste principal de cette analyse (**Section I**), il peut choisir de se faire assister par un expert (**Section II**) ou bien se voir imposer des contraintes d'appréciation préalablement constituées pour la preuve par IA en particulier (**Section III**). Chacun de ces scénarii correspond

¹⁵⁶⁶ CORNU Gérard, *Vocabulaire juridique*, 2022.

¹⁵⁶⁷ FERRAND Frédérique, « Preuve », *Répertoire de procédure civile*, Dalloz, 2024.

¹⁵⁶⁸ Pour rappel, les preuves indiciaires servent de prédicats aux présomptions de faits réalisés par les juges, aussi appelées présomptions judiciaires dans le code civil.

¹⁵⁶⁹ FERRAND Frédérique, *op. cit.*

à de multiples articulations entre les éléments techniques de preuves utilisés pour certifier un système d'IA à haut risque et l'appréciation des résultats produits par ledit système en tant que preuves judiciaires.

Section I : L'appréciation libre par le juge

Annonce de plan. À l'image de l'admissibilité restreinte, il existe deux régimes d'appréciation de la preuve : le régime de libre appréciation de la preuve et le régime de preuves légales. Dans le premier, le juge est libre d'apprécier chaque élément de preuve afin de forger son intime conviction. Dans le second, l'appréciation des éléments de preuve est contrainte par le législateur, liant alors le juge dans sa tâche. Au sein du régime de libre appréciation de la preuve, l'appréciation d'un élément de preuve par un juge est constituée de deux étapes. Dans un premier temps, il évalue sa force probante, à savoir la valeur qu'il accorde à l'élément apprécié¹⁵⁷⁰ (**Sous-section 1**). Dans un second temps, il examine l'adéquation entre l'élément de preuve et l'objet de preuve, on parle alors de la pertinence de la preuve¹⁵⁷¹ (**Sous-section 2**).

Sous-section 1 : L'évaluation de la force probante de la preuve produite par un système d'IA

Annonce de plan. Dans le langage commun, l'adjectif « probant » renvoie au caractère convaincant d'un élément (démonstration, argument, etc.)¹⁵⁷². Utilisé dans le langage juridique, le terme de force probante est défini comme la « *valeur d'un mode de preuve comme élément de conviction ; [la] foi qu'il faut lui attacher [...]* »¹⁵⁷³. Aussi appelée « valeur probante »¹⁵⁷⁴, la force probante est donc une échelle de valeur sur laquelle les preuves sont placées et comparées, par le juge¹⁵⁷⁵. L'examen de la force probante d'une preuve dépend de nombreuses

¹⁵⁷⁰ CORNU Gérard, *Vocabulaire juridique*, 2022.

¹⁵⁷¹ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 437.

¹⁵⁷² « PROBANT : Définition de PROBANT », sur *Cnrtl.fr*.

¹⁵⁷³ CORNU Gérard, *op. cit.*

¹⁵⁷⁴ SAAS Claire, « La valeur probante des procès verbaux », *AJ Pénal*, 2007 Cour de cassation, Chambre criminelle, 6 février 2007, n° 06-83.356 VORMS Marion, « La valeur probante du témoignage », *Cahiers philosophiques*, 142, Réseau Canopé, 2015.

¹⁵⁷⁵ Au sein du régime de libre appréciation de la preuve, à savoir lorsque la force probante de la preuve n'est pas déterminée a priori par le législateur.

caractéristiques qui sont propres à la preuve en question mais aussi au contexte¹⁵⁷⁶. Dans le cadre de la preuve par un système d'IA, il nous semble que cet examen doit prendre en compte, d'une part, la force probante du résultat produit (§1) mais aussi de celle de son mode de preuve, d'autre part (§2).

§1. La force probante de l'élément de preuve

Annnonce de plan. Le résultat du système d'IA est l'élément de preuve que le juge est amené à apprécier. Dans cet exercice, deux critères nous apparaissent pertinents : la préexistence du résultat produit par le système d'IA (**A**) et la crédibilité du résultat, à savoir la valeur accordée par le juge à un élément de preuve (**B**).

A- La préexistence du résultat

La recherche d'un élément existant. À l'heure où les systèmes d'IA génératifs sont au centre de l'actualité médiatique¹⁵⁷⁷ et législative¹⁵⁷⁸, il ne faut pas oublier que les technologies sont aussi capables d'analyser des quantités colossales de données. Ainsi, les systèmes d'IA sont des outils particulièrement utilisés dans la gestion de la modération des plateformes du fait de la quantité de contenus à analyser et de la rapidité avec laquelle les règles de modération doivent être appliquées. Par exemple, les fournisseurs de services de partage de contenu en ligne ont recours à des systèmes d'IA afin de détecter les contenus correspondant à des œuvres protégées

¹⁵⁷⁶ Pour l'exemple de la preuve testimoniale, voir notamment : QUIVIGER Pierre-Yves, « Remarques sur les preuves testimoniales en droit », *Cahiers philosophiques*, 142, Réseau Canopé, 2015 VORMS Marion, *op. cit.*

¹⁵⁷⁷ ChatGPT, Midjourney, Dall-E, etc.

¹⁵⁷⁸ SÉNÉCHAL Juliette, « Les dynamiques actuelles de la future régulation de l'IA, aux niveaux européen et français : entre complexité et angle mort », *Dalloz actualités*, 2023. PETITJEAN Sophie, « ChatGPT, l'invité surprise des négociations sur le règlement pour l'IA - Contexte ».

par le droit d'auteur qui ont été mis à disposition sans le consentement de leur auteur¹⁵⁷⁹. En matière pénale, les systèmes d'IA peuvent également être utilisés pour détecter des cyberattaques au travers de l'analyse des actions réalisées sur les appareils de la victime¹⁵⁸⁰ ou pour détecter des comportements de fraude¹⁵⁸¹.

Dans ce contexte, le système d'IA est utilisé dans une fonction de recherche de la preuve. Le résultat obtenu est une information préexistante à l'utilisation du système d'IA mais difficile à obtenir pour des êtres humains en raison des contraintes de temps ou de ressources (financières et humaines). Cette information aurait parfaitement pu être trouvée par un individu manuellement. Dès lors, l'élément identifié par le système d'IA peut être vérifié par un individu humain *a posteriori*, qu'il s'agisse du déployeur du logiciel ou du juge, sans qu'il lui soit nécessaire de comprendre le fonctionnement du système d'IA. L'individu peut ainsi aisément constater que l'élément trouvé par le système d'IA est exact. En effet, il est possible que l'élément recherché par le système d'IA soit identifié à tort, on parle alors de faux positifs. Dans l'exemple de la détection de piscines non déclarées par le biais d'une analyse par un système d'IA d'images satellites, il s'agit du cas où le système d'IA confond une piscine avec une mare ou avec une bâche bleue étendue sur le sol¹⁵⁸². Si ces faux positifs peuvent causer un préjudice, du point de vue du droit de la preuve, le résultat produit par le système d'IA a la même force probante que s'il avait été recherché par un humain. Le caractère préexistant de l'élément de preuve trouvé par le système d'IA permet une vérification humaine *a posteriori* du résultat. Par conséquent, le fait le système d'IA se limite à chercher un élément existant, vérifiable par un

¹⁵⁷⁹ « Si aucune autorisation n'est accordée, les fournisseurs de services de partage de contenus en ligne sont responsables des actes non autorisés de communication au public, y compris la mise à la disposition du public, d'œuvres protégées par le droit d'auteur et d'autres objets protégés, à moins qu'ils ne démontrent que : a) ils ont fourni leurs meilleurs efforts pour obtenir une autorisation; et b) ils ont fourni leurs meilleurs efforts, conformément aux normes élevées du secteur en matière de diligence professionnelle, pour garantir l'indisponibilité d'œuvres et autres objets protégés spécifiques pour lesquels les titulaires de droits ont fourni aux fournisseurs de services les informations pertinentes et nécessaires; et en tout état de cause c) ils ont agi promptement, dès réception d'une notification suffisamment motivée de la part des titulaires de droits, pour bloquer l'accès aux œuvres et autres objets protégés faisant l'objet de la notification ou pour les retirer de leurs sites internet, et ont fourni leurs meilleurs efforts pour empêcher qu'ils soient téléversés dans le futur, conformément au point b). » - « Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. Article 17.4.

¹⁵⁸⁰ Finalité poursuivie au sein du projet européen H2020 « Starlight » : « Home | Starlight ».

¹⁵⁸¹ « Le fisc généralise la détection des piscines non déclarées par intelligence artificielle », sur *Les Echos*, publié le 29 août 2022. LAUSSON Julien, « Fraude fiscale », sur *Numerama*, publié le 18 février 2021. « LOI n° 2019-1479 du 28 décembre 2019 de finances pour 2020 », 2019-1479, 2019. Article 154 et « Décret n° 2021-148 du 11 février 2021 portant modalités de mise en œuvre par la direction générale des finances publiques et la direction générale des douanes et droits indirects de traitements informatisés et automatisés permettant la collecte et l'exploitation de données rendues publiques sur les sites internet des opérateurs de plateforme en ligne. ».

¹⁵⁸² Pour ce cas d'application, le taux de faux positif était de 30%. EMAIL et PRINT, « Traque à la piscine non déclarée », sur *ZDNet France*, publié le 5 avril 2022.

individu humain, sépare la nature probabiliste du système d'IA de l'évaluation de la force probante de son résultat. En d'autres termes, la préexistence de l'élément de preuve à l'utilisation d'un système d'IA permet d'évaluer sa force probante indépendamment de la qualité de fonctionnement dudit système.

La génération d'un nouvel élément de preuve. Cette logique disparaît lorsque le système d'IA produit un nouvel élément de preuve. Le système d'IA exerce alors une fonction de génération de la preuve. Par exemple, le système d'IA utilisé pour détecter la somnolence du conducteur de véhicule utilise différentes informations issues de capteurs, installés dans le véhicule, pour générer une probabilité que le conducteur soit dans un état de somnolence. De la même manière, l'identification d'un individu à partir de son comportement en ligne ou l'évaluation d'un candidat à un entretien d'embauche sont des résultats générés par un système d'IA.

Au regard de la quantité d'informations prises en entrée par le système d'IA et de son raisonnement opaque, voire incompréhensible pour la logique humaine, il n'est pas envisageable de reproduire le raisonnement du système d'IA. L'absence de préexistence de l'élément produit par le système d'IA empêche tout humain de retrouver manuellement le résultat pour confirmer sa véracité. Dès lors, la force probante du résultat doit être examinée au regard d'autres informations en lien avec l'élément de preuve en question mais aussi avec le système d'IA l'ayant produit. Il peut s'agir de preuves dites « externes » potentiellement préconstituées, à l'image d'une documentation technique par exemple¹⁵⁸³.

B- La crédibilité de l'élément de preuve

La prise en compte risqué du résultat produit par le système d'IA. Au premier abord, la première étape à réaliser pour évaluer la force probante d'un résultat produit par un système d'IA semble être l'analyse du résultat lui-même. En effet, il est indéniable qu'un score élevé produit par un système d'IA apparaît plus convaincant qu'un score faible. Par exemple, un score de 99% de similarité entre deux empreintes ADN ou deux visages semble bien plus crédible pour considérer qu'il s'agit de la même personne que si le score est de 65%. Étant une technique

¹⁵⁸³ GAUTRAIS Vincent, « Preuve des reproductions : vues d'ailleurs ! », *Cahiers Droit, Sciences & Technologies*, PUP, 2014 VIAL Géraldine, VERGÈS Étienne et GAUTRAIS Vincent, « Preuves scientifiques et technologiques », *Cahiers Droit, Sciences & Technologies*, PUP, 2019.

probatoire très récente, les décisions de justice fondées explicitement sur des éléments de preuves produits par un système d'IA sont rares mais pas inexistantes. En 2022, la Cour d'appel de Rouen s'est prononcée sur une ordonnance rendue par le Juge des libertés et de la détention. Grâce à cet arrêt¹⁵⁸⁴, l'on apprend que ledit juge a pris en compte un score produit par un logiciel de reconnaissance faciale pour caractériser l'identité de l'intéressée : « [...] le juge note (de façon 'honteuse') qu'il y a 68 % de chances que ce soit elle mais il y a donc deux chances sur trois de se tromper. [...] »¹⁵⁸⁵. Cette décision révèle qu'un juge peut, en pratique, prendre en compte une preuve produite par un système d'IA pour forger son intime conviction. Au regard des erreurs d'analyse et d'interprétation réalisées par le juge¹⁵⁸⁶, il ressort de sa décision qu'il s'est appuyé sur le seul score produit par le système d'IA de reconnaissance faciale, sans s'aider, du moins explicitement, de preuves externes pour évaluer la crédibilité du résultat.

Cette illustration permet de mettre en avant les limites importantes de l'évaluation de la force probante d'un résultat d'un système d'IA pris isolément. Tout d'abord, le résultat produit par un système d'IA, bien que le plus souvent représenté par une variable quantitative, n'est pas toujours un score. L'analyse du résultat doit être réalisée systématiquement à la lumière du système d'IA utilisé. Par exemple, le résultat produit par un système d'IA évaluant la capacité de remboursement d'un demandeur de prêt doit être interprété de façon différente selon que le logiciel a été conçu pour noter les individus ou les classer dans des catégories¹⁵⁸⁷. Une évaluation correcte d'un résultat généré par un système d'IA pour présumer un fait nécessite donc d'intégrer le procédé probatoire dans la réflexion.

La garantie de fiabilité du résultat. Toutefois, il existe des informations, propres à chaque résultat produit par un système d'IA, pouvant s'avérer pertinentes pour que le juge puisse se faire une idée de la crédibilité de l'élément de preuve qui lui est soumis. Cette garantie de fiabilité peut alors jouer un rôle semblable à celui de la preuve externe (ou double preuve), au sens qui lui est donnée par le professeur Vincent Gautrais¹⁵⁸⁸. La double preuve est un élément qui permet de garantir certaines conditions exigées de la preuve apportée. Dans le cas d'un écrit électronique, ce sont les informations (une documentation par exemple) qui permettent de

¹⁵⁸⁴ Cour d'appel de Rouen, Chambre des Etrangers, 22 octobre 2022, n° 22/03388.

¹⁵⁸⁵ *Ibid.*

¹⁵⁸⁶ Concernant l'erreur d'analyse, un taux de 68% représenterait seulement 1 chance sur 3 d'erreur, dans le cas d'une interprétation correcte du résultat, ce qui n'est pas le cas dans cette affaire. Pour l'explication de l'erreur d'interprétation, voir supra.

¹⁵⁸⁷ Pour autant, le système d'IA générera dans les deux cas un résultat sous forme quantitative. Mais il existe une différence entre les modèles d'IA dits de classification et ceux dits de régression.

¹⁵⁸⁸ VIAL Géraldine, VERGÈS Étienne et GAUTRAIS Vincent, « Preuves scientifiques et technologiques », *Cahiers Droit, Sciences & Technologies*, PUP, 2019. Par. 90.

garantir l'authenticité et l'intégrité de l'écrit. Dans le cas du résultat de l'IA, cette double preuve poursuivrait la fonction de garantie de fiabilité du résultat produit. Contrairement à l'écrit électronique, la fiabilité du résultat n'est pas une condition imposée par la loi mais la logique nous semble suffisamment similaire pour que le parallèle entre les deux preuves soit pertinent.

Celle-ci peut se matérialiser notamment sous la forme d'un intervalle de confiance, un outil couramment utilisé au sein de la communauté scientifique. Il s'agit d'un intervalle de valeurs au sein duquel se situe le résultat. Plus l'intervalle est important, moins le résultat est fiable. Par exemple, pour un score produit par un système d'IA de 0,82¹⁵⁸⁹, un intervalle de confiance allant de 0.80 à 0.84 signifie qu'il existe une marge d'erreur très faible (plus ou moins 0.2). En revanche, si l'intervalle de confiance s'étend de 0.64 à 1, le juge peut s'appuyer sur cette information pour accorder une force probante moindre au résultat produit.

La confiance dans la garantie de fiabilité. Le recours à une double preuve pour attester de certaines caractéristiques d'une preuve nécessite également que le juge puisse avoir confiance dans cette preuve externe. Ainsi, dans le cas de l'écrit électronique, les preuves externes garantissant l'intégrité et l'authenticité de l'écrit (la signature électronique, horodatage, etc.) présentent également des garanties de confiance réglementées au sein du règlement européen eIDAS¹⁵⁹⁰. Le règlement de 2014 établit un cadre juridique relatif aux mécanismes de signature électroniques, de cachets électroniques, d'horodatages électroniques et d'autres éléments destinés à offrir des garanties relatives aux qualités d'un document électronique (intégrité, authenticité, etc.). En d'autres termes, le règlement eIDAS permet de garantir la confiance dans les preuves externes utilisés pour évaluer la force probante de preuves électroniques, l'écrit électronique notamment.

Bien que cela ne soit pas encore réglementé dans le cadre d'un résultat d'IA, la logique appliquée à l'écrit électronique demeure identique. Le juge qui évalue la force probante d'un résultat produit par un système d'IA ne peut se fier à la preuve externe associée (l'intervalle de

¹⁵⁸⁹ Généralement, le score produit par un système d'IA est situé en 0 et 1. Il peut s'agir du pourcentage de similitude entre deux visages pour une reconnaissance faciale, un niveau de somnolence, de solvabilité ou encore un taux de biais présent dans un système d'IA. Pour une description détaillée du dernier exemple, voir PONS Ronan et RISSER Laurent, « Preuve juridique et systèmes d'intelligence artificielle », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général ; Perspectives comparées*, Bruylant Edition, 2023.

¹⁵⁹⁰ « L'intégrité de la copie résultant d'un procédé de reproduction par voie électronique est attestée par une empreinte électronique qui garantit que toute modification ultérieure de la copie à laquelle elle est attachée est détectable. Cette condition est présumée remplie par l'usage d'un horodatage qualifié, d'un cachet électronique qualifié ou d'une signature électronique qualifiée, au sens du règlement (UE) n° 910/2014 [...] » - « Décret n° 2016-1673 du 5 décembre 2016 relatif à la fiabilité des copies et pris pour l'application de l'article 1379 du code civil », 2016-1673, 2016. Article 3.

confiance ici) que s'il existe des garanties de confiance relativement à cette double preuve. Dans le cas de l'intervalle de confiance, cette garantie prend la forme d'un niveau. Généralement, ce niveau est de 95% mais cela peut varier selon les communautés scientifiques¹⁵⁹¹. Un niveau de 95% signifie que le résultat produit a 95% de chances de se situer dans la marge d'erreur. Dans notre exemple, si l'intervalle de confiance à 95% s'étend de 0,81 à 0,83, le résultat statistique est en effet fiable. Cependant si l'intervalle de confiance présente un niveau de 99%, la marge de manœuvre va s'élargir, disons de 0,78 à 0,86. À l'inverse, un intervalle de confiance très large, (0.64, 1) à 95% par exemple, peut être réduit en diminuant son niveau. Il peut alors s'étendre de 0,81 à 0,83 mais avec un niveau de 30%, ce qui signifie que le résultat a 30% de chances de se trouver dans cet intervalle. Un même intervalle de confiance pris avec deux niveaux différents exprimera deux significations, et donc deux forces probantes, très différentes.

Le niveau de l'intervalle de confiance permet donc de mettre en contexte cette garantie de fiabilité pour le lecteur et d'éviter ainsi de l'induire en erreur. Par conséquent, l'intervalle de confiance, et son niveau, apparaissent comme deux éléments indissociables extrêmement pertinents pour aider le juge à apprécier la force probante d'une preuve par un système d'IA.

§2. La force probante du procédé de preuve

Annonce de plan. Bien qu'intimement lié à son résultat, le système d'IA est un élément distinct de l'analyse de la force probante de la preuve produite par un système d'IA. Nous partons ici de l'hypothèse que le résultat a été généré par le système d'IA (par opposition à l'élément préexistant le recours au système d'IA) et que celui est pris comme une information signifiante¹⁵⁹². Par conséquent, la preuve produite est dépendante de la fiabilité du système d'IA en tant que procédé de preuve¹⁵⁹³. Son influence sur la force probante du résultat produit peut être examinée par le biais de différentes preuves externes, soit en se fondant sur des éléments réglementaires externes **(A)**, soit en examinant la fiabilité technique interne dudit système d'IA **(B)**.

¹⁵⁹¹ PONS Ronan et RISSER Laurent, *op. cit.*

¹⁵⁹² Voir le paragraphe 3 de la sous-section 2 de la section du chapitre 1 du présent titre.

¹⁵⁹³ Sur la distinction entre les résultats dépendant de la crédibilité de leur machine et les autres, voir ROTH Andrea L., « Machine Testimony », *The Yale Law Journal*, 126, 2017.

A- La conformité aux normes

La certification du système d'IA à haut risque. La prise en compte de la fiabilité du système d'IA ayant produit l'élément de preuve renvoie immédiatement au règlement européen sur l'IA. En ce qui concerne les systèmes d'IA à haut risque, le texte impose ainsi des obligations de conception visant à garantir que les logiciels possèdent un certain niveau de qualité, et notamment un certain niveau d'exactitude. Symbolisé par le marquage CE, la conformité d'un système d'IA à haut risque est un élément qui peut être pris en compte par le juge afin d'évaluer la force probante du résultat apprécié. La présence du certificat européen obtenu par le fournisseur peut ainsi être réutilisé comme un argument renforçant la capacité du résultat généré par le système d'IA à emporter la conviction du juge. À *contrario*, un système d'IA à haut risque non certifié, en plus d'être illégal, laisse présager que son fonctionnement est d'une qualité inférieure au niveau attendu par le marché européen.

Le recours à des certifications alternatives pour les autres systèmes d'IA. L'influence du mécanisme européen de certification est toutefois limitée aux systèmes d'IA qualifiés à haut risque. Dès lors, un système d'IA qui ne serait pas produit dans le but d'une utilisation dans un des secteurs à haut risque listés par la Commission européenne n'a pas la possibilité de présenter de cette preuve réglementaire externe de fiabilité. Il existe néanmoins d'autres possibilités pour apporter une preuve externe de fiabilité : la certification privée. La logique est identique à celle prévue par le règlement européen. La conception du système d'IA est réalisée conformément à une norme technique privée adéquate (ISO, IEEE, etc.) et la conformité du système est ensuite évaluée (par le fournisseur lui-même ou par un prestataire tiers) au regard des standards techniques choisis. La différence avec le certificat européen est que l'approbation obtenue par un système de certification privée n'offre aucune garantie relative à la conformité à des obligations légales. De plus, les exigences à atteindre, la procédure à appliquer ou encore les acteurs chargés de vérifier la certification peuvent varier entre une norme privée et une norme européenne harmonisée. L'octroi d'un certificat issu de standards privés, non reconnu par l'Union européenne, ne représente donc pas un élément réglementaire de fiabilité.

L'influence des normes techniques dans la force probante de la preuve. L'influence des normes techniques sur la force probante des instruments probatoires scientifiques n'est pas une spécificité des technologies d'IA. En effet, ces standards privés ont déjà une place importante dans les régimes probatoires spécifiques des modes de preuves scientifiques. Même si le raisonnement par analogie est limité du fait de la différence de nature des régimes probatoires

entre les modes de preuves statistiques fréquemment utilisés et les systèmes d'IA¹⁵⁹⁴, il permet d'éclairer la place des normes techniques comme condition de légalité de ces instruments probatoires. Par exemple, l'identification d'un individu par comparaison des empreintes génétiques ne peut être réalisée que par des laboratoires accrédités, une accréditation décernée au regard du respect de la norme ISO/CEI 17205¹⁵⁹⁵. Le renvoi explicite à une norme privée est un choix fort de la part du législateur français qui confère alors au standard technique la force probante de la loi. Contrairement aux normes harmonisées européennes, il ne s'agit pas ici d'une présomption simple de conformité mais bien d'une conformité directe avec les exigences législatives. Le lien entre le respect de la norme harmonisée et la conformité du mode de preuve au droit est ici direct. Par conséquent, la conformité du système d'IA aux normes européennes harmonisées est donc un facteur d'influence crédible dans l'appréciation de la force probante de l'élément de preuve produit.

B- La fiabilité du fonctionnement du procédé

Le lien entre la fiabilité technique du système d'IA et sa crédibilité comme procédé de preuve. L'hypothèse selon laquelle la fiabilité technique¹⁵⁹⁶ du procédé de preuve qu'est le système d'IA serait un facteur d'évaluation de la force probante repose sur un postulat important : il existe un lien entre l'exactitude d'un système d'IA et la confiance que les individus accordent à ses résultats. Ce lien est présent sous deux aspects. Premièrement, il semble évident qu'un système d'IA qui présente un niveau d'exactitude global de 99% produit des résultats plus probants qu'un système d'IA à 60% de précision. Cette influence de l'exactitude d'un système d'IA sur la confiance a été étudiée par Yin, Vaughan et Wallach dans un article de 2019¹⁵⁹⁷. Dans un premier temps, ils confirment que l'exactitude annoncée d'un

¹⁵⁹⁴ À ce jour, les systèmes d'IA comme modes de preuves sont régis par le droit général de la preuve tandis que les autres instruments scientifiques probatoires pris en exemple sont des preuves spéciales qui font l'objet d'un régime juridique spécifique, aussi bien sur leur admissibilité ou sur leur force probante. Voir supra.

¹⁵⁹⁵ « Les laboratoires où sont exécutées les missions d'identification par empreintes génétiques dans le cadre de procédures judiciaires ou de la procédure extrajudiciaire d'identification des personnes décédées sont accréditées par l'organisme national d'accréditation suivant la norme ISO/ CEI 17025. » - « Décret n° 97-109 du 6 février 1997 relatif aux conditions d'agrément des personnes habilitées à procéder à des identifications par empreintes génétiques dans le cadre d'une procédure judiciaire ou de la procédure extrajudiciaire d'identification des personnes décédées. », 1997. Article 9.

¹⁵⁹⁶ La fiabilité technique du système d'IA est sa capacité à produire des décisions exactes dans différentes circonstances. La notion de fiabilité reprend la signification du principe de fiabilité technique produit dans les différents textes éthiques sur l'IA. Pour plus de détails sur ce principe, voir C— Le principe de fiabilité technique.

¹⁵⁹⁷ ZHANG Yunfeng, LIAO Q. Vera et BELLAMY Rachel K. E., Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2020.

système d'IA influence bien la confiance dans le modèle¹⁵⁹⁸. Dans un second temps, ce lien peut être exacerbé ou réduit en fonction de l'exactitude constatée du logiciel dont l'influence est déterminante dans la confiance de l'utilisateur dans le système d'IA¹⁵⁹⁹. Par conséquent, la fiabilité d'un système d'IA est un facteur empiriquement démontré du développement de la confiance dans les résultats chez l'utilisateur. Deuxièmement, l'hypothèse est soutenue par plusieurs dispositions juridiques qui mettent en avant l'importance de la précision dans les procédés de preuve. Par exemple, la loi française impose aux cinémomètres (les radars routiers) des marges d'erreurs maximales lors de leur mise (ou remise) en service et lors de leur utilisation¹⁶⁰⁰. Par ailleurs, une trop faible fiabilité du procédé de preuve est constitutive d'une atteinte aux droits de la défense et entraîne même l'irrecevabilité des résultats¹⁶⁰¹. Dans une logique semblable, de nombreux rapports européens, textes réglementaires et éthiques énoncent que la fiabilité technique est un facteur important de la confiance dans les technologies d'IA¹⁶⁰². Par conséquent, la définition d'un niveau acceptable d'exactitude, par le biais d'un indicateur et d'un seuil probabiliste, d'un système d'IA est une décision cruciale car elle peut impacter la force probante que le juge accorde au résultat produit par ce système. Cette caractérisation de la notion d'exactitude est aujourd'hui confiée aux organismes de normalisation européens¹⁶⁰³. Il est donc indispensable de s'intéresser aux éléments de preuves composant cette exigence d'exactitude afin de minimiser le risque que les juges, mais aussi les autres utilisateurs de systèmes d'IA, accordent une crédibilité aux résultats sur la base de facteurs non maîtrisés.

Les caractéristiques techniques du système d'IA comme facteurs d'évaluation de la force probante. La conformité du système d'IA au droit n'est pas le seul facteur qui pourrait être pris en compte pour évaluer la force probante du résultat produit. Le juge, ou l'expert judiciaire, pourrait choisir d'approfondir son analyse de force probante en examinant les caractéristiques techniques du logiciel. En d'autres termes, le juge ne limite pas son évaluation de la force

¹⁵⁹⁸ « *We find that a model's stated accuracy on held-out data affects people's trust in the model* » - YIN Ming, VAUGHAN Jennifer et WALLACH Hanna, *Understanding the Effect of Accuracy on Trust in Machine Learning Models*, 2019. P. 10.

¹⁵⁹⁹ *Ibid.*

¹⁶⁰⁰ La loi prévoit différentes marges d'erreurs maximales selon, d'une part, que le cinémomètre soit fixe ou embarqué dans un véhicule en mouvement ou d'autre part que la vitesse mesurée soit inférieure ou supérieure à 100km/h. « Arrêté du 4 juin 2009 relatif aux cinémomètres de contrôle routier », 2009. Articles 5 et 6.

¹⁶⁰¹ Sur le lien entre les droits de la défense et la fiabilité du mode de preuve, voir §2. Le respect des droits de la défense.

¹⁶⁰² « *l'exactitude, la fiabilité et la transparence sont particulièrement importantes pour éviter les conséquences négatives, conserver la confiance du public* » - « Règlement sur l'intelligence artificielle », 2024. Considérant 38. COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Commission européenne, 2020. P.24.

¹⁶⁰³ COMMISSION EUROPÉENNE, « Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence », 2022.

probante à la seule présence du marquage CE mais exploite également d'autres informations, plus précises, concernant le système d'IA. Des caractéristiques techniques pertinentes à l'évaluation de la force probante peuvent être obtenues dans la documentation constituée par le concepteur du logiciel. Lorsque le système d'IA ayant produit la preuve appréciée est un système d'IA à haut risque, ces informations sont notamment contenues dans la documentation technique dont la constitution est imposée au fournisseur par le règlement européen sur l'IA¹⁶⁰⁴. Dans cette documentation se trouve entre autres « *les capacités et les limites du système d'IA sur le plan des performances* »¹⁶⁰⁵ ainsi que « *le niveau global d'exactitude prévu par rapport à la destination du système* »¹⁶⁰⁶. Les éléments de preuve sélectionnés lors de la conception du système d'IA seraient alors réutilisés comme facteurs d'évaluation de la force probante du résultat. Par conséquent, les caractéristiques techniques du système d'IA, en particulier sa précision, exercent une influence sur la force probante du mode de preuve.

L'importance des biais dans l'appréciation du résultat. En raison de l'existence de biais discriminatoires, la prise en compte du contenu de la documentation s'avère déterminante pour une appréciation la plus précise possible. En effet, le niveau d'exactitude du système d'IA peut fortement varier du taux moyen indiqué selon le cas traité. L'exemple le plus illustratif est la reconnaissance de visage par algorithme d'apprentissage automatique. Pour un même système d'IA, le taux d'erreur maximal pour un homme blanc est de 20% tandis qu'une femme noire atteint 50%¹⁶⁰⁷. C'est sur la base de cette fiabilité locale qu'Éric Loomis contestait le résultat rendu à son encontre par COMPAS¹⁶⁰⁸. La question des biais algorithmiques est généralement abordée sous le prisme d'un manquement au droit à la non-discrimination mais plus rarement du point de vue de la fiabilité du résultat produit. Du point de vue probatoire, cette problématique se trouve à l'intersection entre la fiabilité de l'élément de preuve et celle du procédé de preuve. En effet, les biais sont présents dans le fonctionnement du système d'IA et s'expriment par le biais du résultat produit. Par conséquent, la force probante d'un résultat d'IA ne peut se limiter à la prise en compte du niveau moyen d'exactitude du procédé de preuve. Il est préférable que le juge prenne en compte les biais mentionnés par le fournisseur dans la documentation technique afin de nuancer la crédibilité du résultat soumis au dossier. La

¹⁶⁰⁴ « Règlement sur l'intelligence artificielle », 2024. Articles 11 et 18.

¹⁶⁰⁵ *Ibid.* Annexe IV. 3.

¹⁶⁰⁶ *Ibid.*

¹⁶⁰⁷ BUOLAMWINI Joy et GEBRU Timnit, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR, 2018.

¹⁶⁰⁸ Cour suprême du Wisconsin, 13 juillet 2016, n° 881 N.W.2d 749.

connaissance et la prise en compte du taux d'erreur moyen du système d'IA pour chaque catégorie protégée d'individus doit être un prérequis pour le magistrat évaluant une preuve issue d'un système d'IA

Sous-section 2 : L'analyse de la pertinence de la preuve par un système d'IA

Annonce du plan. Une fois l'évaluation de la force probante réalisée, l'appréciation du juge se porte sur la pertinence de la preuve apportée, à savoir le rapport entre la preuve offerte et l'objet de la preuve¹⁶⁰⁹. La pertinence est une composante essentielle de la place accordée aux procédés de preuves scientifiques par la loi et la jurisprudence (§1) qui doit ainsi être étudiée dans le cas du résultat produit par un système d'IA (§2).

§1. Le rôle de la pertinence dans les preuves statistiques en droit positif

Annonce de plan. Depuis des dizaines d'années maintenant, des procédés de preuves scientifiques, plus particulièrement statistiques, sont utilisés pour éclairer le juge dans son intime conviction. Deux modes de preuves seront explorés ici : la preuve par expertise biologique en droit de la filiation (A) et la preuve par comparaison d'empreintes digitales en droit pénal (B). Malgré leur nature statistique commune, ces deux modes de preuves occupent des places très différentes dans leurs procédures respectives. Cette divergence qui se justifie selon nous par leur degré différent de pertinence.

A- Le résultat statistique à haute pertinence : le test de paternité en droit de la filiation

L'importance du test de paternité en droit de la filiation. Inscrit au titre VII du code civil, le droit de la filiation est régi par le principe de la liberté de la preuve¹⁶¹⁰. Concernant

¹⁶⁰⁹ CORNU Gérard, *Vocabulaire juridique*, 2022.

¹⁶¹⁰ « Si une action est engagée en application du chapitre III du présent titre, la filiation se prouve et se conteste par tous moyens, sous réserve de la recevabilité de l'action. » - Article 310-3 du code civil.

l'établissement de la paternité et de la maternité, l'expertise biologique domine le débat judiciaire¹⁶¹¹. En mars 2000, la Cour de cassation énonce que « *l'expertise biologique est de droit, sauf s'il existe un motif légitime de ne pas y procéder* »¹⁶¹². Désormais, l'objet de débat entre les parties passe de l'existence d'un lien de filiation vers le refus de la partie à se soumettre au test biologique¹⁶¹³. Au regard du droit à un procès équitable, le refus de se soumettre à cette expertise biologique ne peut pas suffire au juge pour qualifier l'existence d'un lien de paternité¹⁶¹⁴. Pourtant, ce refus est un élément de preuve pris en compte par le juge¹⁶¹⁵ qui l'a déjà qualifié d'aveu tacite de paternité¹⁶¹⁶. En outre, le refus sans motif légitime de procéder à l'expertise biologique peut constituer un préjudice moral pour l'enfant et sa mère qui ouvre un droit à des dommages et intérêts¹⁶¹⁷.

La pertinence parfaite du résultat de test de paternité. L'expertise biologique représente donc un procédé probatoire d'une importance telle que son refus s'avère préjudiciable pour la personne concernée. La preuve par expertise biologique a une valeur probatoire élevée aux yeux des juges. Cette importance peut s'expliquer par la grande fiabilité du procédé probatoire qui s'approche des 99%. Toutefois, c'est réellement la pertinence probatoire de cette preuve qui lui confère cette place centrale dans le litige sur la paternité. En effet, le test de paternité se révèle être un procédé probatoire parfaitement adapté à l'objet de la preuve juridique. Le résultat obtenu permet de garantir que le lien entre les deux personnes est un lien direct de parentalité. L'objet de la preuve est l'existence de lien de filiation directe entre les individus. Le résultat statistique en tant qu'élément de preuve correspond donc directement à l'objet de la preuve. L'existence du premier permet d'induire directement celle du second, sa pertinence probatoire est donc idéale. Corrélée avec sa fiabilité, c'est cette pertinence qui fonde le rôle central de l'expertise biologique en droit de la filiation.

¹⁶¹¹ LE GUIDE Raymond et CHABOT Gérard, « Filiation : modes judiciaires d'établissement », *Répertoire de droit civil*, 2023. Il peut s'agir de l'expertise génétique, dont il est question de cette partie, ou de l'expertise sanguine.

¹⁶¹² Cour de Cassation, Chambre civile 1, 28 mars 2000, no 98-12.806.

¹⁶¹³ GRANET-LAMBRECHTS Frédérique, « Preuves en droit de la famille : La preuve des liens de filiation », *AJ Famille*, 2007.

¹⁶¹⁴ Cour européenne des droits de l'Homme, 2 juin 2015, n° 22037/13.

¹⁶¹⁵ Cour de cassation, 8 juillet 2020, n° 18-20.961.

¹⁶¹⁶ Cour de cassation, 25 septembre 2013, n° 12-24.588.

¹⁶¹⁷ Cour de cassation, 6 mai 2009, n° 08-10.936.

B- Le résultat statistique à faible pertinence : l'empreinte ADN en droit pénal

La place du test ADN en droit pénal. L'utilisation de méthodes d'analyse et d'identification par empreintes génétiques fait son apparition en droit français avec les lois dites de « bioéthiques » de 1994¹⁶¹⁸. La possibilité de recourir aux empreintes génétiques à des fins d'identification « *dans le cadre de mesures d'enquête ou d'instruction diligentées lors d'une procédure judiciaire* »¹⁶¹⁹ fait alors son apparition dans au sein du code civil. De façon similaire au fichier de traitement d'antécédents judiciaires (TAJ), les empreintes génétiques sont stockées dans le fichier national automatisé des empreintes génétiques (FNAEG). Elles servent alors de données de comparaison afin d'identifier le porteur des empreintes ADN retrouvées sur le terrain. À l'exception du marqueur du sexe, la comparaison porte sur des segments d'ADN non codants¹⁶²⁰, à savoir des éléments qui ne permettent pas de retrouver les caractéristiques de la personne (couleurs des yeux, de la peau, etc.). L'identification d'une empreinte génétique à l'aide du FNAEG est possible lors d'une enquête, conformément à l'article 55-1 du code de procédure pénale¹⁶²¹, ou lors d'une instruction au travers d'une demande d'expertise¹⁶²².

La pertinence limitée de l'expertise ADN. En dépit de règles de preuves spécifiques à ce moyen de preuve, l'empreinte ADN occupe un rôle probatoire bien différent de celle du test de paternité. L'ADN est ainsi presque toujours absent des jugements correctionnels et, s'il est plus important en matière criminelle, l'ADN n'y est pas pour autant déterminant¹⁶²³. Au procès, la preuve par ADN semble servir davantage à légitimer scientifiquement la décision du juge aux yeux de la société civile plutôt que d'emporter l'intime conviction du magistrat¹⁶²⁴. L'apport probatoire de l'ADN est en réalité limité à l'élément matériel de l'infraction. Par exemple, dans une affaire de viol, l'ADN permet d'apporter la preuve de l'acte sexuel mais pas l'absence de

¹⁶¹⁸ « LOI no 94-653 du 29 juillet 1994 relative au respect du corps humain (1) », 94-653, 1994 « Loi n° 94-654 du 29 juillet 1994 relative au don et à l'utilisation des éléments et produits du corps humain, à l'assistance médicale à la procréation et au diagnostic prénatal ».

¹⁶¹⁹ Article 16-11 1°

¹⁶²⁰ « Les empreintes génétiques conservées dans ce fichier ne peuvent être réalisées qu'à partir de segments d'acide désoxyribonucléique non codants, à l'exception du segment correspondant au marqueur du sexe. » - Article 706-54 al. 6.

¹⁶²¹ « Il [l'officier de police judiciaire] procède, ou fait procéder sous son contrôle, aux opérations de relevés signalétiques et notamment de prise d'empreintes digitales, palmaires ou de photographies nécessaires à l'alimentation et à la consultation des fichiers de police selon les règles propres à chacun de ces fichiers. » - Article 55-1 al. 2 du code de procédure pénale.

¹⁶²² Article 156 du code de procédure pénale.

¹⁶²³ PY Bruno, « L'utilisation des caractéristiques génétiques dans les procédures judiciaires - Etude de dix années de pratiques en Meurthe-et-Moselle (2003-2013) », Mission de recherche Droit et Justice, 2017.

¹⁶²⁴ LEONHARD Julie, « La place de l'ADN dans le procès pénal », *Cahiers Droit, Sciences & Technologies*, PUP, 2019. P. 54.

consentement de la victime. Le lien entre l'élément de preuve qu'est l'ADN et la caractérisation de l'infraction est donc lacunaire, ce qui permet une contestation de l'hypothèse par la défense. Les avocats ne contestent alors pas la véracité scientifique des faits mais leur qualification juridique¹⁶²⁵. De la même manière, l'ADN retrouvé sur une scène de crime permet uniquement d'attester que l'individu identifié était sur les lieux. La temporalité de sa présence sur le lieu ainsi que sa participation à l'infraction doivent être prouvées par d'autres moyens. Ces exemples illustrent ici la pertinence limitée de la preuve par ADN qui ne peut en aucun cas caractériser la culpabilité d'un individu pour une infraction. La pertinence du résultat statistique apparaît donc comme un élément déterminant lors de l'appréciation de la preuve par le juge.

§2. L'examen de la pertinence du résultat de systèmes d'IA

Annonce de plan. La pertinence est le rapport qui permet d'induire de l'élément de preuve la réalité de l'objet de la preuve¹⁶²⁶. De ce fait, il implique de comprendre précisément ce qui doit être prouvé dans un litige donné **(A)** et la signification exacte de l'élément de preuve **(B)**.

A- L'objet de la preuve statistique

Les deux aspects du résultat produit par un système d'IA. Lorsque l'on souhaite examiner la pertinence d'une preuve, la première étape est d'identifier précisément cet élément de preuve. Dans notre situation, l'élément de preuve est le résultat produit par un système d'IA. Jusqu'à présent, le résultat produit par un système d'IA est un concept qui a été employé dans son sens courant¹⁶²⁷. Néanmoins, il convient désormais de préciser ce concept pour en distinguer deux composantes.

Le résultat intelligible communiqué. Premièrement se trouve le résultat communiqué, c'est-à-dire le résultat formulé sous une forme littérale aisément compréhensible par tout destinataire dudit résultat (utilisateur, juge, expert, etc.). Par exemple, dans le cas d'un détecteur de somnolence, l'on peut imaginer que les résultats communiqués soient présentés de la manière

¹⁶²⁵ BLANC Alain, « La preuve aux assises: entre formalisme et oralité, la formation de l'intime conviction », *AJ Pénal*, 2005.

¹⁶²⁶ CORNU Gérard, *Vocabulaire juridique*, 2022.

¹⁶²⁷ « Ce qui arrive, ce qui se produit à la suite et comme effet de quelque chose » - « RÉSULTAT : Définition de RÉSULTAT », sur *Cnrtl.fr*.

suivante : « conducteur très éveillé », « conducteur éveillé », « conducteur légèrement somnolent », « conducteur somnolent » et « conducteur endormi ». De la même manière, le logiciel de prédiction de récidive COMPAS produit deux résultats : le score de récidive et trois niveaux de risques (« risque faible », « risque moyen » et « risque élevé »)¹⁶²⁸. Dans ce format, le résultat est présenté sous le prisme de l'objectif public que le système d'IA est censé atteindre. Si un système d'IA doit détecter les courriels spams des courriels légitimes, le résultat annoncé sera « spam » ou « non spam ». Cette présentation du résultat est biaisée par l'objectif que le système d'IA est censé accomplir¹⁶²⁹.

Le résultat statistique calculé. En effet, le résultat tel qu'il est communiqué à son destinataire n'est que rarement le résultat généré par le système d'IA. Il est évident qu'un outil statistique comme un système d'IA ne peut pas produire un résultat en langage naturel facilement compréhensible pour l'utilisateur. Le résultat produit par le système d'IA est en réalité une valeur, le résultat mathématique du processus calculatoire qu'est le raisonnement du système d'IA. Cette composante représente alors une version non-interprétée du résultat du système d'IA. Contrairement au résultat communiqué à l'utilisateur, le résultat statistique produit n'est pas présenté sous le prisme de l'objectif public du système d'IA mais à la lumière de sa variable-cible¹⁶³⁰. Celle-ci est la représentation mathématique de l'objectif que le système d'IA doit optimiser. Elle représente la formalisation informatique de l'objectif annoncé par le fournisseur. Par exemple, un système d'IA de recrutement est présenté comme permettant d'identifier « le meilleur candidat » ou « le candidat le plus adapté à votre entreprise » mais en réalité c'est une fonction mathématique que le système d'IA cherche à optimiser pour reconnaître le « meilleur candidat ». La variable-cible est une interprétation informatique de cet objectif.

La variable-cible comme point de départ de l'examen de la pertinence probatoire. Les risques liés au travail d'interprétation à l'origine de la variable-cible a déjà été mis en avant au regard des biais discriminatoires qu'il peut provoquer¹⁶³¹. Au-delà de ces enjeux pour le droit à la non-discrimination, il s'agit ici de mettre en avant la différence entre la variable-cible et l'objectif communiqué publiquement par le fournisseur du système d'IA du point de vue de la

¹⁶²⁸ « Defendants are ranked low, medium Or, high risk in each category » - JULIA ANGWIN Jeff Larson, « Machine Bias », sur *ProPublica*, ProPublica, publié le 23 mai 2016.

¹⁶²⁹ L'adjectif « biaisé » employé ici ne revêt pas de connotation négative particulière. Il exprime simplement le fait que le résultat sous ce format est communiqué à la lumière de l'objectif officiel du système d'IA par son fournisseur.

¹⁶³⁰ La variable-cible est présente dans les modèles prédictifs. Il n'existe pas de variable-cible dans les techniques descriptives d'intelligence artificielle (la classification par exemple).

¹⁶³¹ BAROCAS Solon et SELBST Andrew D., « Big Data's Disparate Impact », *California Law Review*, 104, 2016.

pertinence probatoire. Pour réaliser les présomptions de fait qui constituent son raisonnement probatoire, le juge doit s'appuyer sur les prémisses les plus certaines possibles afin de limiter les risques d'erreurs. Par conséquent, le point de départ de l'analyse de la pertinence d'un résultat d'un système d'IA doit être sa version informatique et non sa version interprétée par le fournisseur. Il est important que le juge, ou l'expert judiciaire, vérifie que le travail d'interprétation réalisée par le fournisseur du système d'IA est réalisé de façon cohérente. À titre d'exemple, il a été vu que le résultat d'une expertise biologique (test de paternité ou comparaison d'empreintes génétiques) doit être interprété par un expert habilité. L'indice à partir duquel le juge fonde en partie son raisonnement probatoire est la version interprétée par l'expert. Dans le cas d'un système d'IA, l'expert habilité ne doit pas être remplacé par le fournisseur du système d'IA qui ne présente alors aucune garantie, ni de compétence, ni même d'indépendance. Le résultat est déjà le fruit d'une présomption statistique dont le raisonnement est inconnu du juge, il convient donc d'éviter d'ajouter à ce résultat d'autres interprétations, humaines ou statistiques, qui échapperaient au juge. Par exemple, lorsqu'un juge considère qu'un score de 68% produit par un logiciel de reconnaissance faciale signifie qu'il y a deux chances sur trois que les deux visages comparés n'appartiennent pas à la même personne¹⁶³², il commet en réalité deux erreurs. Premièrement, la conversion du pourcentage en fraction est incorrecte¹⁶³³. Deuxièmement, le résultat produit par le logiciel renvoie le pourcentage de similarité entre deux visages, ce qui est très différent du nombre de chances que la photo prise corresponde à l'individu. La différence entre une correspondance à 80% et à 90% est bien plus obscure et complexe que de passer de 8 chances sur 10 que la photo représente l'individu à 9 chances sur 10¹⁶³⁴. Par exemple, Amazon indique qu'un « *niveau de confiance de 90% renvoie à une probabilité de correspondance correcte supérieure à 60%* »¹⁶³⁵. Le système d'IA de reconnaissance faciale produit un résultat de similarité entre une photo prise et une base de données de visage, et non un pourcentage de chance que la photo corresponde à l'individu.

¹⁶³² « Le juge note (de façon 'honteuse') qu'il y a 68 % de chances que ce soit elle mais il y a donc deux chances sur trois de se tromper. » - Cour d'appel de Rouen, Chambre des Etrangers, 22 octobre 2022, n° 22/03388.

¹⁶³³ Voir §1. Le rôle de la pertinence dans les preuves statistiques en droit positif.

¹⁶³⁴ « la concordance des traits du visage d'un suspect à celle d'images de personnes dans une base de données, propose un chiffre de concordance qui n'est pas très clair pour les agents chargés de le décrypter. Que signifie une concordance de 91% par rapport à une concordance de 85%... Les différences sont-elles liées aux traits du visage ou à d'autres facteurs comme la pose, le flou ou la luminosité ? Parfois, une correspondance désigne une photo d'un suspect quand plusieurs du même suspect sont disponibles, sans matcher avec les autres images, alors que celles-ci peuvent être plus récentes ou de meilleures qualités. [...] » - GUILLAUD Hubert, « La reconnaissance faciale en ses limites », sur Hubert Guillaud, publié le 21 novembre 2023.

¹⁶³⁵ « Vue d'ensemble de la détection des visages et de la comparaison faciale - Amazon Rekognition ». Les propos portent sur les logiciels de reconnaissance faciale vendus par Amazon. Sans connaître le modèle exact utilisé par la police, il n'est pas possible de comparer strictement avec le score énoncé par le juge dans notre exemple.

L'élément de preuve doit être analysé au regard de sa signification la moins biaisée possible, et donc la plus brute, qu'est le résultat statistique. Bien que coûteux économiquement et en temps, ce choix garantit une meilleure maîtrise du processus d'évaluation de la pertinence de la preuve. Le recours à un expert judiciaire pour produire une interprétation intelligible du format mathématique peu compréhensible de la variable cible est une solution pertinente pour le juge.

B- L'objet de la preuve juridique

La variabilité de l'objet de preuve visé par le résultat du système d'IA. Si le règlement européen sur l'IA énonce différentes contraintes selon le niveau de risque et le secteur d'utilisation des systèmes d'IA, elle vient encadrer de façon transversale le marché des technologies d'IA. Par conséquent, les systèmes d'IA pourront être commercialisés dans la totalité des secteurs d'activité existants¹⁶³⁶. En plus des normes techniques horizontales, c'est-à-dire destinées à tous les systèmes d'IA, des spécifications verticales, propres à certains systèmes d'IA et certains secteurs d'activités de ces standards techniques, devraient être également élaborées¹⁶³⁷. Outre le fait qu'elles soient encore en cours d'élaboration, ces normes techniques, même une fois publiées dans le Journal officiel de l'Union européenne, ne détiendront pas juridiquement une force contraignante mais seulement une force incitative¹⁶³⁸. À l'exception des dispositions interdisant le recours à un traitement automatisé d'évaluation de la personnalité pour fonder une décision de justice¹⁶³⁹, il n'existe pas de dispositions venant interdire sans exceptions l'utilisation de l'IA¹⁶⁴⁰.

La possibilité d'une utilisation multisectorielle éloigne le système d'IA des autres procédés de preuves pris jusqu'à présent en exemple sur deux aspects. Premièrement, le test de paternité ou

¹⁶³⁶ L'article 5 du règlement liste une série d'usages prohibés des systèmes d'IA mais n'interdit aucunement le recours à un système d'IA dans un secteur d'activité en particulier.

¹⁶³⁷ « *Nonobstant leur nature horizontale générale, les normes européennes ou les documents de normalisation européens peuvent fournir des spécifications verticales, lorsque cela est jugé nécessaire à la lumière des risques spécifiques posés par certains systèmes d'IA.* » - COMMISSION EUROPÉENNE, « Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence », 2022. Annexe II.1. Pour rappel une spécification verticale est un livrable ou une partie de livrable destiné à un cas d'utilisation ou un secteur d'application spécifique – *Ibid.*

¹⁶³⁸ Pour rappel, le respect des normes européennes harmonisées entraîne une présomption de conformité aux dispositions réglementaires auxquelles elles sont associées.

¹⁶³⁹ « Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés », 1978. Article 47 al. 1.

¹⁶⁴⁰ L'article 47 alinéa 2 prévoit une interdiction de prise de décision par un système automatisé. Toutefois, les conditions d'application et les exceptions à cette disposition sont si nombreuses que la plupart des situations sont autorisées.

la comparaison d'empreintes génétiques représentent des cas d'utilisations de modes de preuves statistiques. La preuve produite par un système d'IA est un champ qui contient de très nombreux cas d'applications. Deuxièmement, ces modes de preuves sont utilisés ici à des fins probatoires alors que dans le contexte du système d'IA, ce dernier n'est pas nécessairement utilisé comme un outil de preuve. C'est la réutilisation du résultat comme élément de preuve qui transforme le système d'IA qui l'a produit en procédé de preuve. Dès lors, la conception et l'utilisation d'un système d'IA, pensé pour une finalité commerciale par exemple, ne sont pas nécessairement adaptées pour les exigences du droit de la preuve. Dès lors, le résultat d'un système d'IA peut être utilisé par les parties pour prouver potentiellement n'importe quel fait juridique. Par conséquent, la pertinence du résultat produit par un système d'IA ne peut être prédite ou évaluée de façon unique car elle est dépendante de l'objet de la preuve que l'on cherche à attester.

La nature de l'objet de preuve comme possible critère de la pertinence de la preuve par un système d'IA. La distinction entre l'information communiquée du système d'IA et l'objectif statistique qu'est la variable-cible met en lumière la contrainte de simplification statistique intrinsèque au recours aux systèmes d'IA. Les caractéristiques qualitatives d'une situation ou d'un individu sont quantifiées, les valeurs continues sont discrétisées¹⁶⁴¹ et inversement selon les besoins de la performance algorithmique. En d'autres termes, le réel est transformé statistiquement et cette simplification est plus ou moins importante selon l'objet de la preuve. Ainsi entre 2019 et 2022, une expérimentation a été menée sur une centaine de juges et d'auditeurs de justice afin d'observer leur attitude face à un résultat algorithmique¹⁶⁴². Deux résultats différents issus d'une expertise recourant à un système d'IA sont proposés comme preuve aux participants : le calcul d'une prestation compensatoire, d'une part, et la détermination (indirecte¹⁶⁴³) de la culpabilité d'autre part. Il ressort clairement de l'expérience que les juges et les auditeurs de justice considèrent la décision du système d'IA très utile lorsqu'elle est employée pour résoudre un problème chiffré qu'est l'évaluation du montant de la prestation compensatoire. En revanche, en ce qui concerne la culpabilité de l'individu, les

¹⁶⁴¹ « Opération consistant à remplacer des relations portant sur des fonctions continues, dérivables, etc., par un nombre fini de relations algébriques portant sur les valeurs prises par ces fonctions en un nombre fini de points de leur ensemble de définition. » - LAROUSSE Éditions, « Définitions ». Par exemple, la répartition d'un ensemble de valeurs en deux classes différentes est une opération de discrétisation.

¹⁶⁴² VERGES Etienne, « Le juge face à la boîte noire: l'intelligence artificielle au tribunal », *Recueil Dalloz*, 2022.

¹⁶⁴³ Pour être plus précis, l'expertise réalisée par le système d'IA ne vise pas directement la culpabilité de l'individu mais porte sur la possibilité que le bébé ait été victime du syndrome du « bébé secoué » dont la présence caractérise la réalisation de violence causée par le père.

participants sont beaucoup moins convaincus par l'utilité du résultat du système d'IA¹⁶⁴⁴. Certains d'entre eux soulignent que la valeur statistique produite par le système d'IA est « *une facilité apparente ou, encore, qu'elle révèle des données théoriques qui ne sont pas susceptibles d'être appliquées en pratique dans un dossier.* »¹⁶⁴⁵. Bien que ce ne soit pas le seul argument opposé à l'utilité d'un système d'IA dans la détermination de la culpabilité, il apparaît lorsque l'objet de la preuve est de nature qualitative. Dans ce cas, la simplification statistique du réel opérée par le système d'IA (également présente dans la première situation) devient trop importante aux yeux des juges. Que le résultat produit par le système d'IA soit exact ou non, il ne permet pas aux juges d'inférer l'existence de l'objet de la preuve. Par conséquent, il est possible de suggérer de cette expérimentation que la pertinence du résultat d'un système d'IA est plus faible lorsque l'objet de la preuve est qualitatif tandis que la pertinence s'améliore lorsqu'il est quantitatif.

CONCLUSION DE LA SECTION I :

Pris en tant qu'information signifiante, le résultat généré par un système d'IA est un élément de preuve complexe à apprécier pour le juge. Les caractéristiques de fonctionnement du procédé de preuve, à savoir le système d'IA, apparaissent alors comme des informations pertinentes pour éclairer le juge. En raison du principe de libre appréciation de la preuve, les effets de ces informations dans le raisonnement judiciaire sont aujourd'hui indéterminables, celles-ci pouvant être aussi bien décisives qu'indifférentes selon le litige. Néanmoins, il semble exclu que la qualité du fonctionnement du système d'IA puisse être utile lors de l'examen de la pertinence du résultat généré. L'articulation entre les éléments techniques relatifs au fonctionnement des systèmes d'IA et l'appréciation du résultat généré est circonscrite à l'étape d'évaluation de la force probante. Si elle ne peut pas être garantie, la prise en compte de ces éléments est pertinente à l'évaluation cohérente de la force probante du résultat généré par un système d'IA. Dans le cas des systèmes d'IA à haut risque, le juge peut donc utiliser les preuves existantes de fiabilité du procédé de preuve pour forger son intime conviction. Le juge se reposera alors, au travers du marquage CE ou du contenu de la documentation technique, sur les éléments techniques de preuve constitués par le fournisseur lors de la mise en conformité de son système d'IA à haut risque. La question des conséquences, positives ou négatives, de l'articulation entre la réglementation sur les systèmes d'IA et l'appréciation de la preuve est

¹⁶⁴⁴ VERGES Etienne, *op. cit.*

¹⁶⁴⁵ *Ibid.*

contextuelle et sera laissée à de futures recherches. En revanche, il est clair que l'existence de cette articulation lors du recours à une preuve produite par un système d'IA est une possibilité concrète qui accroît encore davantage l'importance des preuves utilisées pour garantir le fonctionnement des systèmes d'IA.

Section II : L'expertise et la preuve par un système d'IA

Annnonce de plan. Face à la complexité technique des systèmes d'IA, l'expert judiciaire apporte une aide importante au juge dans l'appréciation des résultats générés par un système d'IA. À la manière du juge, l'expert peut contribuer à l'articulation entre les éléments techniques de preuves relatives à la conception du système d'IA, imposés par le règlement européen sur l'IA, et l'appréciation des preuves produites par un système d'IA. Le rôle de l'expert dans cette articulation dépend, d'une part, des missions qui lui sont confiées par le juge (**Sous-section 1**) et, d'autre part, de l'influence que son rapport d'expertise exerce dans le raisonnement du juge (**Sous-section 2**).

Sous-section 1 : Les missions de l'expert

Annnonce de plan. Indépendamment de l'acteur à l'initiative du recours à un expert judiciaire (désignation d'office par le juge ou sur demande des parties), le juge doit établir souverainement la ou les missions que ledit expert doit réaliser¹⁶⁴⁶. Outre le délai accordé à l'expert, la détermination de ses missions permet de délimiter le périmètre des opérations qu'il pourra accomplir. Malgré la liberté du juge dans l'élaboration des missions de l'expert, le législateur français limite cette prérogative en restreignant ces missions aux seuls faits du litige, autant en matière civile que pénale¹⁶⁴⁷. Concernant la preuve produite par un système d'IA, l'expert peut être désigné pour deux types de missions : la production d'une preuve par expertise à l'aide

¹⁶⁴⁶ « [...] Et attendu que la seconde branche du moyen, qui critique la mission donnée à l'expert, se heurte au pouvoir souverain reconnu aux juges du fond pour fixer l'étendue de cette mission, qui ne peut être remise en cause devant la cour de cassation ; » - Cour de Cassation, Chambre civile 1, 26 novembre 1980, n° 79-13870, Cass. Civ. 1re, 26 nov. 1980, pourvoi n°79-13870, Bull. civ. des arrêts Cour de Cassation Civ. 1re N. 308, AHJUCAF.

¹⁶⁴⁷ « Il ne doit jamais porter d'appréciation d'ordre juridique. » - Article 238 al. 3 du code de procédure civile. « La mission des experts qui ne peut avoir pour objet que l'examen de questions d'ordre technique est précisée dans la décision qui ordonne l'expertise » - Article 158 du code de procédure pénale.

d'un système d'IA, d'un côté (§1), et l'évaluation d'une preuve par un système d'IA déjà générée, de l'autre (§2).

§1. La production de la preuve par un système d'IA

L'usage traditionnel de l'expert comme utilisateur du mode de preuve statistique.

L'expertise fait partie de la catégorie des mesures d'instruction exécutées par un technicien avec les constatations et la consultation¹⁶⁴⁸. Le recours à l'expert judiciaire peut donc être choisi dans le but d'aider le juge dans sa décision en produisant de nouveaux éléments de preuves. Dans le cadre des éléments statistiques de preuve, ce type de mission est courant. Les deux principaux exemples de preuves statistiques utilisés dans ce chapitre, que sont le test génétique de paternité en droit de la filiation et la comparaison d'empreintes génétiques en droit pénal, en sont de très bonnes illustrations. Dans les deux situations, le juge prononce le recours à un expert pour établir des faits au travers de preuves statistiques. En cours d'instruction¹⁶⁴⁹, la comparaison d'empreintes génétiques est une mesure d'instruction devant être autorisée par le juge. Elle lui permet alors d'obtenir l'identité de l'individu dont les empreintes génétiques ont été retrouvées et ainsi d'éclairer sa réflexion. De la même manière, en matière de filiation, l'expertise biologique est prononcée par le juge afin d'apporter une preuve décisive sur le lien de paternité entre le père présumé et l'enfant¹⁶⁵⁰. Dans ces deux situations, l'expert a recours à un outil statistique en vue de réaliser la ou les missions que lui a confié le juge.

Vers l'apparition d'une expertise algorithmique ? Cette relation entre la preuve statistique et l'expert judiciaire peut être transposée à la preuve générée par un système d'IA. Deux excellentes illustrations peuvent être trouvées dans l'enquête réalisée par les professeurs Vergès et Vial en collaboration avec le syndicat de la magistrature auprès de juges et auditeurs de justice¹⁶⁵¹. Pour rappel, l'enquête visait à observer l'influence d'un résultat par un système d'IA sur le raisonnement des juges. Les deux scénarios étaient la réalisation d'un diagnostic médical

¹⁶⁴⁸ Voir le chapitre V du sous-titre II du titre VII et livre premier du code de procédure civile. L'expertise est encadrée par la section IV dudit chapitre.

¹⁶⁴⁹ La comparaison entre deux empreintes génétiques pouvant également être réalisée au cours de l'enquête.

¹⁶⁵⁰ Dans un litige sur la paternité de l'enfant, l'expertise biologique est de droit, ce qui réduit considérablement le pouvoir d'appréciation du juge dans le prononcé de la mesure d'instruction. Le juge ne peut en effet s'opposer à la demande d'une partie seulement s'il existe un motif légitime. Voir Cour de Cassation, Chambre civile 1, 28 mars 2000, no 98-12.806.

¹⁶⁵¹ VERGES Etienne, « Le juge face à la boîte noire: l'intelligence artificielle au tribunal », *Recueil Dalloz*, 2022 VIAL Géraldine, « Prise en main d'un outil d'intelligence artificielle par des auditeurs de justice : l'office du juge sous l'influence des algorithmes », 2022.

par un système d'IA, dont le résultat était une preuve concluante de la culpabilité de l'accusé, et la prédiction par un système d'IA du montant d'une prestation compensatoire. Au-delà des conclusions de ce travail de recherche, on remarque que ces deux cas pratiques fictifs présentent tous deux une preuve par un système d'IA réalisée par le biais d'une expertise¹⁶⁵². Plus précisément, l'estimation de la prestation compensatoire est fournie par l'épouse dans le cadre d'une expertise de partie¹⁶⁵³ (dont on suppose qu'elle a respecté le principe du contradictoire¹⁶⁵⁴) tandis que la preuve générée par un système d'IA dans le litige criminel est une expertise réalisée sur la demande des experts médicaux¹⁶⁵⁵. Ces deux illustrations mettent ainsi l'accent sur la diversité des usages d'un système d'IA comme procédé de preuve, celui-ci n'étant pas forcément utilisé à la suite du prononcé d'une mesure d'instruction par le juge. Néanmoins, cette étude présente l'expert uniquement sous une fonction de producteur de la preuve. Alors que cette expertise algorithmique apparaît comme une suite logique aux outils statistiques déjà utilisés, cette représentation nous semble manquer une fonction cruciale de l'expert : l'évaluation de la preuve produite par le système d'IA.

§2. L'évaluation de la preuve produite par un système d'IA par l'expert

Annonce de plan. Que ce soit du fait d'une défiance naturelle envers la technologie ou simplement d'une connaissance des limites des systèmes d'IA, les juges seront amenés à douter de la preuve produite par un système d'IA. Pour les éclairer, ils peuvent décider de se faire assister par un expert judiciaire dans le but d'obtenir des informations sur la crédibilité appropriée à accorder à ce résultat. Le recours à un expert judiciaire comme évaluateur peut potentiellement servir deux finalités : l'évaluation de la fiabilité statistique du procédé de preuve et de son résultat **(A)** et l'évaluation de la crédibilité scientifique du résultat signifiant **(B)**.

¹⁶⁵² Ce contexte est d'autant plus intéressant que les deux scénarii ont été choisis par l'école nationale de la magistrature. VERGES Etienne, *op. cit.*

¹⁶⁵³ « Dans ce dossier civil, la preuve algorithmique était une expertise fournie par une société privée. » - *Ibid.*

¹⁶⁵⁴ Cour de cassation, Chambre mixte, 28 septembre 2012, n° 11-18.710.

¹⁶⁵⁵ « À ce dossier a été ajoutée la preuve algorithmique. Les experts médicaux ont demandé la réalisation d'une expertise statistique pour évaluer la probabilité que les symptômes correspondent au SBS [syndrome du bébé secoué] » - VERGES Etienne, *op. cit.*

A- La contestation de la preuve produite par un système d'IA

Le résultat du système d'IA intégré dans la preuve par expertise. La preuve par expertise algorithmique est une possibilité particulièrement probable à l'avenir. À l'image des expertises biologiques, le système d'IA devient un outil au service des opérations réalisées par l'expert dans le cadre de ses missions. Le résultat du système d'IA est donc un élément de preuve contenu dans le rapport d'expertise. La discussion de cette preuve d'expertise, et notamment la formulation de contestations par les parties, est un point essentiel dans le respect du principe du contradictoire. Toutefois, les modalités de discussions varient selon que le système d'IA a été utilisé dans le cadre d'une expertise amiable ou d'une expertise judiciaire. Ensuite, il apparaît que lorsque le résultat d'un système d'IA est présenté comme une preuve en-dehors du cadre d'une expertise, la capacité des parties à le contester apparaît plus importante que lorsque le résultat est intégré dans une expertise.

La contestation limitée du résultat du système d'IA utilisé dans une expertise amiable.

Contrairement à l'expert judiciaire, l'expert amiable (aussi appelé expert de partie) est un professionnel désigné non pas par le juge mais par une partie, le commanditaire, afin de l'éclairer sur un point technique du litige. Le principe du contradictoire ne s'applique que partiellement à l'expertise de partie. Seul le rapport de l'expert de partie doit être soumis aux débats. La réalisation des opérations de l'expert, quant à elle, ne nécessite pas d'informer et/ou de convoquer la partie adverse¹⁶⁵⁶. Par conséquent, la partie adverse ne peut discuter et contester la valeur de l'expertise qu'*a posteriori* et à la lumière des informations contenues dans le rapport d'expertise amiable. Dans cette situation, le système d'IA n'est qu'un élément utilisé par l'expert pour produire son rapport technique. De ce fait, la limitation de la partie adverse aux seules informations contenues dans le rapport limite inévitablement sa capacité à discuter de certains éléments techniques relatifs à la crédibilité du système d'IA utilisé, comme sa fiabilité technique ou sa pertinence, qui est pourtant essentielle à l'appréciation par le juge de cette preuve par expertise.

En outre, la désignation d'un expert judiciaire ne permet pas non plus d'améliorer la capacité de contestation de l'expertise de partie. D'une part, la partie ayant commandité une expertise amiable n'est pas contrainte de fournir les informations contenues dans son rapport à l'expert judiciaire. Un tel refus n'est aucunement sanctionné par le juge. D'autre part, lorsque le juge

¹⁶⁵⁶ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 675.

élabore les missions de l'expert judiciaire, il ne peut pas lui demander d'émettre des observations sur l'expertise de partie. De ce fait, l'expert judiciaire ne peut servir à la partie adverse pour contester la crédibilité des opérations ou du rapport réalisé par l'expert de partie. Il demeure la possibilité que le juge puisse désigner un expert judiciaire (de sa propre initiative ou de celle de la partie adverse) dans le but d'obtenir un autre avis technique sur une question similaire à celle à laquelle a répondu l'expert de partie. Une telle situation est envisageable dès lors que le juge ne considère pas la mesure, comme toute mesure d'instruction, comme superfétatoire¹⁶⁵⁷. Dans cette éventualité, il est possible que la non-utilisation d'un système d'IA par l'expert judiciaire, ou bien son recours à un système d'IA différent de celui de l'expert de partie, résulte dans la production d'un rapport d'expertise contenant des résultats différents sur un sujet technique similaire. Cela mettrait alors en lumière la place significative jouée par le système d'IA dans l'avis technique de l'expert.

La contestation du résultat du système d'IA utilisé au sein d'une expertise judiciaire. Le second scénario envisagé ici est celui de l'expert judiciaire qui aurait recours à un système d'IA pour mener à bien les missions que lui a confiées le juge. Contrairement à l'expertise amiable, l'expertise judiciaire est soumise au principe du contradictoire aussi bien lors de la réalisation de l'expertise que lors de la communication de l'avis technique aux parties¹⁶⁵⁸. Dès lors, les parties bénéficient d'une plus grande marge de manœuvre pour discuter et contester les aspects liés au système d'IA en tant que procédé de preuve, soit au travers d'observations soit par la demande d'une nouvelle expertise.

Dans un premier temps, les parties peuvent produire des observations lors de la réalisation des opérations d'expertise auxquelles elles doivent être associées. Devant la technicité des opérations, un technicien peut assister les parties lors de l'élaboration de remarques et/ou de contestations sur les actions de l'expert judiciaire¹⁶⁵⁹. S'il n'est pas obligé de les suivre, l'expert judiciaire est toutefois obligé de les prendre en considération en les mentionnant, sur demande des parties, dans son rapport final et en y indiquant les suites qu'il leur a donné¹⁶⁶⁰. La présence des parties à ces opérations est donc essentielle pour s'assurer de la fiabilité et de la pertinence des outils d'IA utilisés par l'expert.

¹⁶⁵⁷ Pour plus de détails sur la notion d'utilité probatoire, voir §1. La théorie de l'utilité probatoire.

¹⁶⁵⁸ Le non-respect du principe de contradictoire lors de la réalisation des opérations par l'expert judiciaire n'est cependant pas sanctionné de nullité du rapport d'expertise. Voir précédemment.

¹⁶⁵⁹ Articles 161 et 162 du code de procédure civile.

¹⁶⁶⁰ Article 276 du code de procédure civile.

Dans un second temps, la contestation de l'expertise par un système d'IA peut *intervenir a posteriori* au regard du rapport produit par l'expert judiciaire. Au-delà des discussions autour de la preuve d'expertise, les parties peuvent demander l'adoption de mesures complémentaires lorsque le principe du contradictoire n'a pas à être respecté pendant la réalisation de l'expertise. Ainsi, en droit pénal, les parties peuvent solliciter une contre-expertise ou un complément d'expertise¹⁶⁶¹. À l'image de la réalisation d'une expertise amiable et d'une expertise judiciaire, la réalisation d'une contre-expertise peut permettre de mettre en lumière, d'une part, les potentielles lacunes de la première expertise mais également l'importance des décisions prises par l'expert. Le recours à un système d'IA différent, ou le non recours à un système d'IA, lors de la contre-expertise peut aboutir à un résultat différent, pouvant alors remettre en question la force probante à accorder à la première expertise.

La contestation du résultat du système d'IA comme preuve par l'expert judiciaire. La dernière situation identifiée ici est le cas où le résultat du système d'IA n'est pas intégré dans un processus d'expertise mais est apportée au dossier directement en tant que preuve. Par exemple, le processus de certification des systèmes d'IA à haut risque, instauré par le règlement européen sur l'IA, entraînera le recours à des audits desdits logiciels avant et pendant leur commercialisation (lors du renouvellement du certificat). Lorsque le système d'IA est opaque par conception, l'évaluation de certaines exigences, notamment celles liées aux propriétés techniques du système d'IA comme la mesure de biais et de l'exactitude, sera réalisée au moyen d'outils algorithmiques pouvant être qualifiés eux-mêmes de systèmes d'IA. Plus généralement, le recours à l'audit algorithmique est une solution sérieusement envisagée par les autorités de contrôle pour évaluer la conformité de ces systèmes d'IA aux exigences énoncées notamment par le règlement sur les services numériques¹⁶⁶² ou le RGPD¹⁶⁶³ au niveau européen et la loi pour une République numérique¹⁶⁶⁴ en droit français. En cas de contestation de la sanction, le résultat de l'audit par un système d'IA sera une preuve que le juge sera amené à apprécier, possiblement à la lumière d'un avis technique produit par un expert judiciaire.

¹⁶⁶¹ Article 167 du code de procédure pénale.

¹⁶⁶² « Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. Le règlement impose plusieurs obligations de transparence relatives aux systèmes algorithmiques utilisés par les fournisseurs de services intermédiaires.

¹⁶⁶³ « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016.

¹⁶⁶⁴ « LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique », 2016-1321, 2016.

De la même manière, les systèmes d'IA sont utilisés en entreprise pour évaluer les candidats au recrutement mais également les salariés de l'entreprise. En droit français, le caractère sérieux d'un licenciement peut être justifié par une insuffisance de résultats¹⁶⁶⁵. La seule insuffisance des résultats n'étant plus une cause de licenciement en soi¹⁶⁶⁶, le juge doit examiner deux conditions : « *d'une part, si les objectifs étaient réalistes, d'autre part, si le salarié était en faute de ne pas les avoir atteints ;* »¹⁶⁶⁷. Si ces objectifs ont été établis de façon personnalisée par un système d'IA, le résultat du système d'IA devient une preuve soumise à l'appréciation du juge pour rechercher le caractère réaliste de ces derniers¹⁶⁶⁸.

Dans ces mises en situations, nous proposons de diviser les missions que le juge peut confier à l'expert judiciaire en deux catégories : d'une part, les missions portant sur la fiabilité du système d'IA ayant produit la preuve et, d'autre part, les missions portant sur l'interprétation du résultat généré. Ces deux types de missions sont directement liés aux deux critères d'appréciation de la preuve par un système d'IA élaborés précédemment. Ils impliquent généralement le recours à des compétences différentes et donc à des experts différents. Pour autant, il n'est pas exclu que le juge désigne deux experts pour l'éclairer à la fois sur la fiabilité technique du procédé de preuve et la pertinence du résultat. Dans un premier temps, la mission de l'expert peut porter sur la fiabilité technique du système d'IA ayant produit la preuve. Le fait à établir est alors la probabilité que la preuve produite soit exacte, au regard de la qualité de fonctionnement du système d'IA et des conditions d'utilisations de ce dernier lors de la production de la preuve. Dans un second temps, l'expert désigné peut vérifier la pertinence du système d'IA. Cette question nécessite davantage de compétence métier car il faut, en plus de solides connaissances en informatique, une bonne compréhension de ce que le système d'IA est censé faire, à savoir ici l'établissement d'objectifs à atteindre pour un employé. Dans le cas du système d'IA servant à l'évaluation des employés, il s'agirait par exemple d'examiner si le système d'IA ne procède pas à une simplification statistique excessive de la réalité complexe du salarié. Le rapport pourrait alors mettre en évidence la différence entre l'objectif communiqué par l'entreprise et l'objectif statistique réellement calculé par le système d'IA. La qualification juridique ou non de ces divergences entre ces deux éléments comme « objectifs réalistes » revient enfin au juge.

¹⁶⁶⁵ Cour de Cassation, Chambre sociale, 13 janvier 2004, n° 01-45.931; 01-45.932.

¹⁶⁶⁶ Cour de Cassation, Chambre sociale, 11 juillet 2001, n° 99-42.927.

¹⁶⁶⁷ Cour de Cassation, Chambre sociale, 13 janvier 2004, n° 01-45.931; 01-45.932.

¹⁶⁶⁸ FABRE Alexandre, « Contrat de travail à durée indéterminée », *Répertoire du droit du travail*, 2023.

B- Les rôles de l'expert judiciaire dans l'utilisation des preuves de conception devant le juge

L'expert judiciaire favorisant l'utilisation des preuves produites à l'aide de systèmes d'IA devant le juge. Lorsque l'expert judiciaire a recours à un système d'IA pour réaliser sa mission, il s'appuie sur les résultats algorithmiques obtenus pour forger son rapport d'expertise. Son statut d'expert présente une certaine garantie de compétences¹⁶⁶⁹ et de confiance¹⁶⁷⁰ dans les conclusions qu'il soumet au juge. En utilisant un système d'IA dans sa méthodologie d'analyse, l'expert judiciaire devient ainsi une porte d'entrée du système d'IA comme procédé de preuve au sein d'un tribunal, et permet l'utilisation de preuves produites par un système d'IA dans ce contexte¹⁶⁷¹. En d'autres termes, l'expert judiciaire favorise l'utilisation des preuves par un système d'IA devant les juridictions sous la forme du rapport d'expertise. Par conséquent, l'expert imprègne d'une partie de sa légitimité scientifique à la fois le fonctionnement du système d'IA utilisé et le résultat généré.

L'expertise judiciaire comme espace de questionnement de la qualité du système d'IA comme mode de preuve. L'expertise judiciaire est aussi un espace de discussions et de contestation d'éléments particulièrement techniques. Le recours à l'expert peut alors devenir un moyen d'interroger les garanties et les limites du système d'IA comme procédé de preuve. La preuve par un système d'IA étant le résultat d'une présomption statistique, il est essentiel que soit discuté le degré de certitude des faits qui ont fondé cette présomption et de la justesse des conséquences qui en ont été tirées¹⁶⁷². L'expert désigné doit alors appliquer une méthodologie d'évaluation de la fiabilité du système d'IA ayant produit l'élément de preuve. Les critères d'évaluation peuvent alors s'aligner sur ceux contenus dans les normes techniques

¹⁶⁶⁹ La garantie dans la compétence de l'expert dans sa spécialité est qualifiée ici de « relative » au sens où l'appréciation de ces capacités professionnelles et le choix de l'inscription de l'expert sur les listes sont réalisés librement par l'assemblée générale des magistrats de la cour d'appel : Cour de cassation, Chambre civile 2, 21 septembre 2006, n° 05-21.978.

¹⁶⁷⁰ Comme nous le préciserons ultérieurement, la confiance du juge dans le rapport d'expertise est variable en fonction de la situation et du juge. Même si l'expert peut orienter le juge dans son travail d'appréciation juridique, « *le juge n'est pas lié par les constatations du technicien.* » - Article 246 du code de procédure civile.

¹⁶⁷¹ Le degré d'importance des résultats du système d'IA dans les conclusions de l'expertise n'est pas fixe et peut se révéler plus ou moins grand. Le terme de preuve par système d'IA désigne à la fois la preuve pour laquelle le résultat du système d'IA a été déterminante que celle où le résultat du système d'IA a eu un impact faible sur le rapport d'expertise.

¹⁶⁷² « *la certitude ou l'incertitude des présomptions [...] dépend de la certitude ou l'incertitude des faits dont on tire les présomptions, et de la justesse des conséquences qu'on tire de ces faits pour la preuve de ceux dont il s'agit.* » - REMY Joseph, *Oeuvres complètes de J. Domat*, Firmin-Didot père et fils (Paris), 1828. P. 49.

élaborés par les organismes européens de normalisation, ou l'expert peut décider de s'en éloigner.

Dans un premier temps, l'expert peut décider de réutiliser les mécanismes d'évaluation (et donc les choix probatoires) sélectionnés par les organismes de normalisation européens au sein des normes techniques. La réutilisation des preuves techniques de certification peut être directe, l'expert procédant alors à une vérification du fonctionnement du système d'IA au regard des exigences prévues par le règlement européen sur l'IA. Elle peut également être indirecte, par la simple vérification de la présence du marquage CE apposé sur le système d'IA utilisé. Dans les deux cas, l'expert est alors un vecteur de réutilisation de ces preuves techniques considérées comme pertinentes, dans le cadre du règlement sur l'IA, dans le contexte judiciaire. À son niveau, l'expert valide les interprétations des exigences de conception du texte européen réalisées par les acteurs¹⁶⁷³. Le niveau de biais, d'exactitude, de transparence qualifié d'acceptable pour l'utilisation du système d'IA ayant produit la preuve est ainsi validé par l'expert. En outre, il renforce, du fait de son statut d'expert, l'influence des preuves techniques de certification considérées comme pertinentes sur la force probante des preuves par un système d'IA.

Le recours par l'expert des preuves techniques énoncées dans les standards européens ne doit pas être nécessairement considéré comme un phénomène négatif. Si le périmètre de la mission et les prérogatives d'investigation qui lui sont confiés par le juge le lui permettent, l'expert éclaire le juge sur la qualité de la formalisation informatique des exigences juridiques réalisée au sein du fonctionnement du système d'IA. Cependant, une utilisation trop peu critique des standards européens par l'expert peut conduire le juge à accorder sa confiance à un système d'IA respectant une interprétation des exigences juridiques différente de celle de la jurisprudence. Par conséquent, l'expert est un acteur important dans l'interprétation des éléments techniques de preuve pertinents dans la réglementation sur l'IA. Selon le degré de validation du résultat qu'il transmet dans son rapport d'expertise, il met en lien les critères de qualité de fonctionnement du mode de preuve avec la force probante du résultat. En outre, les experts judiciaires peuvent développer une méthodologie d'évaluation du système d'IA probatoire qui précisera ou corrigera les potentielles limites et zones de flou contenues dans les standards techniques.

¹⁶⁷³ Pour une étude des acteurs impliqués dans la détermination des preuves techniques pertinentes au sein du règlement européen sur l'IA, voir Titre II : La place de la preuve dans les exigences de conception d'un système d'IA.

Sous-section 2 : La place de l'expert dans l'appréciation du juge

Annonce de plan. Le rôle de l'expert dans l'articulation entre les éléments techniques de preuve de conception, exigés par le règlement sur l'IA, et l'examen de la fiabilité du système d'IA comme procédé de preuve ne permet cependant pas de conclure des conséquences sur l'appréciation de la force probante du résultat du système d'IA par le juge. Le lien entre le fonctionnement du système d'IA et la preuve qu'il produit doit être mis en perspective avec la place accordée aux conclusions de l'expert par le juge. L'étude de l'influence de l'expertise réalisée sur le fonctionnement du système d'IA dans l'appréciation d'un de ses résultats comme preuve judiciaire nécessite au préalable que le juge accepte de désigner un expert **(A)**. Ensuite, l'influence du rapport d'expertise dépend de sa capacité à affecter la décision du juge **(B)**.

§1. Le choix du juge de recourir à un expert judiciaire

Le recours à l'expert judiciaire. L'expertise judiciaire appartient à la catégorie des mesures d'instructions. Sauf exceptions¹⁶⁷⁴, le juge accepte une demande d'expertise ou initie de lui-même une telle action seulement s'il la considère comme utile¹⁶⁷⁵. Pour rappel¹⁶⁷⁶, la théorie de « l'utilité probatoire » désigne la qualité de faits capables d'influencer la résolution du procès¹⁶⁷⁷. En droit positif, « l'utilité probatoire » est la condition pour le prononcé de la mesure par le juge : « *Les faits dont dépend la solution du litige peuvent, à la demande des parties ou d'office, être l'objet de toute mesure d'instruction légalement admissible.* »¹⁶⁷⁸. Par conséquent, le juge ne désignera un expert judiciaire que s'il considère que son expertise peut établir des faits au minimum pertinents pour résoudre le différend.

¹⁶⁷⁴ En droit de la filiation, l'expertise biologique est de droit (Cour de Cassation, Chambre civile 1, 28 mars 2000, no 98-12.806). Par conséquent, le juge ne jouit que d'une liberté réduite quant à l'appréciation de l'utilité probatoire de cette mesure. Pour plus d'informations sur les conditions de prononcé de l'expertise biologique en droit de la filiation, voir A. Le résultat statistique à haute pertinence : le test de paternité en droit de la filiation. De la même manière, le juge voit sa liberté de prononcer une expertise réduite lors d'une contre-expertise en droit pénal si la situation d'espèce est d'une particulière gravité : voir not. LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. PP. 730-731.

¹⁶⁷⁵ Articles 10 et 143 du code de procédure civile.

¹⁶⁷⁶ La théorie de « l'utilité probatoire » a été définie et expliquée en détails précédemment, au sein de la sous-section 2 de la section 2 du chapitre 2 du titre I de la partie 2 de la présente thèse.

¹⁶⁷⁷ Bien qu'absentes du droit positif, il existe des nuances d'utilité probatoire des faits. Selon Aurélie Bergeaud, un fait qui exerce une influence sur l'issue du litige est dit « pertinent » tandis que si son influence est décisive pour la solution au litige il sera qualifié de « concluant ». BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010. P. 234.

¹⁶⁷⁸ Article 143 du code de procédure civile.

À l'image de tout procédé probatoire, l'utilité du recours à un expert judiciaire est analysée au regard d'éléments intrinsèques et extrinsèques¹⁶⁷⁹. Premièrement, il faut que la mesure d'instruction puisse permettre d'obtenir les éléments de preuves pertinents pour le litige. Si l'expert n'est pas en capacité matérielle de produire les preuves attendues, le juge refusera la demande d'expertise de la partie¹⁶⁸⁰. Deuxièmement, l'utilité de l'expertise est évaluée au regard des preuves déjà portées au dossier. Ainsi, le juge dont la conviction est formée par les éléments de preuve déjà présents et qui considère que son avis ne peut être modifié jugera inutile le prononcé d'une mesure d'expertise judiciaire. Dans cette situation, l'expertise est qualifiée de superfétatoire. Par conséquent, des éléments de preuve externes à la mesure d'instruction peuvent exercer une influence sur son utilité probatoire.

L'interaction entre le marquage CE et le recours à un expert judiciaire. La présence du marquage CE sur le système d'IA à haut risque ayant généré la preuve appréciée pourrait constituer un de ces éléments externes. En d'autres termes, l'existence d'une certification européenne d'un système d'IA à haut risque peut-elle réduire l'intérêt probatoire de désigner un expert pour éclairer l'évaluation de la force probante des éléments de preuve produits par ledit logiciel ? Si oui, dans quelle mesure cette influence s'exerce-t-elle ? La question est délicate et la réponse est prospective. En effet, elle interroge frontalement les arguments du juge dans sa décision de diligenter un expert judiciaire pour un procédé de preuve qui n'est que peu utilisé aujourd'hui. Le prononcé d'une expertise judiciaire est une décision d'avant-dire droit et, par conséquent, cette décision d'acceptation ou de refus de recourir à un expert échappe à l'obligation législative de motivation¹⁶⁸¹. Seule la loi pénale impose aux juges de motiver leur refus de prononcer une mesure d'expertise¹⁶⁸². Néanmoins, la jurisprudence, aussi bien civile qu'administrative, invite les juges à motiver le choix ou le refus de l'expertise. Ainsi, la Cour de cassation met en avant l'intérêt pour la cour d'appel d'avoir motivé son choix de refuser la demande de complément d'expertise demandée par une des parties¹⁶⁸³. Le Conseil d'Etat reprend également dans ses arrêts les motivations de la cour administrative d'appel concernant

¹⁶⁷⁹ À nouveau, ces éléments sont détaillés plus précisément au sein du chapitre 2 du titre 1 de la partie 2 de la présente thèse. Pour une analyse approfondie de ces éléments, voir également BERGEAUD Aurélie, *op. cit.*

¹⁶⁸⁰ Cour de cassation, Chambre civile 2, 20 mars 2014, n° 13-14.985, p. 985.

¹⁶⁸¹ « *Le jugement doit exposer succinctement les prétentions respectives des parties et leurs moyens. Cet exposé peut revêtir la forme d'un visa des conclusions des parties avec l'indication de leur date. Le jugement doit être motivé.* » - Article 455 alinéa 1 du code de procédure civile. LECLERC Olivier, *Le juge et l'expert*, LGDJ, 2005. P. 280.

¹⁶⁸² « *Lorsque le juge d'instruction estime ne pas devoir faire droit à une demande d'expertise, il doit rendre une ordonnance motivée au plus tard dans un délai d'un mois à compter de la réception de la demande.* » - Article 156 du code de procédure pénale.

¹⁶⁸³ Cour de cassation, Chambre civile 2, 26 juin 2008, n° 07-13.875.

le refus de prononcer la mesure d'expertise demandée par une partie¹⁶⁸⁴, allant même jusqu'à sanctionner le recours inutile à une expertise judiciaire¹⁶⁸⁵.

Une autre difficulté de cette question est l'absence de jurisprudence spécifique aux systèmes d'IA comme modes de preuve. Tout juste votées, les obligations du règlement européen sur l'IA à haut risque ne sont pas applicables avant trois ans¹⁶⁸⁶ et les normes européennes harmonisées ne seront pas publiées avant 2025¹⁶⁸⁷. De plus, les possibilités de raisonnement par analogie se révèlent peu convaincantes. Premièrement, les modes de preuves statistiques auxquels a été comparé le système d'IA, à savoir l'expertise biologique en droit de la filiation et en droit pénal, font tous deux l'objet d'un régime de preuve spéciale. Dès lors, leurs utilisations sont encadrées plus strictement à la fois par la loi et par une jurisprudence conséquente. Deuxièmement, les modes de preuves que sont les preuves par enregistrement¹⁶⁸⁸ (sonore, image, localisation), s'ils présentent une diversité plus proche de celle des systèmes d'IA, ne possèdent pas sa nature probabiliste. Ainsi, le risque de preuves inexactes du fait de la conception du procédé de preuve est plus faible. Le recours à un expert judiciaire pour vérifier le bon fonctionnement d'un micro ou d'un appareil photo est bien moins pertinent que pour le fonctionnement d'un système d'IA. Au regard de ces limites, il ne sera pas tenté d'induire une logique entre l'existence d'une conception et d'un fonctionnement conforme d'un produit et l'utilité de recourir à un expert judiciaire pour évaluer la preuve qu'il a fournie. Par conséquent, nous explorerons d'abord l'hypothèse où un juge se limite à la vérification de la conformité du système d'IA ayant servi de procédé de preuve et ensuite celle où le juge diligente un expert pour l'éclairer sur la force probante du système d'IA.

L'expert superfétatoire face aux preuves de conception du fonctionnement du système d'IA. Dans un premier temps, le juge peut fonder son appréciation de la preuve par un système d'IA sur les garanties de conception et de fonctionnement du logiciel. En d'autres termes, le juge se considère comme suffisamment informé sur le système d'IA et son résultat en raison de l'existence ou de l'absence de garanties formelles. Par conséquent, le recours à un expert

¹⁶⁸⁴ « *La demande d'expertise est dépourvue d'utilité dès lors que l'ensemble des données météorologiques, seules utiles à la solution du litige, figurent au dossier* » - Conseil d'État, 12 décembre 2007, n° 298155.

¹⁶⁸⁵ « *En estimant qu'une expertise était utile à la solution du litige, le tribunal administratif a dénaturé les pièces du dossier* » - Conseil d'État, 23 octobre 2013, n° 360961.

¹⁶⁸⁶ « *L'article 6, paragraphe 1, et les obligations correspondantes du présent règlement s'appliquent à partir du ... [36 mois après l'entrée en vigueur du présent règlement].* » - « Règlement sur l'intelligence artificielle », 2024. Article 113.c).

¹⁶⁸⁷ COMMISSION EUROPÉENNE, « Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence », 2022.

¹⁶⁸⁸ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

judiciaire sur la pertinence ou la fiabilité de fonctionnement du système d'IA est jugé inutile¹⁶⁸⁹. Dès lors, l'expert n'a plus l'opportunité d'exercer son rôle d'évaluateur du fonctionnement du système d'IA comme procédé de preuve. De même, il ne peut plus offrir un espace de discussion et d'accès à des informations particulièrement techniques pour les parties quant au fonctionnement du système d'IA.

L'expert judiciaire associé aux technologies n'intervient alors que dans son rôle d'utilisateur d'un système d'IA. Ce dernier utilise le système d'IA comme un outil au service d'une expertise médicale, bancaire, environnementale ou autre. En tant que déployeur, l'expert s'appuie sur les résultats produits par son système d'IA pour rédiger son rapport d'expertise. Par conséquent, les choix de conception (acceptabilité du taux d'erreur, de l'atteinte aux droits fondamentaux, interprétation des exigences juridiques par le fournisseur du système d'IA, etc.¹⁶⁹⁰) ne sont pas remis en question et sont transmis au juge au travers du rapport d'expertise.

Cette hypothèse met en avant deux éléments essentiels. Premièrement, elle illustre les effets de l'encadrement *ex ante* d'un produit sur l'appréciation de ses résultats lorsqu'il est utilisé comme procédé de preuve. Qu'elles soient d'origine réglementaire ou non, la satisfaction ou non des normes de conception par le système d'IA est alors un élément suffisant pour le juge quant à l'appréciation du résultat. D'une part, cela signifie que le résultat produit dans un cas précis est apprécié au regard du fonctionnement global du système. Or, la problématique des biais algorithmiques illustre parfaitement les limites de ce raisonnement. D'autre part, en utilisant le système d'IA pour réaliser son expertise, l'expert judiciaire vient intégrer la preuve par un système d'IA au sein de la preuve par expertise. Or, l'expert déployeur d'un système d'IA n'est en rien un spécialiste des technologies d'IA et n'offre donc aucune garantie quant au bon fonctionnement du système d'IA utilisé. La confiance du juge dans l'expert qu'il a désigné est étendue de fait à la méthodologie qu'il utilise. Il est alors essentiel pour les parties de discuter le rapport d'expertise non seulement au regard des opérations réalisées par l'expert mais également des outils qu'il a utilisés.

Deuxièmement, en appréciant la preuve générée par un système d'IA au regard de sa conformité aux exigences de fonctionnement, réglementaires ou non, le juge valide la pertinence de la certification dans le cadre du litige. Dans le cas des systèmes d'IA à haut risque, cela signifie que le juge confirme la pertinence des choix méthodologiques appliqués pour certifier le

¹⁶⁸⁹ Conseil d'État, 12 décembre 2007, n° 298155.

¹⁶⁹⁰ Pour une explication détaillée des enjeux juridiques derrière la conception des systèmes d'IA, Section I : L'importance des preuves au sein des normes techniques.

système d'IA ayant produit la preuve. Pour rappel, la logique de certification établie par le règlement sur l'IA contient à la fois l'interprétation d'exigences juridiques, dont les droits et libertés fondamentaux, et une évaluation de l'acceptabilité des atteintes du système d'IA à ces mêmes droits et libertés au travers de preuves quantitatives¹⁶⁹¹. Les interprétations juridiques sous-jacentes aux choix des preuves techniques sont déterminées par des organismes de normalisation dont la légitimité est contestable. Le contrôle de ces choix par les juges européens se révèle très réduit alors que ces-derniers sont les gardiens des droits et libertés¹⁶⁹². Par conséquent, l'utilisation du mécanisme de certification européen par le juge pour apprécier la preuve produite par un système d'IA revient à valider implicitement, dans le contexte spécifique de la preuve judiciaire, son bien-fondé au regard des libertés et des droits fondamentaux. La certification européenne devient un critère indirect d'appréciation de la preuve par un système d'IA.

L'expert mandaté pour vérifier la qualité réelle du fonctionnement du système d'IA ayant produit la preuve discutée. Dans un second temps, le juge diligente un expert judiciaire pour évaluer la qualité de fonctionnement du système d'IA à haut risque ayant servi de procédé de preuve, malgré la présence d'un marquage CE. L'expertise judiciaire pourra alors servir d'espace de discussion et de contestation de la fiabilité du système d'IA utilisé pour produire la preuve. Dans cette hypothèse, l'expert judiciaire peut examiner le fonctionnement du système d'IA selon la méthodologie qu'il considère la plus adaptée, qu'il s'agisse de celle contenue dans les normes européennes harmonisées ou non. En ayant l'opportunité de tester différemment le fonctionnement du système d'IA, l'expert peut mettre en avant auprès du juge et des parties certaines limites du logiciel qui seraient invisibles avec d'autres procédures de test. Indirectement, il met en lumière les enjeux autour des éléments techniques de preuve utilisés pour certifier le système d'IA à haut risque. Par exemple, c'est en confrontant différentes méthodes de test sur le logiciel d'IA COMPAS¹⁶⁹³ qu'a été mise en avant l'importance du choix d'indicateurs d'équité utilisés pour mesurer les biais du système d'IA. En effet, la métrique de

¹⁶⁹¹ Sur les valeurs contenues dans le processus de certification des systèmes d'IA, en particulier leurs exigences de fonctionnement, voir Section I : L'importance des preuves au sein des normes techniques.

¹⁶⁹² Voir Titre II : La place de la preuve dans les exigences de conception d'un système d'IA. Voir aussi VEALE Michael et ZUIDERVEEN BORGESIJUS Frederik, « Demystifying the Draft EU Artificial Intelligence Act », *Computer Law Review International*, 22, 2021.

¹⁶⁹³ Pour rappel, COMPAS est un logiciel de prédiction de récidive utilisé au sein de certains tribunaux états-unien.

détection des biais discriminatoires utilisée par Northpoint, le fournisseur de COMPAS, entraînait des conclusions différentes de celles utilisées par les journalistes de ProPublica¹⁶⁹⁴.

Dans cette situation, l'application du droit de la preuve peut alors avoir des conséquences sur les règles de conception des systèmes d'IA, et remettre ainsi en question les choix réalisés par les acteurs des normes techniques ou du fournisseur du système d'IA lui-même. Dès lors que le juge octroie des prérogatives d'investigation suffisantes à l'expert, le rapport d'analyse peut révéler un niveau de biais discriminant, un taux d'erreur différent de celui contenu dans la documentation technique communiquée par le fournisseur du système d'IA. En présence de résultats différents, le juge peut ainsi forger son évaluation de la force probante de la preuve par un système d'IA au regard de la méthodologie. En ce qui concerne le respect des droits fondamentaux par le système d'IA, l'expert peut décider de choisir la méthodologie qui lui apparaît comme la plus conforme à l'évaluation de la proportionnalité des droits fondamentaux telle que réalisée par les tribunaux. Par le recours à un expert malgré la présence d'un marquage CE, le juge évite de valider implicitement les choix de conception prononcés par des acteurs privés et leurs interprétations juridiques sous-jacentes. Si aujourd'hui les experts en technologies d'IA peuvent manquer sur les listes fournies par les tribunaux, cette lacune devrait être comblée rapidement¹⁶⁹⁵ et la désignation d'un expert compétent par le juge sera rapidement facilitée.

§2. L'appréciation du rapport d'expertise par le juge

Annonce de plan. L'intensité de l'action de l'expert sur l'articulation entre les preuves techniques de certification et l'évaluation de la force probante dépend de la force probante que le juge va accorder à son rapport d'expertise. De façon générale, l'influence de l'avis de l'expert sur le juge peut être constatée grâce à l'étude du code civil et de la jurisprudence **(A)**. De façon spécifique, et en l'absence d'une jurisprudence conséquente, cette articulation dans le cadre de l'IA sera étudiée en s'inspirant des preuves scientifiques déjà utilisées en droit positif **(B)**.

¹⁶⁹⁴ JULIA ANGWIN Jeff Larson, « Machine Bias », sur *ProPublica*, ProPublica, publié le 23 mai 2016. MATTU Jeff Larson Julia Angwin, Lauren Kirchner, Surya, « How We Analyzed the COMPAS Recidivism Algorithm », sur *ProPublica*.

¹⁶⁹⁵ À titre d'exemple, le réseau européen des instituts de la police scientifique (*European Network of Forensic Sciences Institutes* ou *ENFSI*) met l'étude des technologies d'IA comme techniques criminalistiques dans ses priorités. EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES, « Vision of the European Forensic Science Area - 2030 ». P.2.

A- L'influence concrète du rapport d'expertise sur l'appréciation du juge.

Le rapport d'expertise comme simple élément d'aide à la décision. « *Le juge n'est pas lié par les constatations ou les conclusions du technicien.* »¹⁶⁹⁶. Par cette formulation, le code de procédure civile énonce une idée claire : le rapport d'expertise ne présente aucune force contraignante sur l'appréciation des faits par le juge. Outre l'absence de contrainte sur la liberté d'appréciation du juge, l'article n'offre aucune information sur l'influence exercée par l'expertise dans le rendu du jugement si ce n'est que cette dernière peut n'en avoir aucune.

L'influence reconnue du rapport d'expertise sur l'appréciation de l'expert. À la différence du code de procédure civile, la jurisprudence, notamment européenne, a reconnu à plusieurs reprises l'influence et l'importance de l'expertise dans l'appréciation des faits par le juge. Dès 1986, la CEDH fait remarquer dans l'arrêt *Feldbrugge c. Pays-Bas* que le juge s'est limité « à entériner l'avis de l'expert »¹⁶⁹⁷. De façon similaire dans l'arrêt *Mantovanelli c. France*, la Cour qualifie de « prépondérante¹⁶⁹⁸ » l'influence des conclusions de l'expert sur l'appréciation des faits par le juge. À chaque fois, la CEDH reconnaît l'absence de contrainte juridique mais ne ferme pas non plus les yeux sur l'influence concrète des résultats de l'expertise. C'est d'ailleurs la reconnaissance de l'influence de l'expertise judiciaire qui a conduit la CEDH à reconnaître l'obligation de respecter le principe du contradictoire pendant et après la réalisation de l'expertise¹⁶⁹⁹. Comme mentionné précédemment, un jugement ne peut pas être considéré comme contradictoire si les opérations d'expertise ordonnées par le juge n'ont pas été réalisées de façon contradictoire. Par conséquent, il est nécessaire d'explorer la manière dont le juge pourrait interpréter le rapport d'expertise portant ou utilisant une preuve par un système d'IA afin de mettre en lumière les conséquences des preuves de fonctionnement du système d'IA sur le processus d'appréciation du résultat par le juge.

¹⁶⁹⁶ Article 246 du code de procédure civile.

¹⁶⁹⁷ Cour européenne des droits de l'Homme, 29 mai 1986, n° 53268/08, 53273/08, ..., 25125/12. §44.

¹⁶⁹⁸ « bien que le tribunal administratif ne fût pas juridiquement lié par les conclusions de l'expertise litigieuse, celles-ci étaient susceptibles d'influencer de manière prépondérante son appréciation des faits. » - Cour européenne des droits de l'Homme, 18 mars 1997, n° 29266/08, 32422/08, ..., 27813/13. §36.

¹⁶⁹⁹ LECLERC Olivier, *Le juge et l'expert*, LGDJ, 2005. P. 355.

B- Le comportement du juge face à l'expertise relative à la preuve par un système d'IA

Les comportements envisageables du juge face au rapport d'expertise. Une partie significative du rôle de l'expert dans l'appréciation des preuves produites par un système d'IA dépend de la manière dont le juge utilise son rapport d'expertise dans son raisonnement probatoire. Du fait de l'aspect prospectif de notre analyse sur les preuves par un système d'IA, le choix a été fait de procéder à un raisonnement par analogie à partir des réactions observées du juge vis-à-vis du rapport d'expertise scientifique produit dans le contexte d'un contentieux en droit de la responsabilité.

Dans une étude de 2017, les professeurs L. Khoury et E. Vergès réalisent une analyse de la jurisprudence québécoise et française et mettent en lumière cinq niveaux de réactions du juge face à des preuves scientifiques¹⁷⁰⁰. Deux d'entre eux sont pertinents pour notre analyse prospective de la réaction du juge face au rapport d'expertise relatif à des preuves par un système d'IA¹⁷⁰¹.

Premièrement, le juge s'appuie directement sur les conclusions du rapport d'expertise qu'il approuve, c'est « *l'adhésion à une opinion expertale* »¹⁷⁰². Aussi bien en droit québécois qu'en droit français¹⁷⁰³, ce type de comportement du juge est très fréquent. Il se décompose en deux étapes avec, tout d'abord, l'appropriation des conclusions de l'expert judiciaire par le juge et, ensuite, la qualification juridique des conclusions. Au sein d'un litige en responsabilité, cela peut prendre la forme d'un juge qui se fonde, dans un premier temps, sur les analyses d'un expert judiciaire en médecine pour constater, dans un second temps, l'absence du préjudice pour le demandeur¹⁷⁰⁴. Dans le contexte de l'appréciation d'une preuve produite par un système d'IA, ce raisonnement pourrait ainsi prendre plusieurs formes. Par exemple, un juge se fonde

¹⁷⁰⁰ VERGÈS Etienne et KHOURY Lara, « Le traitement judiciaire de la preuve scientifique », *cd1*, 58, Faculté de droit de l'Université Laval, 2017.

¹⁷⁰¹ Le périmètre de l'étude est plus large que celui de la présente analyse. Ainsi, les faits scientifiques sont classés par les auteurs en trois catégories : les faits bruts (dont les statistiques), les faits interprétés (à savoir l'opinion de l'expert) et les raisonnements scientifiques. À cette étape de notre démonstration, seule la seconde catégorie de faits scientifiques est jugée pertinente. Par conséquent, seules les réactions du juge associées à une opinion d'expert seront traitées.

¹⁷⁰² VERGÈS Etienne et KHOURY Lara, *op. cit.* P. 533.

¹⁷⁰³ Pour un exemple de ce type de comportement du juge, voir notamment : Aix-en-Provence, 24 juin 2011, n° 09/18929 : *Juris-Data no 2011-014016* ; Toulouse, 21 juill. 2011, n° 08/04148 : *Juris-Data no 2011-027552*.

¹⁷⁰⁴ « [l]e professeur [X], expert commis par la cour par l'arrêt du 26 janvier 2010, conclut dans son rapport que les différents examens réalisés le 24 mars 2010 lors de l'expertise — audiométries tonale et vocale, tympanogramme, potentiels évoqués auditifs et vidéonygstamographique — ne permettent pas de retrouver de trouble auditif ou de trouble vestibulaire chez [le demandeur] » - Toulouse, 21 juill. 2011, no 08/04148 : *Juris-Data no 2011-027552*. Cité dans *Ibid.*

sur le rapport d'expertise, qui énonce la conception et le fonctionnement correct du système d'IA aux exigences de l'état de l'art et qui présente un taux d'exactitude moyen de 95%, pour en déduire que l'identification du suspect sur la base de données biométriques (démarche ou visage par exemple) à l'aide dudit logiciel est un élément de preuve probant participant à la reconnaissance de la culpabilité de l'individu.

La seconde réaction observée par les auteurs est celle de « *l'émancipation du juge* » par rapport aux conclusions de l'expert¹⁷⁰⁵. Dans cette situation, le juge agit alors de façon autonome et vient apprécier les preuves scientifiques brutes indépendamment de leur interprétation par l'expert judiciaire. Cela ne signifie pas pour autant que le juge ignore complètement le rapport d'expertise, il peut en tenir compte et, de façon simultanée, se forger sa propre opinion à partir des éléments de preuve bruts. Il produit alors sa propre interprétation des faits scientifiques bruts, en parallèle de celle de l'expert. Ce type de réaction a été notamment observée lorsque le droit disposait de standards de comportement pour qualifier les faits. Dès lors, le juge possédait une grille de lecture juridique des faits scientifiques et a ainsi décidé de l'appliquer en complément de celle de l'expert¹⁷⁰⁶. Cette prise d'autonomie quant à l'avis de l'expert va dans le sens de la jurisprudence européenne et de son application du principe de précaution en cas d'incertitude scientifique¹⁷⁰⁷. Malgré les avancées de la recherche, les risques liés à l'utilisation des technologies d'IA sont encore empreints d'une incertitude scientifique. De plus, lesdits risques dépassent la seule compétence de l'expert évaluant le système d'IA (ou bien utilisant le système d'IA à des fins probatoires) car ils peuvent englober des risques juridiques, éthiques, politiques et informatiques. Dès lors, le juge peut décider de s'écarter du rapport d'expertise car l'expert n'est pas compétent pour répondre aux problèmes dans toutes ses dimensions¹⁷⁰⁸. En effet, la détermination d'un fonctionnement acceptable d'un système d'IA, et donc de la fiabilité de son résultat en tant que preuve, impose de valider les métriques utilisées et les seuils statistiques sélectionnés lors de sa conception. Ces choix cristallisent des enjeux juridiques importants (la formalisation informatique de la discrimination ou de la liberté d'expression par

¹⁷⁰⁵ VERGÈS Etienne et KHOURY Lara, *op. cit.* P. 540.

¹⁷⁰⁶ Aix-en-Provence, 29 févr. 2012, no 10/04437 : Juris-Data no 2012-004869. En l'espèce, un arrêté ministériel limitait l'utilisation d'un test de tuberculose aux enfants de moins de trois ans. C'est au travers de ce standard réglementaire que le juge a interprété les éléments scientifiques contenus dans le dossier, aboutissant aux mêmes conclusions que l'expertise judiciaire.

¹⁷⁰⁷ Tribunal de l'Union européenne, 11 septembre 2002, n° T-13/99 Tribunal de l'Union européenne, 11 septembre 2002, n° T-70/99.

¹⁷⁰⁸ LECLERC Olivier, *Le juge et l'expert*, LGDJ, 2005. P. 182.

exemple) qui nous semble entrer dans les contextes où le tribunal de l'Union européenne reconnaît l'intérêt à se détacher de l'avis de l'expert¹⁷⁰⁹.

Les conséquences du comportement du juge sur l'articulation entre les preuves techniques du fonctionnement conforme du système d'IA sur l'appréciation du résultat généré comme preuve. Qu'il utilise un système d'IA pour réaliser son expertise ou qu'il évalue la fiabilité d'un système d'IA ayant produit une preuve, l'expert judiciaire fait usage de sa légitimité scientifique pour valider (ou à l'inverse mettre en doute) un élément de preuve produit par un système d'IA. Pour reprendre les termes du professeur O. Leclerc, cet aspect du travail de l'expert consiste alors à stabiliser la connaissance scientifique par son expérience afin de faciliter l'appréciation de la preuve par le juge¹⁷¹⁰. Dans le cas des systèmes d'IA à haut risque soumis à une certification préalable, cette stabilisation de la connaissance scientifique vient se superposer à celle déjà débutée par les organismes de normalisation par le biais des normes européennes harmonisées et autres normes techniques de conception des systèmes d'IA. Les deux processus peuvent alors s'aligner, renforçant ainsi la stabilité de la connaissance scientifique sur la fiabilité du fonctionnement d'un système d'IA, ou diverger, le juge devant alors décider quelle expertise privilégier pour l'appréciation de l'élément de preuve. Par conséquent, il est clair que l'expert judiciaire contribue à l'articulation entre les preuves techniques utilisées pour certifier les systèmes d'IA et le processus d'appréciation du résultat comme élément de preuve.

CONCLUSION DE LA SECTION II :

Face à un élément technique de preuve particulièrement complexe, le recours par le juge à un expert judiciaire pour éclairer son travail d'appréciation est une solution usuelle. Dès lors, l'expertise apparaît comme un lieu intermédiaire privilégié d'interactions entre des éléments scientifiques et le processus d'appréciation de l'élément de preuve. Selon la mission qui lui est confiée par le juge, l'expertise judiciaire peut jouer différents rôles dans cette articulation. En mettant le système d'IA au service de sa mission, l'expert contribue à valider et légitimer scientifiquement la fiabilité du fonctionnement du système d'IA auprès du juge en fondant, au moins partiellement, le rapport d'expertise sur le résultat généré. En revanche, l'expert saisi pour interpréter ou évaluer une preuve produite par un système d'IA représente une occasion

¹⁷⁰⁹ Tribunal de l'Union européenne, 11 septembre 2002, n° T-13/99 Tribunal de l'Union européenne, 11 septembre 2002, n° T-70/99. Dans les deux arrêts mentionnés, l'expertise est séparée de la décision publique mais cette prise de recul par rapport à l'expertise peut tout à fait être appliquée au juge et à l'expertise judiciaire.

¹⁷¹⁰ LECLERC Olivier, *op. cit.*

de discuter de la fiabilité du fonctionnement du procédé de preuve. Son travail contribue à mettre en lumière les choix des éléments techniques de preuves réalisés par le fournisseur du système d'IA, que lui ou les parties peuvent alors décider de discuter. Dans le cadre d'un litige impliquant une preuve générée par système d'IA à haut risque, le recours à un expert judiciaire constitue une reconnaissance d'une interaction entre les éléments techniques de preuves, issus du règlement sur l'IA, avec le processus d'appréciation pour le litige en cours. Néanmoins, le choix de recourir à un expert, comme la force probante du rapport d'expertise, est laissé à la libre appréciation du juge et demeure donc à ce jour impossible à déterminer précisément.

Section III : La détermination préalable de règles d'appréciation de la preuve par un système d'IA

Introduction. Au sein du droit de la preuve, il existe une dichotomie doctrinale entre la théorie générale de la preuve et le droit des preuves spéciales. D'un côté, la théorie générale de la preuve renvoie aux règles qui s'appliquent l'ensemble des modes de preuves existants. De l'autre, pour reprendre la définition des professeurs Vergès, Vial et Leclerc, « *le droit des preuves spéciales désigne les règles particulières qui s'appliquent à chaque mode de preuve* »¹⁷¹¹. Les règles spécifiques à ces preuves spéciales peuvent alors préciser les règles générales ou y déroger. L'existence de règles d'appréciation pour un mode de preuve spécifique limite la liberté du juge dans son évaluation de la force probante dudit mode de preuve. Les réflexions menées jusqu'à présent dans ce chapitre étaient présentées dans un contexte de libre appréciation des preuves par un système d'IA par le juge. Dans cette ultime étape de notre démonstration, il sera abordé la possibilité que des règles d'appréciation de la force probante soient élaborées spécifiquement pour les preuves produites par un système d'IA. Les éléments techniques de preuves de la conformité au règlement sur l'IA pourraient alors contribuer à l'élaboration d'un régime de preuve spéciale pour le résultat généré par un système d'IA.

Annnonce de plan. Les conséquences des caractéristiques techniques d'un procédé de preuve dans l'élaboration de règles spécifiques à l'appréciation des résultats qu'il produit est un phénomène qui peut prendre deux formes distinctes. Premièrement, les règles peuvent interdire au juge la prise en compte desdites propriétés techniques, fixant alors de façon systématique la

¹⁷¹¹ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 442.

force probante des preuves concernées (**Sous-section 1**). Secondement, le juge peut également se voir imposer ces éléments techniques comme critères d'appréciation de l'élément de preuve généré (**Sous-section 2**).

Sous-section 1 : L'exclusion du fonctionnement du système d'IA dans l'appréciation de son résultat

Annonce de plan. La fixation d'une valeur probante pour un mode de preuve est un mécanisme fort qui réduit drastiquement la liberté d'appréciation du juge¹⁷¹². Pour autant, le législateur français y a déjà eu recours à plusieurs reprises et de nombreux exemples existent en droit positif (**§1**). Ainsi, il doit être envisagé la possibilité qu'il décide de limiter les risques de variation de l'appréciation judiciaire des résultats produits par un système d'IA en décidant lui-même de la force probante à accorder aux résultats produits par des systèmes d'IA (**§2**).

§1. La fixation de la force probante des modes de preuve en droit positif

Annonce de plan. L'élaboration, généralement par le législateur, d'une force probante globale à un mode de preuve peut résulter en deux situations. Premièrement, le cas le plus fréquent est l'augmentation de la force probante d'un mode de preuve afin qu'il bénéficie d'une autorité supérieure aux autres modes de preuve (**A**). Secondement, la fixation de la force probante peut aussi entraîner une diminution de la force probante du mode de preuve en question (**B**).

A- L'augmentation de la force probante

Les raisons d'une augmentation de la force probante. L'augmentation de la force probante d'un mode de preuve particulier le fait prévaloir sur d'autres modes de preuve, dépourvus de force probante accrue, qui lui seraient contradictoires. Cette décision peut être justifiée par un

¹⁷¹² La liberté d'appréciation ne disparaît pas totalement pour autant. Lorsque deux preuves contraires ont une force probante légale identique, le juge retrouve sa liberté d'évaluation de leur force probante pour choisir celle qui emporte son intime conviction.

besoin de sécurité juridique pour les parties¹⁷¹³. En d'autres termes, la création d'une preuve légale offre une prévisibilité plus grande quant à l'issue d'un litige. La réduction de la liberté d'appréciation du juge et l'élaboration d'une hiérarchie probatoire relativement claire sont ainsi deux moyens efficaces de réduire l'incertitude judiciaire en cas de futur litige.

Une seconde raison de l'accroissement de la force probante d'un mode de preuve est liée la fiabilité dudit mode de preuve. Celle-ci serait si importante qu'elle surpasserait celle des modes de preuves dont la force probante est appréciée librement. Par exemple, cette confiance dans la fiabilité renvoie aux preuves émises par un officier public ou un avocat. Ainsi, depuis l'ordonnance de 2016¹⁷¹⁴, l'acte produit par un officier public est un acte authentique¹⁷¹⁵ qui ne peut être contesté que par le biais d'une procédure d'inscription en faux¹⁷¹⁶. La procédure particulière de contestation d'un acte authentique implique que celui-ci ne peut pas être contesté par n'importe quel élément de preuve, ces derniers ayant une force probante plus faible. Dans la même logique, l'acte sous signature privée contresigné par les avocats des parties possède une force probante supérieure en ce qui concerne l'identité des parties par rapport à un acte sous signature privé qui n'aurait pas été contresigné. Le code civil décrit cet acte comme faisant « *foi de l'écriture et de la signature des parties* »¹⁷¹⁷. Cette augmentation de la force probante de ces preuves lie le juge à la fois pour les contentieux de droit civil et pour les contentieux administratifs. Par exemple, en droit commercial, la force probante accrue de l'acte authentique vient se heurter et déroger au principe de liberté totale de la preuve¹⁷¹⁸. En droit administratif, les juges accordent aussi une force probante accrue aux actes d'officier public¹⁷¹⁹ mais également aux actes et aux mentions « *apportées sous la responsabilité du maire* »¹⁷²⁰ qui font alors foi jusqu'à la preuve du contraire. De façon similaire en droit pénal, les procès-verbaux et les rapports d'officier font foi jusqu'à la preuve du contraire¹⁷²¹, limitant la preuve contraire à

¹⁷¹³ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.* PP. 439-440.

¹⁷¹⁴ « Ordonnance n° 2016-131 du 10 février 2016 portant réforme du droit des contrats, du régime général et de la preuve des obligations », 2016.

¹⁷¹⁵ « *L'acte authentique est celui qui a été reçu, avec les solennités requises, par un officier public ayant compétence et qualité pour instrumenter.* » - Article 1369 alinéa 1 du code civil.

¹⁷¹⁶ « *L'acte authentique fait foi jusqu'à inscription de faux de ce que l'officier public dit avoir personnellement accompli ou constaté.* » - Article 1371 alinéa 1 du code civil.

¹⁷¹⁷ Article 1374 alinéa 1 du code civil.

¹⁷¹⁸ « *Le principe de liberté de la preuve, qui permet de prouver outre et contre un écrit par tous moyens, ne peut que céder le pas lorsque cet écrit consiste en un acte authentique* » - GRYNBAUM Luc, « Preuve », *Répertoire de droit commercial*, Dalloz, 2020.

¹⁷¹⁹ « *Considérant que les mentions d'un acte d'huissier font foi jusqu'à inscription de faux* » - Conseil d'État, 17 octobre 2012, n° 351833.

¹⁷²⁰ « *les mentions apportées, sous la responsabilité du maire, pour certifier le caractère exécutoire des actes des autorités communales font foi jusqu'à la preuve du contraire* » - Conseil d'État, 27 août 2014, n° 374473.

¹⁷²¹ Article 537 du code de procédure pénale.

l'écrit et au témoignage¹⁷²². En raison de la confiance dans la fiabilité de ces agents publics, les éléments qu'ils produisent se voient octroyer une force probante supérieure aux autres éléments de preuves et lient, par conséquent, le juge¹⁷²³. Enfin le critère de la fiabilité pour octroyer un accroissement de la valeur de la preuve se retrouve également dans les preuves électroniques. Dès lors qu'il présente des garanties d'intégrité¹⁷²⁴, l'écrit électronique se voit octroyer une force probante accrue. La fiabilité accordée à l'auteur humain d'une preuve est désormais exigée aussi à la technologie à l'origine de l'écrit électronique.

Les conséquences d'une augmentation de la force probante. Au fil du temps, le phénomène d'accroissement de la force probante a donné lieu à une hiérarchie entre les différents modes de preuve en droit positif. En effet, même parmi les modes de preuve bénéficiant de ce mécanisme, certains ont une force probante plus importante que les autres. Ainsi les actes authentiques présentent une force probante supérieure aux actes contresignés par avocats. L'aveu judiciaire et le serment judiciaire se trouve au sommet de cette pyramide des modes de preuves car leur force probante est telle qu'il n'est possible d'y opposer une preuve contraire. L'établissement de cette hiérarchie entraîne une automatisation du raisonnement probatoire pour le juge. L'objectif premier est l'efficacité de la procédure, la recherche de la vérité devient alors secondaire¹⁷²⁵.

B- La réduction de la force probante : la demi-preuve

La notion de demi-preuve. La notion de demi-preuve trouve son origine dans la hiérarchie des preuves mise en place du XIIème au XIVème siècle au cours de ce qui est appelé « le droit savant médiéval »¹⁷²⁶. Chaque mode de preuve existant alors se voyait octroyer un degré de force probante propre, fondant alors un système probatoire hiérarchique. Afin d'établir un fait contesté, c'est-à-dire pour convaincre l'autorité judiciaire, les parties avaient à leur disposition

¹⁷²² « La preuve contraire ne peut être rapportée que par écrit ou par témoins. » - Article 537 alinéa 3 du code de procédure pénale.

¹⁷²³ Cour de Cassation, Chambre criminelle, 18 octobre 2000, n° 00-82.68 Cour de cassation, Chambre criminelle, 13 juin 2007, n° 06-85.441.

¹⁷²⁴ Article 1366 du code civil. Pour plus de détails sur les critères de qualification de l'écrit électronique d'un acte authentique et la possibilité que cet écrit soit produit par un système d'IA, voir §2. La catégorie de l'écrit électronique difficilement envisageable.

¹⁷²⁵ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 440.

¹⁷²⁶ Jean-Philippe Lévy, *La hiérarchie des preuves dans le droit savant du Moyen-Âge depuis la Renaissance du Droit Romain jusqu'à la fin du XIVe siècle*, Paris, Sirey, « Annales de l'Université de Lyon », 1939, p. 17.

des preuves dites « pleines », des preuves dites « semi-pleines » et enfin des « indices ». La preuve pleine désignait la preuve qui offrait une certitude suffisante pour convaincre à elle seule¹⁷²⁷. Possédant ensuite une force probante moindre venait la preuve semi-pleine, dont est issue la notion de demi-preuve. Les preuves semi-pleines sont les modes de preuves incapables de fonder un jugement lorsqu'elles sont prises seules. Cette catégorie de modes de preuves rassemblait la présomption, le témoignage d'un individu, la comparaison d'écritures, etc¹⁷²⁸. Ces preuves imparfaites pouvaient être associées l'une avec l'autre afin de former une preuve pleine. La logique est ici très simple : deux moitiés de preuves forment une preuve complète. Enfin, à la base de la pyramide des modes de preuve, la catégorie des indices permettait de corroborer d'autres éléments de preuve.

La notion de demi-preuve découle de celle de preuve semi-pleine. Cependant, sa réapparition dans le droit positif¹⁷²⁹ s'accompagne d'un changement important par rapport au droit savant médiéval. Les demi-preuves ne peuvent plus être cumulées dans le but de former une preuve complète (ou une preuve parfaite). Cependant, les demi-preuves désignent toujours les preuves qui ne peuvent pas emporter, prise seule, la conviction du juge. Depuis le début de la démonstration, nous avons déjà abordé certaines de ces exceptions que sont les demi-preuves. En droit civil, le juge « *ne peut se fonder exclusivement sur une expertise réalisée à la demande de l'une des parties* »¹⁷³⁰. L'expertise réalisée en violation du principe du contradictoire subie le même traitement et voit sa force probante diminuée¹⁷³¹. Il en va de même pour les témoignages anonymes¹⁷³² et pour l'examen radiologique osseux (dans le but de déterminer l'âge de l'individu) dont la force probante est explicitement réduite par le législateur¹⁷³³. Ces

¹⁷²⁷ Cette catégorie rassemblait notamment les actes écrits rédigés par une personne publique, les actes publics ou privés rédigés dans l'intérêt de l'adversaire, les témoignages issus de deux témoins. LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *op. cit.* P. 57.

¹⁷²⁸ *Ibid.* P.57. La liste de preuves semi-pleines n'est pas et ne pourrait pas être exhaustive car le droit savant médiéval n'a jamais établi de liste finie des preuves semi-pleines.

¹⁷²⁹ La notion de demi-preuves a disparu à la Révolution française. L'article 353 du code de procédure pénale pose ainsi le principe pour les jurés de la Cour d'assises que « *la loi ne demande pas compte à chacun des juges et jurés composant la cour d'assises des moyens par lesquels ils se sont convaincus, elle ne leur prescrit pas de règles desquelles ils doivent faire particulièrement dépendre la plénitude et la suffisance d'une preuve ; [...]* »

¹⁷³⁰ Cour de cassation, Chambre mixte, 28 septembre 2012, n° 11-18.710.

¹⁷³¹ Cour de cassation, Chambre civile 1, 11 juillet 2018, n° 17-17.441 17-19.581.

¹⁷³² « *Attendu que le juge ne peut fonder sa décision uniquement ou de manière déterminante sur des témoignages anonymes ;* » - Cour de cassation, Chambre sociale, 4 juillet 2018, n° 17-18.241.

¹⁷³³ « *Les conclusions de ces examens, qui doivent préciser la marge d'erreur, ne peuvent à elles seules permettre de déterminer si l'intéressé est mineur. Le doute profite à l'intéressé.* » - Article 388 alinéa 3 du code civil.

exemples sont également présents en matière pénale où, à titre d'exemple¹⁷³⁴, l'aveu du suspect dont les droits de la défense n'ont pas été respectés ne peut fonder à lui seul la condamnation¹⁷³⁵.

Finalité de la réduction de la force probante d'un mode de preuve. Qu'elle soit d'origine législative ou jurisprudentielle, la qualification de demi-preuve est un mécanisme de protection contre les modes de preuve auxquels il n'est pas possible d'avoir une pleine confiance. Premièrement, cette défiance se justifie par le manque de fiabilité du mode de preuve qui entache la preuve produite. En ce qui concerne l'examen radiologique osseux pour déterminer l'âge d'un individu par exemple, le conseil constitutionnel, par une décision n° 2018-768 du 21 mars 2019, mentionne la présence d'une marge d'erreur significative qui justifie l'impossibilité pour le juge de se fonder uniquement sur le résultat de ce test pour rendre sa décision¹⁷³⁶. Deuxièmement, l'amoindrissement de la force probante est la conséquence de l'atteinte aux droits de la défense. Ainsi, la méconnaissance du principe du contradictoire par l'expert transforme le rapport d'expertise en demi-preuve. Contrairement à l'augmentation de la force probante, la notion juridique de demi-preuve pose la recherche de la vérité comme objectif principal.

§2. L'élaboration de règles spécifiques d'appréciation des preuves par un système d'IA

Annonce de plan. Au regard de l'exactitude limitée, du risque de biais discriminatoire, de l'imprévisibilité ou encore de l'opacité des technologies d'IA, le système d'IA ne remplit aucunement les critères justifiant une augmentation de la force probante de son résultat. *A contrario*, ces caractéristiques techniques permettent de raisonnablement envisager l'hypothèse

¹⁷³⁴ Pour plus d'exemples, voir LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P. 454.

¹⁷³⁵ « En matière de crime ou de délit, le droit de se taire sur les faits qui lui sont reprochés est notifié à toute personne suspectée ou poursuivie avant tout recueil de ses observations et avant tout interrogatoire, y compris pour obtenir des renseignements sur sa personnalité ou pour prononcer une mesure de sûreté, lors de sa première présentation devant un service d'enquête, un magistrat, une juridiction ou toute personne ou tout service mandaté par l'autorité judiciaire. Aucune condamnation ne peut être prononcée sur le seul fondement de déclarations faites sans que ledit droit ait été notifié. » - Article préliminaire du code de procédure pénale.

¹⁷³⁶ « Les dispositions contestées autorisent le recours à un examen radiologique osseux aux fins de contribuer à la détermination de l'âge d'une personne. En l'état des connaissances scientifiques, il est établi que les résultats de ce type d'examen peuvent comporter une marge d'erreur significative. » - Conseil constitutionnel, 21 mars 2019, n° 2018-768. Par. 7.

d'une réduction de la force probante lorsque le résultat est utilisé comme une information signifiante **(A)** et d'explorer les conséquences qu'emporterait une telle décision **(B)**.

A- Le résultat généré par un système d'IA réduit à une demi-preuve

Les caractéristiques de la preuve par un système d'IA en faveur d'une réduction de leur force probante. Lorsqu'il est employé comme information signifiante, le résultat généré par un système d'IA partage plusieurs propriétés techniques en commun avec les demi-preuves existantes en droit positif.

La première d'entre elles est la fiabilité du système d'IA. Au préalable, il est important de reprendre la distinction entre le résultat du système d'IA pris en tant qu'élément existant et le résultat du système d'IA pris en tant qu'information signifiante¹⁷³⁷. Le présent titre s'intéresse à la portée des preuves issues de la conformité *ex ante* sur la preuve produite par un système d'IA. Par conséquent, seuls les résultats pris comme informations significatives sont pertinents à traiter en l'espèce. Pour ce type de résultat de systèmes d'IA, la fiabilité de la preuve peut être extrêmement variable selon le type de systèmes d'IA et la technologie utilisée. Pour les cas d'application simples, un système d'IA peut obtenir des résultats très fiables. Néanmoins, les technologies plus avancées d'IA, telle que l'apprentissage automatique, sont généralement employées dans des situations complexes là où des technologies plus traditionnelles révèlent leurs limites. Par conséquent, la fiabilité de ces systèmes d'IA reste variable mais ne sera jamais parfaite. Un taux de précision de 95% est un taux excellent dans l'état de l'art et pour autant il signifie, qu'en moyenne, un résultat sur vingt est une erreur. Cette fiabilité très relative offre un parallèle intéressant avec l'examen radiologique osseux dont la fiabilité très moyenne a conduit le législateur français à imposer une réduction de sa force probante¹⁷³⁸. À l'image du système d'IA, l'examen radiologique osseux a aussi une fiabilité variable selon les situations. Cette variabilité est prise en compte par l'article 388 du code civil, qui impose que les conclusions de l'examen « *doivent préciser la marge d'erreur* »¹⁷³⁹, sans que cela affecte pour autant sa qualification en tant que demi-preuve. En appliquant un raisonnement par analogie, il apparaît

¹⁷³⁷ La distinction est notamment inspirée de l'article suivant : ROTH Andrea L., « Machine Testimony », *The Yale Law Journal*, 126, 2017. Pour rappel, la première catégorie de résultat par système d'IA ne dépend pas de la qualité de fonctionnement du système d'IA, contrairement à la seconde. Pour plus de détails, voir §3. Le résultat d'un système d'IA comme indice au service d'une présomption de fait.

¹⁷³⁸ Article 388 alinéa 3 du code civil.

¹⁷³⁹ *Ibid.*

donc que la fiabilité relative des systèmes d'IA est un argument pouvant justifier une intervention législative visant à réduire la force probante des preuves par un système d'IA.

La seconde caractéristique à mettre en lumière est l'opacité des technologies d'IA. Tout d'abord, l'opacité intrinsèque au mode de fonctionnement des systèmes d'IA limite très souvent la compréhension par les tiers, voire du concepteur lui-même, du fonctionnement du système d'IA quant à la production d'un résultat en particulier. Cet état de fait pourrait soulever des difficultés quant à son respect du principe du contradictoire. Par exemple, c'est l'impossibilité d'interroger le témoin anonyme qui a conduit les juridictions pénales¹⁷⁴⁰ et civiles¹⁷⁴¹ à considérer ce mode de preuve comme une demi-preuve. Au moyen d'un raisonnement par analogie, il est donc envisageable de vouloir aligner la force probante de la preuve par un système d'IA à celui des témoignages anonymes car tous deux portent atteinte au principe du contradictoire. De plus, la preuve par un système d'IA peut être produite préalablement au litige et donc sans la présence des différentes parties. Par exemple, un système d'IA chargé de détecter l'état de somnolence d'un conducteur produit son résultat au moment de faits. Un autre exemple serait un particulier ou une entreprise qui, à la suite d'un cambriolage, utiliserait son propre logiciel afin de collecter des informations identifiantes sur l'auteur de l'infraction. Dans ces situations, la preuve est produite de façon non contradictoire. Le parallèle avec l'expertise de partie et l'expertise judiciaire non contradictoire est alors intéressant, ces dernières étant réalisées sans les garanties associées au principe du contradictoire. À nouveau, le manquement aux droits de la défense par un mode de preuve est une justification pour une réduction globale de sa force probante, aussi bien en matière civile que pénale¹⁷⁴². Dès lors, l'opacité et la fiabilité des systèmes d'IA sont deux caractéristiques, déjà retenues en droit positif, pouvant permettre de justifier une réduction globale de la force probante des systèmes d'IA comme mode de preuve.

L'influence de la conformité *ex ante* des systèmes d'IA dans la réduction de leur force probante. L'élaboration de dispositions réglementaires, le règlement européen sur l'IA¹⁷⁴³ en

¹⁷⁴⁰ Cour européenne des droits de l'Homme, 20 novembre 1989, n° 62171/00. Par la suite, le législateur français a intégré cette réduction de la force probante des témoignages anonymes en droit pénal dans la loi en 2001 : « Loi n° 2001-1062 du 15 novembre 2001 relative à la sécurité quotidienne », 2001. La disposition est aujourd'hui codifiée à l'article 706-62 du code de procédure pénale.

¹⁷⁴¹ Ce raisonnement est appliqué par la chambre sociale de la Cour de cassation : Cour de cassation, Chambre sociale, 4 juillet 2018, n° 17-18.241.

¹⁷⁴² C'est notamment le cas de l'aveu fournis par un individu sans la présence de son avocat : « *En matière criminelle et correctionnelle, aucune condamnation ne peut être prononcée contre une personne sur le seul fondement de déclarations qu'elle a faites sans avoir pu s'entretenir avec un avocat et être assistée par lui.* » - Article préliminaire du code de procédure pénale.

¹⁷⁴³ « Règlement sur l'intelligence artificielle », 2024.

étant la figure de proue, peut et pourra affecter ces caractéristiques. L'existence d'exigences relatives au fonctionnement des systèmes d'IA à haut risque pourrait réduire l'intensité du besoin d'intervention, législative ou jurisprudentielle, sur la force probante des preuves produites par des systèmes d'IA de manière globale. L'influence est particulièrement envisageable en ce qui concerne la fiabilité du système d'IA. Dès lors que la mise sur le marché d'un système d'IA est conditionnée à des exigences de fiabilité, cela offre une certaine garantie quant à l'exactitude des résultats générés. En ce qui concerne l'opacité, une influence similaire pourrait être constatée avec les exigences relatives à la transparence¹⁷⁴⁴ et à l'explicabilité des systèmes d'IA¹⁷⁴⁵.

Cependant la prise en compte de garanties réglementaires relatives à la fiabilité et à l'opacité des systèmes d'IA pour imposer une force probante aux preuves par un système d'IA doit être nuancée.

Premièrement, la portée des garanties réglementaires est limitée aux systèmes d'IA contenus dans leur champ d'application. Pour prendre l'exemple du règlement sur l'IA, seuls les fournisseurs de systèmes d'IA à haut risque pourraient se targuer de garantir un certain niveau de fiabilité et de transparence de leurs produits. La certification obtenue pour commercialiser ou mettre en service leur système d'IA leur permettrait de se défendre contre une décision de réduction de la force probante des résultats produits par leurs systèmes d'IA. Dans cette hypothèse, seuls les résultats de systèmes d'IA qui ne sont soumis à aucune exigence de conformité *ex ante*, relative à la fiabilité et à l'opacité, pourraient être visés par le statut de demi-preuve.

Deuxièmement, les exigences de conformité *ex ante* ne doivent avoir une influence sur l'imposition d'une force probante seulement si elles disposent d'un niveau de garantie minimum. Par exemple, le règlement européen sur l'IA impose « *d'atteindre un niveau approprié d'exactitude, de robustesse et de cybersécurité [...]* »¹⁷⁴⁶. Formulé ainsi, l'exigence ne garantit aucun niveau d'exactitude minimal pour tous les systèmes d'IA à haut risque. Il est possible que ce constat change avec l'élaboration des normes européennes harmonisées qui

¹⁷⁴⁴ Voir, par exemple : *Ibid.* Article 13. « Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. Article 22.

¹⁷⁴⁵ « Règlement sur l'intelligence artificielle », 2024. Article 86. En droit national, les obligations imposées aux décisions individuelles administratives prises par un système algorithmique peuvent également entrer dans la réflexion.

¹⁷⁴⁶ *Ibid.* Article 15.1.

pourraient imposer un seuil statistique minimal à atteindre en termes d'exactitude. De la même manière, la production de documentation par les différents acteurs de la phase de conformité *ex ante* (autorités de contrôle, autorités de surveillance du marché, autorité notifiante, organismes notifiés, etc.) peut mettre en avant un seuil statistique d'exactitude minimal pour qu'un système d'IA puisse obtenir sa certification. Cette fixation d'un seuil statistique minimal doit s'accompagner de lignes directrices sur le choix des indicateurs à utiliser, sans quoi l'élaboration d'un seuil minimal n'a aucun sens. Par conséquent, en l'absence d'accord sur les éléments techniques de preuve relatifs à une qualité de fonctionnement des systèmes d'IA, les exigences réglementaires ne peuvent constituer une réelle garantie au niveau probatoire. Par conséquent, elles ne pourraient pas apparaître comme un argument solide contre le choix de réduire la force probante des preuves générées par un système d'IA.

Troisièmement et dernièrement, les spécificités des technologies d'IA posent des limites qui ne pourront pas être dépassées malgré la présence de garanties réglementaires sérieuses. Ce constat peut être effectué particulièrement pour les approches reposant sur l'apprentissage automatique. Malgré une amélioration de l'exactitude des systèmes d'IA, ces derniers raisonneront toujours en termes de rapports de corrélation et non de rapports de causalité. Le résultat produit par un système d'IA est une présomption statistique tirée de données. Il ne s'agit en aucun cas d'un fait certain. Pour reprendre la formule de Domat sur la justesse des conséquences tirées de présomption : *« la certitude ou l'incertitude des présomptions [...] dépend de la certitude ou l'incertitude des faits dont on tire les présomptions, et de la justesse des conséquences qu'on tire de ces faits pour la preuve de ceux dont il s'agit »*. L'utilisation d'un résultat produit par un système d'IA comme seul argument d'une présomption judiciaire est alors très discutable. Par conséquent, la décision du juge serait alors fondée sur une double présomption, affectant directement la qualité et la confiance dans le raisonnement du juge.

B- Les conséquences du statut de demi-preuve pour le résultat d'un système d'IA

La hiérarchisation de la preuve par un système d'IA. La réduction de la force probante du résultat généré par un système d'IA la placerait automatiquement au bas de la hiérarchie des preuves existantes. En d'autres termes, le statut de demi-preuve impose au juge de faire prévaloir toute preuve contraire au résultat d'un système d'IA, à l'exception des autres demi-preuves. Prenons par exemple le cas d'un système d'IA de reconnaissance faciale. Avec le statut

de demi-preuve, même si le système d'IA identifie à 99,9% un individu sur une caméra, n'importe quel témoignage contraire à cette version empêchera le juge de prendre en compte le résultat, pris seul, dans son raisonnement. Dans un premier temps, cela signifie que la preuve par un système d'IA peut être utilisée seulement pour contredire une autre demi-preuve. Lorsque deux preuves ayant une force probante similaire s'oppose, le juge retrouve sa liberté d'appréciation et tranche celle qu'il considère comme la plus convaincante¹⁷⁴⁷. Dans un second temps, cela signifie que la fiabilité du résultat individuel n'est plus, en principe, un élément à prendre en compte pour le juge¹⁷⁴⁸. En d'autres termes, le fait que le système d'IA soit confiant à 95% ou à 50% du résultat qu'il a produit et que ce résultat, pour une identification par exemple, soit de 20% ou 80% n'entraîne aucune conséquence sur l'appréciation de sa force probante. Celle-ci serait alors fixe. La force probante du résultat deviendrait indépendante de la fiabilité technique du système d'IA. Par conséquent, la décision de qualifier la preuve produite par un système d'IA de demi-preuve supprime, dans la plupart des situations, l'influence des règles et preuves techniques de conception *ex ante* des systèmes d'IA sur l'appréciation de la force probante par le juge.

La réduction de la place du résultat d'un système d'IA dans le processus décisionnel. La réduction de la force probante des résultats d'un système d'IA entraînerait des conséquences au-delà du seul droit de la preuve. Selon nous, attribuer ce statut de demi-preuve influencerait directement le rôle du système d'IA dans la prise de décision. Dans un premier temps, cette force probante atténuée pourrait confiner le résultat à un rôle de preuve supplémentaire facultative. Puisqu'elle ne peut emporter la conviction du juge seule, la preuve par un système d'IA deviendrait un élément de preuve supplémentaire que les parties ajouteraient pour corroborer d'autres preuves au dossier. Dans un second temps, la limitation de la force probante pourrait rendre moins fréquente l'automatisation totale de processus décisionnels dont les conséquences pourraient faire l'objet de poursuites judiciaires. En effet, en cas de contestation de la décision finale, l'utilisateur ne pourrait pas la justifier seulement avec le résultat produit par le système d'IA. Par conséquent, le statut de demi-preuve pourrait inciter les utilisateurs à maintenir un contrôle humain effectif, voire à limiter l'utilisation des systèmes d'IA à la seule aide à la décision. Cette hypothèse prospective présente des bons et des mauvais aspects. D'un

¹⁷⁴⁷ LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022. P

¹⁷⁴⁸ Comme dit précédemment, en cas de concurrence entre deux preuves à la force probante identique, le juge retrouve sa liberté d'appréciation. Il pourra alors prendre en considération, s'il le souhaite, les éléments techniques comme la fiabilité du système d'IA pour forger sa conviction personnelle.

côté, cette décision renforcerait l'effectivité des dispositions imposant un contrôle humain du système d'IA. De l'autre côté, une telle mesure pourrait complexifier les procédures dans lesquelles le système d'IA représente un outil probatoire intéressant. C'est notamment le cas pour les audits de systèmes d'IA particulièrement opaques pour lesquels le recours à un système d'IA plus simple est une solution prometteuse envisagée par les autorités de contrôle¹⁷⁴⁹.

Sous-section 2 : L'imposition du fonctionnement du système d'IA dans l'appréciation de son résultat

Annonce de plan. Au lieu d'imposer une force probante unique pour tout un procédé de preuve, il est également possible d'imposer certains critères d'appréciation au juge. Ce dernier est alors contraint de prendre en compte certaines propriétés techniques du procédé de preuve dans l'appréciation de l'élément de preuve produit. D'origine législative ou jurisprudentielle, cette imposition de la prise en compte du fonctionnement du procédé de preuve est déjà mise en place pour certaines preuves technologiques en droit pénal (§1). Il est alors possible d'envisager l'application de cette logique aux résultats générés par des systèmes d'IA. Pour les systèmes d'IA à haut risque, une telle hypothèse conduirait alors à la réutilisation des éléments techniques de preuves, constituées lors de la conception du logiciel conformément au règlement sur l'IA, comme critères contraignants d'appréciation du résultat (§2).

§1. La qualité de fonctionnement comme critère de détermination de la force probante en droit pénal

Annonce de plan. Par le passé, le législateur français a déjà fait le choix d'encadrer spécifiquement le fonctionnement d'outils technologiques à finalité probatoire. Le droit pénal contient deux exemples particulièrement éclairants à ce sujet : les cinémomètres (aussi appelés radars routiers) et les éthylomètres. D'une part, ils sont couramment utilisés pour les preuves de contraventions. D'autre part, ils ont fait l'objet de plusieurs décisions de justice, concernant la méthode d'appréciation de ces preuves technologiques par les juges. Ces deux avantages en

¹⁷⁴⁹ « Le projet-pilote REGALIA au service de la régulation des algorithmes | Inria », publié le 8 décembre 2020. « [Algaudit] 1 - Choisir une solution d'audit algorithmique | Linc ».

font des références intéressantes pour fonder notre raisonnement par analogie avec les résultats produits par les systèmes d'IA.

A- L'interprétation des résultats des cinémomètres

L'encadrement du fonctionnement des cinémomètres destinés au contrôle routier. Le 30 janvier 1974, l'arrêté n°74-74 vient réglementer une catégorie particulière d'instrument de mesurage : les cinémomètres de contrôle autoroutier¹⁷⁵⁰. Le législateur décide déjà d'encadrer précisément le fonctionnement de ces outils probatoires. L'arrêté précise le taux d'erreur maximal autorisé pour l'utilisation des cinémomètres :

« Les erreurs maximales tolérées en vérification primitive sont de :

- *3 km/h en plus ou en moins, pour les vitesses inférieures à 100 km/h ;*
- *3 centièmes de la vitesse, en plus ou en moins, pour les vitesses égales ou supérieures à 100 km/h »*¹⁷⁵¹

Le texte dispose également des modalités de tests de ces cinémomètres, à la fois sur la forme¹⁷⁵² (par exemple la fréquence des contrôles est d'un par an) que sur le fond. Sur les tests à réaliser, l'arrêté indique les données à obtenir. Ces-dernières sont toutes relatives à la qualité du fonctionnement du cinémomètre :

« Essais en laboratoire permettant de déterminer :

- *La courbe d'erreur en fonction de la vitesse ;*
- *L'insensibilité aux variations des caractéristiques de l'alimentation électrique ;*
- *L'influence du temps de mise sous tension ;*
- *L'influence de la température et des variations de l'état hygrométrique de l'air ambiant ;*
- *L'exactitude de la valeur de contrôle*

¹⁷⁵⁰ « Décret n° 74-74 du 30 janvier 1974 réglementant la catégorie d'instruments de mesurage », 74-74, 1974.

¹⁷⁵¹ *Ibid.* Article 8.

¹⁷⁵² Par exemple, la fréquence des contrôles est d'un par an et ces contrôles ne peuvent être réalisés que par le constructeur ou des réparateurs agréés. Voir *Ibid.* Articles 9 et 10.

- [...] »¹⁷⁵³

Les caractéristiques techniques relatives au fonctionnement ont été enrichies au fur et à mesure des innovations techniques et des nouvelles utilisations des cinémomètres. Ainsi, l'arrêté du 4 juin 2009 relatif aux cinémomètres de contrôle routier ajoute une marge d'erreur maximale tolérée pour les cinémomètres installés dans une véhicule en mouvement¹⁷⁵⁴. De façon intéressante, cet arrêté crée une distinction entre le taux d'erreur minimal tolérée au moment de la mise (ou remise) en service du cinémomètre et celui pour les cinémomètres en service. Par exemple, un cinémomètre à poste fixe neuf ou venant d'être réparé doit avoir une marge d'erreur maximale de 3 km/h tandis que celui en service qui est contrôlé doit présenter une marge d'erreur inférieure à 5 km/h¹⁷⁵⁵. Au-delà de créer un régime de conformité technique des cinémomètres, ces informations peuvent s'avérer extrêmement utiles pour apprécier les résultats produits par ces appareils.

L'interprétation contrainte des résultats générés par les cinémomètres. La marge tolérée des cinémomètres a très tôt été intégrée dans l'interprétation de leurs résultats. Les dispositions des articles 5 et 6 de l'arrêté du 4 juin 2009 relatives aux erreurs maximales tolérées deviennent alors des « règles de tolérance » qu'il est indispensable d'appliquer pour constater l'infraction. La première apparition de l'imposition de ces règles de tolérance remonte à la circulaire du 30 novembre 1992 relatives aux nouvelles dispositions en matière de circulation et de sécurité routières¹⁷⁵⁶. Produite à la suite de la loi du 10 juillet 1989 relative à diverses dispositions en matière de sécurité routière et en matière de contraventions¹⁷⁵⁷, la circulaire précise la méthode de constatation des infractions d'excès de vitesse :

« en ce qui concerne la constatation des infractions relatives à la vitesse au moyen de cinémomètres, les réflexions engagées dans le cadre des discussions sur le permis à points ont conduit, en accord avec le Garde des Sceaux, à prendre en compte la marge d'erreur des appareils dès la constatation des infractions d'excès de vitesse. Dans ces conditions, à côté de la vitesse lue sur l'appareil, la vitesse retenue du véhicule après déduction

¹⁷⁵³ *Ibid.* Article 7.2.

¹⁷⁵⁴ « Les erreurs maximales tolérées applicables aux instruments neufs ou réparés sont les suivantes : [...] pour les cinémomètres installés dans un véhicule en mouvement : - plus ou moins 7 km/h, pour les vitesses inférieures à 100 km/h ; - plus ou moins 7 % de la vitesse, pour les vitesses égales ou supérieures à 100 km/h. » « Arrêté du 4 juin 2009 relatif aux cinémomètres de contrôle routier », 2009. Article 5.

¹⁷⁵⁵ Arrêté du 4 juin 2009 relatif aux cinémomètres de contrôle routier, *supra* note 44. Articles 5 et 6.

¹⁷⁵⁶ « Circulaire relative aux nouvelles dispositions en matière de circulation et de sécurité routières applicables à compter du 1er décembre 1992 par les services de police », 1992.

¹⁷⁵⁷ « Loi n° 89-469 du 10 juillet 1989 relative à diverses dispositions en matière de sécurité routière et en matière de contraventions (1) », 1989.

des tolérances prévues par l'arrêté du 7 janvier 1991, devra figurer sur les procès-verbaux. Seule cette dernière vitesse servira de fondement à l'exercice de l'action publique. »¹⁷⁵⁸

La déduction de la marge d'erreur maximale tolérée est donc une étape obligatoire par les agents des forces de l'ordre. La notion de « vitesse retenue » est la conséquence directe de la prise en compte de cette caractéristique technique dans l'appréciation de la preuve d'un excès de vitesse. La preuve appréciée par le juge est donc un résultat transformé du cinémomètre. Cette obligation de prendre en compte la marge d'erreur est contrôlée par les magistrats qui s'assure que celle-ci est respectée conformément à la loi¹⁷⁵⁹. En cas de manquement à cette obligation, le juge ne peut pas se fonder sur la vitesse lue mais seulement sur la vitesse retenue. Par conséquent, le juge vient interpréter une preuve déjà modifiée au regard de la fiabilité du mode de preuve.

B- L'interprétation des résultats produits par les éthylomètres

L'encadrement du fonctionnement des éthylomètres. Un éthylomètre est défini comme un instrument de mesure de la concentration d'alcool dans l'air expiré¹⁷⁶⁰. De la même manière que les cinémomètres de contrôles routiers, les performances du fonctionnement des éthylomètres ont très tôt été encadrées par des arrêtés ministériels. Pour pouvoir être mis (ou remis) en service, les éthylomètres devaient respecter les spécifications techniques de la norme privée NF X 20-701 produite par l'AFNOR (Association française de normalisation)¹⁷⁶¹. Le renvoi à ce standard a ensuite été retiré en 2003 lors de l'arrêté du 8 juillet relatif au contrôle des éthylomètres¹⁷⁶². Le texte renvoie désormais à la recommandation internationale R126 qui énonce notamment la manière de calculer les erreurs maximales tolérées (EMT) pour les éthylomètres. Dans le respect des indications de la recommandation internationale, l'arrêté fixe ses propres niveaux d'erreurs maximales tolérés :

¹⁷⁵⁸ « Circulaire relative aux nouvelles dispositions en matière de circulation et de sécurité routières applicables à compter du 1er décembre 1992 par les services de police », 1992.

¹⁷⁵⁹ Cour de Cassation, Chambre criminelle, 24 janvier 1996, n° 95-83.574.

¹⁷⁶⁰ « Vocabulaire des transports et de l'urbanisme (liste de termes, expressions et définitions adoptés) », 2017.

¹⁷⁶¹ « Les instruments neufs ou réparés doivent satisfaire aux spécifications de la norme NF X 20-701 : *Ethylomètres*. » - « Arrêté du 31 décembre 1985 relatif à la construction, la vérification et l'utilisation des éthylomètres. », 1985. Article 2.

¹⁷⁶² « Arrêté du 8 juillet 2003 relatif au contrôle des éthylomètres. », 2003.

« Les erreurs maximales tolérées, en plus ou en moins, applicables lors de la vérification périodique ou de tout contrôle en service sont :

- 0,032 mg/l pour les concentrations en alcool dans l'air inférieures à 0,400 mg/l ;
- 8 % de la valeur mesurée pour les concentrations égales ou supérieures à 0,400 mg/l et inférieures ou égales à 2,000 mg/l ;
- 30 % de la valeur mesurée pour les concentrations supérieures à 2,000 mg/l. »¹⁷⁶³

De nombreuses autres informations doivent être fournies par le concepteur. Cependant, ces niveaux d'EMT représentent le niveau d'exactitude minimal afin qu'un éthylomètre soit utilisé. Cette information s'avère particulièrement intéressante pour l'appréciation de leurs résultats par les juges.

L'interprétation contrainte des résultats générés par les éthylomètres. Au premier abord, l'appréciation des résultats d'éthylomètre semble être une tâche évidente à réaliser pour les juges. Le résultat (ou les résultats si le conducteur demande un second test) obtenu se situe entre 0,25mg et 0,40 mg d'alcool par litre d'air expiré, le conducteur est contraventionné sur le fondement de l'article R. 234-1-I.2° du code de la route¹⁷⁶⁴. En revanche, si le résultat de l'éthylomètre est égal ou supérieur à 0,40 mg d'alcool par litre d'air expiré, alors il s'agit d'un délit puni de deux ans d'emprisonnement et de 4 500€ d'amende¹⁷⁶⁵.

Pour autant, certains juges choisissaient de prendre en compte les EMT (le taux d'Erreur Maximale Tolérée) dans l'appréciation du résultat de l'éthylomètre. Plus précisément, ils retiraient le taux d'erreur maximal inscrit dans l'arrêté du 8 juillet 2003 au résultat produit. Prenons par exemple le cas d'un conducteur contrôlé à un taux de 0,41 mg d'alcool par litre d'air. Le résultat permettait alors de qualifier son comportement de délictuel. Certains juges appliquaient alors les 8% de taux d'erreur en faveur du conducteur. Ainsi, le résultat pris en compte par le juge était alors de 0,38¹⁷⁶⁶. Ce choix d'appliquer l'EMT inscrit dans la loi change

¹⁷⁶³ *Ibid.* Article 15.

¹⁷⁶⁴ « Même en l'absence de tout signe d'ivresse manifeste, est puni de l'amende prévue pour les contraventions de la quatrième classe le fait de conduire un véhicule sous l'empire d'un état alcoolique caractérisé par : [...] 2° Une concentration d'alcool dans le sang égale ou supérieure à 0,50 gramme par litre ou par une concentration d'alcool dans l'air expiré égale ou supérieure à 0,25 milligramme par litre et inférieure aux seuils fixés à l'article L. 234-1 [0.4], chez les autres conducteurs. » - « Décret n° 2019-871 du 21 août 2019 relatif au droit de conduire limité aux seuls véhicules équipés d'un éthylotest anti-démarrage », 2019-871, 2019. Article 1.

¹⁷⁶⁵ « Loi n° 2003-495 du 12 juin 2003 renforçant la lutte contre la violence routière (1). », 2003.Article 11(V) codifié à l'article L. 234-1 du code de la route.

¹⁷⁶⁶ Ce résultat a été obtenu suite au calcul suivant : 1) 0.41 x 0.08 = 0.033 2) 0.41 – 0.033 = 0.38.

radicalement les conséquences pour le conducteur qui n'encourait désormais qu'une simple contravention de 4^{ème} classe.

Jusqu'en 2019, l'appréciation du résultat d'éthylomètre au regard des EMT prévus par la loi était laissé à la discrétion des juges. La chambre criminelle de la Cour de cassation expliquait alors que « *l'interprétation des mesures du taux d'alcoolémie effectuées au moyen d'un éthylomètre constitue pour le juge une faculté et non une obligation* »¹⁷⁶⁷. La diversité d'appréciation entre les juges du fond, que cette liberté entraînait, portait cependant atteinte au principe posé par l'article préliminaire du code de procédure pénale : « *Les personnes se trouvant dans des conditions semblables et poursuivies pour les mêmes infractions doivent être jugées selon les mêmes règles.* »¹⁷⁶⁸. Par conséquent, la chambre criminelle a changé sa position en imposant la prise en compte favorable des EMT sur le résultat obtenu par l'éthylomètre dans un arrêt du 26 mars 2019 :

*« Attendu qu'il se déduit en conséquence de l'article 15 de l'arrêté du 8 juillet 2003 précité que le juge, lorsqu'il est saisi d'une infraction pour conduite sous l'empire d'un état alcoolique, doit vérifier que, dans le procès-verbal qui fonde la poursuite, il a été tenu compte, pour interpréter la mesure du taux d'alcool effectuée au moyen d'un éthylomètre, des marges d'erreur maximales prévues par ce texte ; »*¹⁷⁶⁹

Les juges sont donc contraints de prendre en compte les EMT et de les déduire du résultat produit par l'éthylomètre, une jurisprudence constante depuis lors¹⁷⁷⁰. En d'autres termes, le juge judiciaire doit prendre en compte les caractéristiques techniques exigées par le législateur pour apprécier les preuves par éthylomètres. De plus, cette jurisprudence rejoint la position adoptée par les juridictions administratives depuis 2018. En effet, dans un arrêt du 14 février 2018, le Conseil d'Etat indique clairement l'obligation pour le représentant de l'état de prendre en compte la marge d'erreur maximale tolérée inscrite dans l'arrêté du 8 juillet 2003 :

« il appartient au représentant de l'Etat dans le département, lorsqu'il entend prononcer la suspension de permis de conduire prévue par l'article L. 224-2 du code de la route au titre d'une conduite sous l'empire d'un état alcoolique caractérisé par une concentration d'alcool dans le sang égale ou supérieure à 0,80 gramme par litre ou par une concentration d'alcool dans l'air expiré égale ou supérieure à 0,40 milligramme par litre, de s'assurer qu'il est établi que ces seuils ont été effectivement dépassés ; qu'il lui appartient, par suite,

¹⁷⁶⁷ Cour de cassation, Chambre criminelle, 24 juin 2009, n° 09-81.119.

¹⁷⁶⁸ Article préliminaire du code de procédure pénale, I alinéa 3.

¹⁷⁶⁹ Conseil d'Etat, 21 mars 2016, n° 390023.

¹⁷⁷⁰ Cour de cassation, Chambre criminelle, 14 mai 2019, n° 18-84.552 Cour de cassation, Chambre criminelle, 14 décembre 2021, n° 20-86.969.

de prendre en compte la marge d'erreur maximale tolérée en vertu de l'arrêt du 8 juillet 2003 précité, sauf si le résultat qui lui a été communiqué mentionne que le chiffre indiqué tient déjà compte de la marge d'erreur, ou fait état d'une marge d'erreur de la technique utilisée inférieure à cette marge maximale ; »¹⁷⁷¹

Cette décision est le point de départ d'une jurisprudence aujourd'hui constante¹⁷⁷². L'obligation d'appliquer la marge d'erreur maximale tolérée s'impose donc, en plus des juges judiciaires, au représentant de l'État qui prend les sanctions associées à l'infraction réalisée par le conducteur. La prise en compte d'une caractéristique technique est donc une contrainte d'interprétation du résultat de l'éthylomètre appliquée à l'ensemble des acteurs impliquées dans la procédure.

§2. La qualité de fonctionnement des systèmes d'IA comme critère d'appréciation de la force probante de leurs résultats

Annonce de plan. Imposer au juge la prise en compte des caractéristiques de fonctionnement des systèmes d'IA est une solution intéressante. Elle permet de garantir que la fiabilité relative du système d'IA en tant qu'outil probatoire est intégrée dans le raisonnement du juge. L'exploration de cette hypothèse pour les preuves par un système d'IA met en lumière l'importance du choix des éléments techniques utilisées pour prouver la conformité réglementaire du procédé de preuve **(A)**. Toutefois, les systèmes d'IA présentent des défis spécifiques qui limitent la portée de l'analogie. Par conséquent, la mise en place d'un système de preuve légale pour les résultats d'un système d'IA nécessiterait certaines adaptations par rapport aux exemples existants **(B)**.

A- L'application aux preuves produites par un système d'IA

Les conditions à l'application du raisonnement probatoire. La logique de raisonnement extraite de l'utilisation des preuves par éthylomètre et par cinémomètre nécessite un prérequis indispensable. Dans les deux situations, la loi dispose d'un niveau de qualité minimale d'exactitude que doit atteindre le procédé probatoire afin d'être mis en service. Ce minimum

¹⁷⁷¹ Conseil d'État, 14 février 2018, n° 407914.

¹⁷⁷² Cour de cassation, Chambre criminelle, 14 mai 2019, n° 18-84.552 ; Cour de cassation, Chambre criminelle, 14 décembre 2021, n° 20-86.969.

de qualité de fonctionnement est désigné sous le nom de marge d'erreur maximale tolérée (marge d'EMT). La marge d'EMT est un seuil statistique posé comme une des conditions de légalité du recours aux éthylomètres et aux cinémomètres. Cette marge d'EMT remplit deux caractéristiques indispensables pour l'application de la logique de raisonnement en question.

Le besoin de seuils statistiques connus et précis. Premièrement, il est essentiel que le seuil statistique servant de référence pour l'interprétation soit précis et connu. Tout d'abord, si le taux d'erreur d'un système d'IA est ignoré par le juge, il lui sera impossible d'interpréter le résultat du système d'IA au regard de cette information. Cela signifie que l'information relative au taux d'erreur du système d'IA doit systématiquement être associée au résultat. Il doit s'agir d'un automatisme, voire d'une obligation, comme c'est le cas pour le cinémomètre. Dans le cas des systèmes d'IA à haut risque, la documentation technique et/ou la notice d'utilisation apparaissent comme deux outils pertinents permettant de satisfaire l'exigence de connaissance du seuil statistique. Cependant, ce n'est pas le cas. La logique discutée ici impose d'appliquer le même taux d'EMT à tous les outils probatoires visés par la législation. Pour les cinémomètres et les éthylomètres, les règles d'interprétation sont identiques quel que soit le taux d'erreur réel de l'appareil qui a produit la preuve. Le critère d'appréciation est celui inscrit dans la loi, pas celui affiché par l'appareil en question. Cela permet d'appliquer une règle uniforme.

Ensuite, l'information utilisée comme référentiel doit être précise sans pour autant devoir être unique. Que ce soit pour les éthylomètres ou les cinémomètres, la marge d'EMT prend la forme d'une valeur chiffrée. Les conditions d'utilisation peuvent amener la marge d'EMT à varier mais sa valeur n'est jamais incertaine pour autant. Pour l'éthylomètre, la quantité d'alcool mesurée fait varier la fiabilité du résultat produit¹⁷⁷³. Pour le cinémomètre, c'est la mobilité de l'appareil contrôleur et la vitesse du véhicule contrôlé qui causent ce problème¹⁷⁷⁴. Ce besoin de précision du seuil statistique de conformité n'est aujourd'hui pas satisfait par le règlement européen sur l'IA. Les exigences de fonctionnement désignent un seuil de qualité uniquement au travers de notions floues¹⁷⁷⁵. Le niveau de risque résiduel global du système d'IA mis sur le

¹⁷⁷³ « Les erreurs maximales tolérées, en plus ou en moins, applicables lors de la vérification périodique ou de tout contrôle en service sont : - 0,032 mg/l pour les concentrations en alcool dans l'air inférieures à 0,400 mg/l ; - 8 % de la valeur mesurée pour les concentrations égales ou supérieures à 0,400 mg/l et inférieures ou égales à 2,000 mg/l ; - 30 % de la valeur mesurée pour les concentrations supérieures à 2,000 mg/l. » - « Arrêté du 8 juillet 2003 relatif au contrôle des éthylomètres. », 2003. Article 15.

¹⁷⁷⁴ « Arrêté du 4 juin 2009 relatif aux cinémomètres de contrôle routier ». Article 6.

¹⁷⁷⁵ Aussi appelées notions cadre ou notions à contenu variable, il s'agit d'une « critère vague mais chargé d'évocation dont il appartient au juge, sur la force de l'idée directrice qui s'en dégage, de déterminer le contenu variable évolutif au gré des espèces et au fil du temps » - CORNU Gérard, *Vocabulaire juridique*, 2022. Pour plus d'informations sur cette notion, voir notamment : CASTETS-RENARD Céline, *Notions à contenu variable et droit d'auteur*, L'Harmattan, 2004.

marché doit être « *acceptable* »¹⁷⁷⁶. L'exactitude, la robustesse et la cybersécurité doivent quant à elles « *atteindre un niveau approprié* »¹⁷⁷⁷. Cette lacune est volontaire de la part du législateur européen¹⁷⁷⁸ et les spécifications techniques de ces exigences seront probablement contenues dans les normes techniques à venir. En effet, la détermination des seuils statistiques de conformité est dépendante de la finalité de l'outil. Pour les cinémomètres et les éthylomètres, la finalité est celle du contrôle routier. En ce qui concerne le règlement européen sur l'IA, l'approche transversale choisie ne permet pas au texte d'entrer dans ce niveau de détails.

Le besoin d'un seuil statistique uniforme entre les produits. La seconde condition pour imposer la prise en compte de seuils statistiques d'un outil dans l'appréciation de ses résultats est l'uniformité desdits seuils. L'uniformité des seuils désigne ici le fait que la ou les indicateurs de calcul du seuils (ou des seuils) sont choisis et fixés pour tous les fournisseurs de l'appareil. Pour les cinémomètres, la vitesse est mesurée en km/h quel que soit l'appareil utilisé pour contrôler le véhicule. La métrique « km/h » de la marge d'EMT est utilisée uniformément pour tous les appareils et dans toutes les situations. De la même manière, la métrique utilisée pour les éthylomètres est le « milligramme par litre d'air expiré ». En revanche, le règlement sur l'IA ne satisfait pas lesdites conditions et ce sont les normes techniques qui apporteront probablement les premiers éléments de réponse. Si l'on prend l'exactitude comme caractéristique centrale de l'appréciation d'un résultat de systèmes d'IA, plusieurs métriques sont envisageables. Il est possible d'utiliser le taux de précision général, le taux de faux positifs ou le taux de faux négatifs. La finalité du système d'IA est encore une fois un élément déterminant dans ce choix. Pour la détection de cancers sur une radiographie ou la reconnaissance faciale utilisée pour retrouver un suspect, la métrique importante peut différer. Quelle que soit la finalité, cette méthode d'appréciation de la preuve par un système d'IA nécessitera que la métrique soit uniforme pour tous les systèmes d'IA poursuivant la même finalité.

¹⁷⁷⁶ « Règlement sur l'intelligence artificielle », 2024. Article 8.4.

¹⁷⁷⁷ *Ibid.* Article 15.1.

¹⁷⁷⁸ Sur la stratégie législative adoptée par le législateur européen pour le règlement européen sur l'IA, voir Section II : Une conformité juridique ex ante pour les systèmes d'IA à haut risque.

B- Les spécificités des preuves produites par un système d'IA.

L'existence de biais algorithmiques. La mise en place d'un système de preuve légale pour l'appréciation des résultats de systèmes d'IA nécessiterait certaines adaptations par rapport aux règles appliquées aux éthylomètres et aux cinémomètres. Tout d'abord, ces derniers sont exempts de biais discriminatoires. Un cinémomètre mesure la vitesse d'un véhicule avec une précision identique quel que soit le modèle du véhicule. De façon identique, les éthylomètres mesurent le taux d'alcoolémie de façon identique quel que soit l'individu contrôlé. En revanche, les biais algorithmiques présents dans les technologies d'IA affectent l'exactitude des systèmes lors de leur fonctionnement. Dès lors, le critère d'uniformité du seuil statistique à prendre en compte devient plus complexe à satisfaire. Prenons par exemple le cas d'un résultat produit par un système d'IA de reconnaissance faciale. Le visage analysé correspond à 85% à celui de la personne poursuivie. Ensuite, disons que le seuil légal d'EMT pour ce type de systèmes d'IA serait établi à 10%. Si l'on applique strictement la logique issue des éthylomètres, le résultat à apprécier comme élément de preuve devrait avoir une valeur de 75%. Cependant, il s'agit de l'application du niveau moyen d'erreur du système d'IA. Or, ce taux d'erreur peut varier grandement selon le sexe et la couleur de peau de l'individu. Ainsi cette valeur pourrait être de 5% pour les hommes blancs mais 35% pour les femmes noires. Par conséquent, l'application du taux d'erreur moyen devient contestable. D'un côté, il avantage le cas des hommes blancs qui bénéficieront d'une tolérance d'erreur plus élevée que la réalité (le résultat retenu et apprécié étant de 75% au lieu de 80%). De l'autre côté, cela désavantage toutes les populations sur lesquelles le taux moyen d'erreur est supérieur à 10%. Ici, l'appréciation de du résultat pour une femme noire devrait conduire à retenir une valeur de 50% et non de 75%. Il s'agit ici d'une différence qui pourrait avoir un effet considérable dans l'interprétation, voire la qualification juridique, réalisée par le juge.

La présence de biais algorithmiques ne rend pas pour autant impossible l'établissement d'un seuil précis et uniforme à prendre en compte lors de l'appréciation de la preuve par un système d'IA. Cependant, elle exige une grande rigueur de la part du législateur s'il souhaite emprunter cette voie. D'une part, il faudra définir clairement le ou les indicateurs à utiliser pour évaluer l'exactitude d'un système d'IA afin que tous les fournisseurs emploient les mêmes. D'autre part, il faudra également définir clairement les indicateurs des biais qui influenceront ensuite ceux de l'exactitude. Ces choix devront être réalisés de façon générale pour tous les systèmes d'IA partageant la même finalité, de préférence par un organisme central ou par le législateur.

Une solution envisageable serait d'appliquer par défaut le taux d'EMT le plus faible, toute population confondue. En d'autres termes, le taux d'erreur toléré posé par la loi n'est plus l'erreur moyenne du système d'IA mais le taux d'exactitude le plus bas enregistré. Par exemple, la loi pourrait exiger que le taux d'erreur, toute population confondue, ne soit pas supérieur à 15%. Si le législateur ne s'empare pas de la question, les juges pourraient décider, à l'image des éthylotests, de prendre en compte la marge d'erreur la plus importante inscrite dans la documentation technique ou la notice d'utilisation communiquée par le fournisseur. C'est ce seuil qui serait retiré du résultat observé pour obtenir le résultat retenu et apprécié par le juge. De cette manière, le mécanisme d'appréciation de la preuve pourrait exercer une influence considérable sur les exigences d'exactitude et de biais au moment de la conception des systèmes d'IA à finalité probatoire.

Les modalités de prise en compte des caractéristiques techniques. La conséquence de la marge d'erreur sur l'appréciation du juge est également une limite du raisonnement par analogie. Nos deux exemples sont issus du code de la route qui répond à une logique de droit pénal. Comme le veut l'adage « *in dubio pro reo* », à savoir « le doute profite à l'accusé », le taux d'erreur d'un mode de preuve doit logiquement être retiré au résultat produit. Cependant, cette logique n'a pas vocation à s'appliquer systématiquement pour des litiges en droit des assurances, en droit administratif, etc. La manière d'appliquer la marge d'erreur revient à discuter de la charge du risque lié à cette erreur. La manière de calculer le résultat retenu revient à favoriser une des parties entre la personne affectée et le déployeur du système d'IA. Pour reprendre l'exemple du cinémomètre, la marge d'erreur prévue de 5 km/h pourrait être appliquée de façon défavorable au conducteur. Ainsi, un conducteur qui serait détecté comme roulant à 48 km/h se verrait appliquer les 5 km/h de marge d'erreur en sa défaveur et le résultat retenu, preuve de l'infraction, serait de 53 km/h. Cette décision signifie que les résultats sont toujours perçus comme de potentiels faux positifs mais jamais comme des faux négatifs¹⁷⁷⁹. Si ce n'est pas le cas pour nos exemples en droit pénal où le choix est évident, ce ne serait probablement pas le cas pour les systèmes d'IA qui seraient appliqués à d'autres finalités que celles prévues dans les activités répressives. Par conséquent, ce principe d'application de ces règles d'appréciation du résultat produit par un système d'IA ne pourra pas être transposé de

¹⁷⁷⁹ À nouveau le taux de faux positifs et de faux négatifs peut différer dans les systèmes d'IA. Par conséquent, il peut être intéressant d'inscrire ces taux directement dans la loi plutôt que de faire référence à l'erreur moyenne tolérée.

façon générale à tous les systèmes d'IA. Une réflexion approfondie sur la personne devant supporter la marge d'erreur devra être menée au préalable.

CONCLUSION DE LA SECTION III :

La fiabilité d'un procédé de preuve est un critère sous-jacent dans l'élaboration d'un régime de preuve spéciale. L'élaboration préalable de règles d'appréciation d'un mode de preuve peut conduire à la fixation d'une force probante au mode de preuve, dépossédant alors le juge de sa prérogative d'appréciation des éléments de preuve. Dans cette situation, l'articulation entre l'appréciation du résultat d'un système d'IA et les propriétés techniques imposées par le règlement européen sur l'IA serait alors quasiment absente¹⁷⁸⁰. Cependant, l'instauration d'un régime de preuve spéciale pourrait également venir imposer au juge la prise en compte du fonctionnement du système d'IA dans l'interprétation faite du résultat. À ce jour, le critère obligatoire d'interprétation est celui de l'exactitude, prenant généralement la forme d'une marge d'erreur inscrite préalablement dans la loi. La dynamique de cette articulation peut alors varier selon que la loi (ou la jurisprudence) reprend les caractéristiques techniques contenues dans les textes relatifs à la conception des systèmes d'IA. Cette solution conduit alors à une articulation forte entre le règlement européen sur l'IA, plus particulièrement les éléments techniques de preuve, et le nouveau droit spécial de la preuve par système d'IA. Si le législateur (ou la jurisprudence) reprend les caractéristiques techniques imposées par le texte européen, et plus particulièrement les normes techniques associées, alors il apparaîtra clairement une influence des exigences de conception sur l'appréciation de la preuve produite par un système d'IA. Sinon, les valeurs énoncées dans les règles d'appréciation auront probablement une influence sur les exigences de conception des systèmes d'IA destinés au contexte concerné par le régime de preuve spéciale. Dans ces deux cas hypothèses, des articulations importantes entre les éléments de preuves de conception et les règles d'interprétation peuvent être constatées.

¹⁷⁸⁰ Pour rappel, lorsque deux preuves contraires possédant la même force probante sont apportées lors d'un litige, le juge retrouve sa capacité d'appréciation des éléments de preuves.

CONCLUSION DU CHAPITRE 2 :

Les informations relatives au fonctionnement du système d'IA se révèlent être particulièrement utiles lors de l'appréciation du résultat généré par ledit logiciel. Au sein du système de libre appréciation de la preuve, l'utilisation de ces éléments techniques dans le contexte d'une preuve produite par un système d'IA dépend aujourd'hui exclusivement du juge. Ce dernier est libre de décider du besoin d'utiliser les indicateurs de fonctionnement du système d'IA à haut risque en tant que critères d'évaluation de la force probante du résultat. L'articulation peut être directe lorsque le juge emploie lui-même les informations techniques relatives au système d'IA dans son raisonnement. Elle peut aussi être indirecte lorsque le juge fait le choix de recourir à un expert judiciaire afin de l'éclairer sur la fiabilité du système d'IA par le biais d'un rapport d'expertise. Enfin, dans l'hypothèse d'un futur régime de preuve spéciale pour la preuve générée par un système d'IA, le besoin et la manière d'articuler le fonctionnement du système d'IA avec l'appréciation de son résultat peuvent être imposés ou interdits au juge. Dans le premier cas, l'étude du droit positif révèle un usage systématique d'indicateurs de fiabilité du procédé probatoire comme critères d'appréciation des éléments de preuve produits.

Pour les systèmes d'IA à haut risque, lesdits critères d'appréciation du résultat généré seront vraisemblablement les éléments techniques de preuve constitués par le fournisseur lors de la conception du système d'IA et contenus dans la documentation technique, conformément au règlement européen sur l'IA. En effet, la préexistence de ces informations et la confiance générée par la certification européenne offre à ces éléments une crédibilité et une disponibilité plus importante pour le juge. L'hypothèse d'une réutilisation de ces informations comme critères d'évaluation de la force probante du résultat accroît considérablement l'importance de ces éléments dans la réglementation applicable aux systèmes d'IA. Dans le cas contraire, la décision de solliciter d'autres informations relatives à la fiabilité du système d'IA à haut risque que celles validées par le droit européen met en avant l'importance du choix des indicateurs dans le règlement européen sur l'IA. Les critères d'appréciation sélectionnés pourront alors avoir des conséquences sur les exigences de conception de certains systèmes d'IA, mettant en lumière l'importance du choix des indicateurs dans la conception du logiciel. Le choix du juge, de l'expert ou du législateur de réemployer ou non ces mêmes informations dans le processus d'appréciation de l'élément de preuve ne remettent pas en question l'existence d'une articulation mais seulement la forme qu'elle prend.

CONCLUSION DU TITRE 2 :

L'articulation entre les éléments techniques de preuve, constitués lors de la conception d'un système d'IA à haut risque, et le traitement judiciaire du résultat produit par ce même logiciel se révèle très variable. Tout d'abord, la qualité de fonctionnement d'un système d'IA à haut risque n'entraîne que peu de conséquences sur l'admissibilité de son résultat. Il ne s'agit ni d'un critère pertinent pour sa qualification en tant que mode de preuve ou pour l'évaluation de sa licéité. Seul le principe du respect des droits de la défense semble affecté par la qualité de fonctionnement du système d'IA. En revanche, nous constatons qu'il existe des interactions significatives dans le cadre du processus d'appréciation de la preuve, plus précisément lors de l'évaluation de la force probante du résultat produit par le système d'IA. Dans ce contexte, les éléments techniques de preuve utilisés pour certifier le logiciel d'IA à haut risque apparaissent pertinents pour servir de critères d'évaluation de la force probante. Que l'évaluation de la force probante soit réalisée par le juge, l'expert judiciaire ou fixée au préalable par des règles, le constat demeure identique. Toutefois, cette articulation ne peut avoir lieu que si le besoin de prendre en compte la qualité de fonctionnement du procédé probatoire est reconnu, que ce soit par le juge ou par de futures règles établissant un régime de preuve spéciale. Ensuite, les conséquences d'une articulation entre l'évaluation de la force probante du résultat d'un système d'IA et les éléments garantissant sa qualité de fonctionnement sont encore incertaines. D'un côté, il est possible que le processus de certification mis en place par le règlement européen sur l'IA mette en avant des indicateurs de fonctionnement qui seront réutilisés pour apprécier la preuve produite par le système d'IA à haut risque. De l'autre côté, il est aussi envisageable que le juge, l'expert, ou même le législateur, se détachent de ces éléments élaborés au niveau européen pour recourir à des indicateurs alternatifs. Ces deux hypothèses, qui ne sont par ailleurs pas exclusives, mettent en lumière et accroissent encore davantage l'importance juridique de la sélection des éléments techniques de preuves lors du processus de certification. Enfin, les dynamiques et les conséquences de cette articulation représentent définitivement un approfondissement à réaliser dans les années qui suivront l'application du texte européen et l'élaboration des normes techniques associées.

CONCLUSION DE LA PARTIE 2 :

En attendant l'application du règlement européen sur l'IA et des législations associées, l'ampleur de l'articulation des éléments de preuves issus de la certification des systèmes d'IA à haut risque et les textes applicables lors de l'utilisation de ces logiciels demeure incertaine. En revanche, l'existence de cette articulation est désormais démontrée et incontestable. Le phénomène permet ainsi d'affirmer que la place de ces éléments de preuve de certification dépasse le seul périmètre du règlement européen sur l'IA. Les interactions entretenues avec les autres législations offrent au juge l'opportunité d'évaluer la pertinence juridique des indicateurs, seuils probabilistes et autres éléments techniques de preuve ayant servi à certifier le logiciel. Par ailleurs, l'analyse de ce phénomène d'articulation dans le cadre, d'une part, des régimes de responsabilité *ex post* du fait de l'IA et, d'autre part, du traitement judiciaire de la preuve produite par un système d'IA, révèle que celui-ci ne se limite pas aux liens textuels entre les législations. Les possibilités d'interactions, et plus particulièrement de réutilisation de ces preuves, sont multiples et dépendantes de la volonté des acteurs sollicités. Ainsi, la possibilité pour ces acteurs d'accepter ou de refuser la réutilisation de ces éléments de preuves confirme clairement que leur détermination est un travail éminemment juridique. Enfin, la réutilisation effective des preuves de conception apparaît comme un indicateur pertinent de la réussite de l'objectif nécessaire, affiché par l'Union européenne, de cohérence entre le règlement sur l'IA et les législations applicables lors de l'utilisation des systèmes d'IA à haut risque.

CONCLUSION GENERALE

Avec le règlement sur l'IA, le législateur européen souhaite protéger, d'une part, le marché européen et, d'autre part, la sécurité, la santé et les droits fondamentaux des résidents européens. Pour cela, il a choisi de soumettre les acteurs participant à la conception des systèmes d'IA, en particulier les systèmes d'IA à haut risque, à des obligations juridiques. En d'autres termes, l'Union européenne emploie la norme juridique pour garantir que la norme technique soit cohérente avec les principes juridiques posés. Cependant, la présence de nombreuses notions floues, habituelles en droit, au sein du règlement sur l'IA interroge l'effectivité de ce texte pour s'appliquer à des systèmes d'IA intégrant des mesures quantifiées et précises. Le recours à des tests techniques, exigés par le règlement, pour garantir la conformité des systèmes d'IA à haut risque aux normes juridiques conduit à l'apparition d'une marge de manœuvre importante lors de l'application du texte par les acteurs concernés. Le premier apport de la présente thèse est de démontrer que les enjeux et les difficultés relatives à l'application du règlement sur l'IA se cristallisent principalement autour du choix, inévitable et subjectif, des preuves techniques de conception.

Dans ce contexte, l'étude du rôle desdites preuves en droit des systèmes d'IA nous a alors permis de faire émerger plusieurs contributions à la compréhension de ce nouveau droit des systèmes d'IA.

Premièrement, il a été démontré que le choix des preuves de conception n'est pas un acte neutre. À l'image des systèmes d'IA eux-mêmes, les preuves techniques de conception sont le reflet des valeurs de l'acteur en charge de leur élaboration. Lorsqu'elles parviennent à convaincre leur destinataire, ces éléments techniques pénètrent le droit et amènent les valeurs qu'elles portent dans le système juridique. La preuve juridique des exigences du règlement sur l'IA est alors un vecteur de transmission des valeurs entre la norme technique et la norme juridique. Cumulé à la marge de manœuvre dont bénéficie les acteurs pour appliquer le règlement, les preuves techniques de conception dépassent leur rôle traditionnel visant à convaincre de la conformité juridique du système d'IA à haut risque. Ainsi, elles contribuent grandement à l'interprétation des notions floues qui composent les obligations substantielles contenues dans le règlement européen. En d'autres termes, les preuves de conception contribuent à définir la conformité juridique d'un système d'IA presque autant qu'elles en

attestent le respect. Le choix de l'élément de preuve détermine donc en partie l'objet de la preuve.

Par conséquent, le contrôle de la preuve juridique, de son élaboration à l'appréciation de son caractère convaincant, est un enjeu de pouvoir. Au-delà des prérogatives juridiques accordées à chacun de ces acteurs, l'étude a également permis d'identifier les acteurs qui seraient concrètement en capacité de déterminer les preuves juridiques considérées comme pertinentes, à savoir celles permettant de convaincre de la conformité du système d'IA à une exigence donnée du règlement sur l'IA. En raison de leur aspect technique, le pouvoir de détermination des preuves techniques de conception pertinentes est confié principalement à des acteurs privés dépourvus de légitimité démocratique et de garanties d'indépendance, que sont les organismes de normalisation et les fournisseurs des systèmes d'IA. À l'inverse, les institutions européennes, les autorités administratives et les juges risquent de jouer un rôle secondaire sur le sujet de la preuve.

La sélection de preuves techniques de conception cohérentes avec les exigences du règlement sur l'IA est en outre un enjeu crucial. Un contrôle insuffisant sur le choix desdites preuves pourra exposer à une perte de cohérence des notions juridiques, une diminution de la sécurité juridique des acteurs du secteur ainsi qu'une réduction de l'effectivité du texte européen. Par conséquent, nous avons recommandé le renforcement du contrôle concret des preuves techniques de conception utilisées pour certifier les systèmes d'IA à haut risque, notamment en systématisant le recours à une certification par un tiers.

Secondement, notre étude a mis en lumière la possibilité pour les preuves de conception d'entraîner des conséquences au-delà de leur périmètre d'application initial, à savoir le règlement sur l'IA. Tout d'abord, il a été analysé que la réglementation relative à la conception des systèmes interagira, lors de sa mise en application, avec les textes relatifs à l'utilisation des systèmes d'IA. Ainsi, l'étude de la preuve a démontré la possibilité de nombreuses relations entre, d'un côté, le règlement européen sur l'IA et, de l'autre côté, le droit de la responsabilité civile et le droit de la preuve. La possibilité de réutiliser des preuves de conception dans le contexte de l'utilisation d'un système d'IA représente un aspect majeur de cette articulation. En d'autres termes, les preuves de conception produisent des effets sur le reste du droit des systèmes d'IA qui dépassent largement les seuls renvois textuels présents dans chacune des réglementations. Elles sont donc des vecteurs d'articulation entre les textes.

De plus, la présente thèse a révélé que cette articulation par la preuve est une dynamique à double sens. Ainsi, nous avons décrit avec précision comment le règlement européen sur l'IA produira des effets sur les réglementations relatives à l'utilisation des systèmes d'IA et réciproquement. Par l'étude des propositions de directives européennes sur la responsabilité en matière d'IA et sur la responsabilité du fait des produits défectueux, il a été démontré que la réutilisation des preuves techniques de conception (produites par le fournisseur du système d'IA à haut risque) dans le contexte d'une action en réparation est un phénomène encouragé et récompensé par le législateur européen. En dépit de ce souhait législatif, nous avons recommandé aux acteurs concernés (juges, experts, défenseurs, etc.) d'évaluer systématiquement la pertinence juridique des preuves techniques de conception communiquées par le fournisseur et de rechercher, le cas échéant, des preuves techniques alternatives.

La découverte de ce phénomène d'articulation par la preuve entre la réglementation relative à la conception des systèmes d'IA et celle relative à leur utilisation révèle deux conséquences majeures. Dans un premier temps, les preuves techniques de conception jouent un rôle majeur non seulement dans le règlement européen sur l'IA mais aussi dans l'encadrement des systèmes d'IA de façon générale. En effet, l'étude a mis en lumière que le choix desdites preuves pourra réduire ou renforcer considérablement la protection des droits des personnes affectées par un système d'IA à haut risque ainsi que leur capacité à obtenir réparation en cas de préjudice. De la même manière, lorsque qu'un résultat généré par un système d'IA est utilisée en tant que preuve judiciaire, les preuves techniques de conception peuvent être réutilisées en tant que preuve externe. De ce fait, leur réutilisation pourra affecter grandement l'appréciation de la preuve générée par un système d'IA et, dans une moindre mesure, son admissibilité en justice.

Dans un second temps, l'étude a mis en lumière les acteurs en capacité de déterminer les preuves de conception pertinentes, en particulier le juge. Lors du traitement judiciaire d'une preuve par système d'IA ou lors d'un litige en responsabilité du fait d'un système d'IA, le juge pourra s'emparer du sujet des preuves techniques de conception et apprécier la pertinence de leur réutilisation dans de tels contextes. Ainsi, le juge retrouvera une capacité concrète d'appréciation des preuves de conception, et donc indirectement une capacité d'évaluer l'interprétation des exigences du règlement sur l'IA réalisée lors de la phase de conception.

Au départ de la présente thèse, nous nous questionnions sur le rôle joué par la preuve dans la réglementation sur les systèmes d'IA et les conséquences de ce rôle. À l'issue de notre

étude, il a été clairement démontré que les preuves exigées par le règlement européen sur l'IA, plus précisément les preuves techniques de conception, ne se limitent pas au rôle traditionnel de la preuve, à savoir emporter la conviction de leur destinataire. L'étude a révélé deux autres rôles majeurs de la preuve issue du règlement sur l'IA. Tout d'abord, la preuve est, dans le cadre du droit des systèmes d'IA, un vecteur de valeurs de la norme informatique vers la norme juridique. Ensuite, la preuve est un aspect décisif de l'articulation entre les législations sur la conception des systèmes d'IA et celles sur leur utilisation. Du fait de ce double rôle, la preuve juridique exigée dans le règlement européen sur l'IA occupe une place centrale pour l'ensemble du droit des systèmes d'IA. Par conséquent, la sélection des preuves de conception considérées comme pertinentes est une tâche encore plus cruciale et déterminante que ne pouvait révéler la seule analyse du règlement sur l'IA. Les acteurs en charge de cette tâche ont donc un pouvoir considérable dans la définition concrète de la future conformité européenne des systèmes d'IA.

Au jour de l'écriture de ces lignes, le règlement européen sur l'IA n'est pas encore entré en application. Si l'étude de la preuve offre un prisme d'analyse inédit sur le droit des systèmes d'IA, ce choix présente une limite inévitable. L'analyse prospective de la place de la preuve juridique au sein du droit des systèmes d'IA devra donc être complétée à l'avenir par des recherches sur la place concrète occupée par la preuve dans ce contexte une fois que les dispositions traitées dans l'étude seront entrées en application. Du fait de l'horizontalité de la présente étude et de la multiplicité des branches du droit étudiées, les prochains travaux de recherche seront probablement nombreux. De l'étude des normes européennes harmonisées à celles des décisions de justice impliquant un système d'IA en passant par les décisions rendues par les autorités de surveillance du marché, les objets d'études ne manqueront pas. En raison de la diversité des secteurs d'applications de l'IA, ces travaux devront intégrer des collaborations interdisciplinaires aussi bien entre différentes branches du droit qu'entre les sciences juridiques et d'autres disciplines scientifiques. De nombreuses recherches devront être menées afin d'analyser la place effective de la preuve dans le droit des systèmes d'IA. En particulier, l'étude des procédures de contrôle qui seront élaborées par les autorités de surveillance de marché ou, en interne, par les fournisseurs de systèmes d'IA à haut risque, est particulièrement prometteuse. Dans une perspective différente, l'étude de l'importance accordée par les différents acteurs à la preuve est un aspect qui n'a pas été abordé ici. La réalisation de recherche quantitatives ou qualitatives, notamment par le biais d'entretiens, impliquant des juges, des experts judiciaires, des professionnels de l'IA ou encore des membres

d'autorité de contrôle nous apparaît indispensable pour traiter le sujet le plus largement possible.

Au sein de la présente thèse, nous nous sommes concentrés sur la preuve au sein du droit des systèmes d'IA mais cette étude permet d'éclairer plus globalement les interactions entre le droit et la technique. De façon générale, la recherche en droit des systèmes d'IA offre une perspective nouvelle sur les défis de l'encadrement des outils des technologies numériques. Elle interroge la capacité du droit à protéger nos droits et libertés lors de la conception et de l'utilisation d'outils technologiques et participe à concevoir l'étendue des conséquences de telles réglementations sur notre système juridique. Alors que le règlement européen 2023/1230 sur les machines¹⁷⁸¹ doit entrer en vigueur en 2027 et que l'Union européenne travaille actuellement sur l'élaboration d'un cadre sur l'identité numérique¹⁷⁸², une bonne compréhension et maîtrise du droit systèmes d'IA est désormais indispensable pour espérer répondre aux enjeux de ces futurs textes mêlant également droit et technologie.

¹⁷⁸¹ « Règlement (UE) 2023/1230 du Parlement européen et du Conseil du 14 juin 2023 sur les machines, abrogeant la directive 2006/42/CE du Parlement européen et du Conseil et la directive 73/361/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2023. Article 54.

¹⁷⁸² COMMISSION EUROPÉENNE, *Proposition de règlement du Parlement européen et du Conseil modifiant le règlement (UE) n° 910/2014 en ce qui concerne l'établissement d'un cadre européen relatif à une identité numérique*, 3 juin 2021. EYNARD Jessica, *L'identité numérique*, Larcier, 2020.

BIBLIOGRAPHIE

Dictionnaires et encyclopédies

CABRILLAC Rémy, *Dictionnaire du vocabulaire juridique* 2022, Edition 2022, Paris, LEXISNEXIS, 2021, 568 p.

CORNU Gérard, *Vocabulaire juridique*, Presses Universitaires de France-P.U.F., 2022, 1106 p.

DEBARD Thierry et GUINCHARD Serge, *Lexique des termes juridiques 2022-2023*, Dalloz, 2022.

Ouvrages généraux, manuels et traités

Ouvrages généraux

ALBIGÈS Christophe, *Introduction au droit*, 9^e éd., Bruylant, 2024, 338 p.

AUBERT Jean-Luc et SAVAUX Éric, *Introduction au droit et thèmes fondamentaux du droit civil*, Sirey, 2023, 520 p.

ARISTOTE, *Éthique de Nicomaque*, Paris, FLAMMARION, 2004, 574 p.

BARBIER Hugo et GHESTIN Jacques, *Introduction générale*, vol. 2, 5^e éd., LGDJ, 642 p.

BERGEL Jean-Louis, *Théorie Générale du Droit*, 5e éd., Dalloz, 2012, 400 p.

BACACHE-GIBEILI Mireille, *Droit civil - Les obligations, la responsabilité civile extracontractuelle*, 4^{ème}, Economica, 2021, 1176 p.

BRUN Philippe, *Responsabilité civile extracontractuelle*, 6e édition, Paris, LEXISNEXIS, 2023, 733 p.

CABRILLAC Rémy, *Introduction générale au droit*, 15^e éd., Dalloz, 2023, 312 p.

CADIET Loic et JEULAND Emmanuel, *Droit judiciaire privé*, 11^{ème}, LexisNexis, 2020.

CARBONNIER Jean, *Droit civil. Introduction*, 27e éd., Presses Universitaires de France - PUF, 2002.

CARBONNIER Jean, *Essais sur les lois*, 2^{ème} édition, Dalloz, 1995.

CARVAL Suzanne, JOURDAIN Patrice et VINEY Genevieve, *Les régimes spéciaux et l'assurance de responsabilité*, 4e édition, LGDJ, 2017, 688 p.

CARVAL Suzanne, VINEY Genevieve et JOURDAIN Patrice, *Les conditions de la responsabilité*, 4^{ème} édition, LGDJ, 2013.

CHABAS François, MAZEAUD Henri, MAZEAUD Jean *et al.*, *Traité théorique et pratique de la responsabilité civile délictuelle et contractuelle*, vol. 1, LGDJ, 2013, 936 p.

COLIN Ambroise et CAPITANT Henri, *Cours élémentaire de droit civil français.*, 1930. ↗
<https://gallica.bnf.fr/ark:/12148/bpt6k5435367j>

CORNU Gérard, *Droit civil. Introduction au droit*, LGDJ, 2007, 277 p.

DEUMIER Pascale, *Introduction générale au droit*, 5^e éd., LGDJ, 400 p.

DOMAT Jean et DE BOUCHEVRET Benoît l'Usurier, *Les loix civiles dans leur ordre naturel, le droit public et Legum delectus.*, 1756. ↗

DUMONT Étienne et BENTHAM Jeremy, *Traité des preuves judiciaires*, vol. Tome 1, 1823. ↗

GLASSON Ernest, *Précis théorique et pratique de procédure civile*, vol. Tome 1, 1908. ↗

LARROUMET Christian, *Traité de droit civil*, vol. 1, 6^e éd., Economica, 2013, 438 p.

LE TOURNEAU Philippe, *Droit de la responsabilité et des contrats - Régimes d'indemnisation*, Dalloz.

LECLERC Olivier, VERGES Etienne et VIAL Géraldine, *Droit de la preuve*, Presses Universitaires de France - PUF, 2022.

MALINVAUD Philippe et BALAT Nicolas, *Introduction à l'étude du droit*, 24^e éd., Paris, LexisNexis, 2024, 598 p.

MOLFESSIS Nicolas et TERÉ François, *Introduction générale au droit*, 15^e éd., Dalloz, 2023, 906 p.

MORVAN Patrick et MALAURIE Philippe, *Introduction au droit*, vol. 1, 9^e éd., LGDJ, 2022, 536 p.

PLANIOL Marcel (1853-1931), *Traité élémentaire de droit civil*, vol. Tome 2, 9^{ème}, 1922. ↗

RAWLS John, *Théorie de la justice*, Paris, POINTS, 2009, 672 p.

REMY Joseph, *Oeuvres complètes de J. Domat*, vol. 4, Firmin-Didot père et fils (Paris), 1828, 582 p.

ROCHFELD Judith, *Les grandes notions du droit privé*, 3^e éd., Presses Universitaires de France - PUF, 2022, 714 p.

SAINT-PAU Jean-Christophe, *Droits de la personnalité*, LexisNexis, 2013.

THOMAT-RAYNAUD Anne-Laure et BEIGNIER Bernard, *Introduction au droit*, 7^e éd., LGDJ, 2024, 628 p.

VINEY Genevieve, *Introduction à la responsabilité*, 4^e éd., LGDJ, 2019, 704 p.

VINEY Genevieve, *Les conditions de la responsabilité*, 4^e éd., LGDJ, 2013, 1320 p.

Thèses

BERGEAUD Aurélie, *Le droit à la preuve*, LGDJ, 2010, 588 p.

CASTETS-RENARD Céline, *Notions à contenu variable et droit d'auteur*, L'Harmattan, 2004, 518 p.

HUTTNER Liane, *La décision de l'algorithme*, vol. 235, Dalloz, 2024, 500 p.

MERABET Samir, *Vers un droit de l'intelligence artificielle*, Dalloz, 2020, 558 p.

RENONDIN DE HAUTECLOCQUE Stanislas, *Intelligence artificielle, la recherche d'un régime juridique. Contribution à l'étude de la compatibilité d'une approche par les risques et d'une approche par les droits*, Université Pais Cité, 2024, 755 p.

Ouvrages spécialisés, thèses et mémoires

Ouvrages spécialisés

BRADFORD Anu, *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020, 424 p.

BENSAMOUN Alexandra et LOISEAU Grégoire, *Droit de l'intelligence artificielle*, vol. 15, L.G.D.J., 2019, 444 p.

BRENNER Susan, *Law in an Era of « Smart » Technology*, Oxford University Press, 2007, 200 p.

CALO Ryan, FROOMKIN A. et KERR Ian, *Robot Law*, Edward Elgar Publishing, 2016, 424 p.

CARDON Dominique, *A quoi rêvent les algorithmes : Nos vies à l'heure des big data*, Paris, SEUIL, 2015, 112 p.

CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général, perspectives comparées*, 1re éd., Bruylant, 2023, 996 p.

COLLINGRIDGE David, *The Social Control of Technology*, Frances Pinter, 1980, 208 p.

EUBANKS Virginia, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor*, New York, NY, St Martin's Press, 2018, 272 p.

EYNARD Jessica, *L'identité numérique: Quelle définition pour quelle protection ?*, Larcier, 2020, 212 p.

FRYDMAN Benoit et VAN WAEYENBERGE Arnaud, *Gouverner par les standards et les indicateurs. De Hume aux rankings / Frydman Benoît et van Waeyenbergen Arnaud (dir.)*, Bruylant, 2013.

GORZ Andre, *Stratégie ouvrière et néocapitalisme*, Editions du seuil, 1964.

GRYNBAUM Luc, LE GOFFIC Caroline et PAILLER Ludovic, *Droit des activités numériques*, 2^e éd., Dalloz, 2023, 1144 p.

KEANE Adrian et MCKEOWN Paul, *The Modern Law of Evidence*, 9th edition, Oxford, Oxford University Press, 2012, 760 p.

LE TOURNEAU Philippe, *Contrats du numérique ; informatiques et électroniques*, 12^{ème}, Dalloz, 2022, 808 p.

LESSIG Lawrence, *Code: And Other Laws of Cyberspace, Version 2.0*, New York, Basic Books, 2006, 431 p.

LLORENS-FRAYSSE Françoise, *La présomption de faute dans le contentieux administratif de la responsabilité*, Mazères Jean-Arnaud (dir.), L.G.D.J., 1985, 372 p.

MCCORDUCK Pamela, *Machines Who Think*, 2e édition, Boca Raton London New York, Routledge, 2004, 598 p.

MENECEUR Yannick, *L'intelligence artificielle en procès. Plaidoyer pour une réglementation internationale et européenne*, Bruylant, 2020, 547 p.

MINSKY Marvin et PAPERT Seymour A., *Perceptrons: An Introduction to Computational Geometry*, The MIT Press, 1969, 258 p.

NORVIG Peter et RUSSELL Stuart, *Intelligence artificielle - Une approche moderne*, 4e édition, Paris, PEARSON, 2021, 976 p.

O'NEIL Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Reprint édition, New York, Crown, 2017, 288 p.

PASQUALE Frank, *The Black Box Society – The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015, 320 p.

PICKOVER Clifford A., *La fabuleuse histoire de l'intelligence artificielle : Des automates aux robots humanoïdes*, Dunod, 2021, 304 p.

OGRON Joseph-Adrien, *Code de procédure civile expliqué par ses motifs et par des exemples*, 3^{ème}, 1837, 315 p.

SARANGI Saswat et SHARMA Pankaj, *Artificial Intelligence: Evolution, Ethics and Public Policy*, London, Routledge India, 2018, 164 p.

SCHEPEL Harm, JOERGES Christian, TEUBNER Gunther *et al.*, *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets*, Oxford ; Portland, Or, Hart Publishing, 2005, 504 p.

SIMONDON Gilbert, *Du mode d'existence des objets techniques*, Paris, AUBIER, 2012, 368 p.

SUPIOT Alain, *La Gouvernance par les nombres*, Nantes : Paris, Fayard, 2015, 512 p.

TRUILHÉ-MARENGO Eve, *Preuve scientifique, preuve juridique*, Larcier, 2011, 358 p.

TURNER Jacob, *Robot Rules: Regulating Artificial Intelligence*, 2019, 377 p.

ZUBOFF Shoshana, *L'Âge du capitalisme de surveillance*, Zulma, 2020, 864 p.

Mémoires

PONS Ronan, *Le droit et la discrimination algorithmique*, 2018, 232 p.

Articles

ABBOTT Kenneth et LEVI-FAUR David, « Introducing Regulatory Intermediaries », *The ANNALS of the American Academy of Political and Social Science*, 670, mars 2017, p. 6-13.

ABBOTT Kenneth et LEVI-FAUR David, « Enriching the RIT framework », *The ANNALS of the American Academy of Political and Social Science*, 670, mars 2017, p. 280-288.

ABBOTT Kenneth W., LEVI-FAUR David et SNIDAL Duncan, « Theorizing Regulatory Intermediaries: The RIT Model », *The ANNALS of the American Academy of Political and Social Science*, 670, SAGE Publications Inc, mars 2017, n° 1, p. 14-35.

ACKOFF Russell, « From Data to Wisdom », *Journal of Applied Systems Analysis*, 1989, p. 3-9.

AFFONSO SUZA Carlos, PARRA CANO Flavia et BRANCO Sergio, « La réglementation de l'intelligence artificielle au Brésil : la décortication des propositions législatives actuelles », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle. Entre règles sectorielles et régime général. Perspectives comparées.*, 1re éd., Bruylant, 2023, p. 845-866.

ALHELT K., « The Applicability of the EU Product Liability Directive to Software », *Comparative and International Law Journal of Southern Africa*, 34, 2001, n° 2, p. 188-209.

ALLEN Craig H., « The Seabots are Coming Here: Should they be Treated as 'Vessels'? », *The Journal of Navigation*, 65, Cambridge University Press, octobre 2012, n° 4, p. 749-752.

ANDREOU Athanasios, VENKATADRI Giridhari, GOGA Oana *et al.*, *Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations*, 2018.

BAILLOUX Cassandre, « The Average Consumer in European Consumer Law », *Exeter Law Review*, 44, 2017, p. 158-179.

BALKIN Jack M., « The Path of Robotics Law », *California Law Review*, 6, juin 2015, p. 45-60.

BARBIER Hugo, « L'admissibilité de la preuve illicite », *Revue trimestrielle de droit civil*, juin 2021, n° 2, p. 413-414.

BAROCAS Solon et SELBST Andrew D., « Big Data's Disparate Impact », *California Law Review*, 104, 2016, no 3, p. 671-732.

BARRAUD Boris, « La corégulation d'Internet (ou comment répondre à la plurinormativité par l'internormativité) — Une contribution française », *Les Cahiers de droit*, 59, Faculté de droit de l'Université Laval, 2018, n° 1, p. 85-116.

BATYA Friedman et HELEN Nissenbaum, « Bias in computer systems », *Proceedings of the 1996 Conference on Fairness, Accountability, and Transparency*, 1996.

BENEZECH Ludovic, « L'affermissement de la justiciabilité des actes de droit souple devant le Conseil d'État », *Droit administratif*, LexisNexis, 2020, n° 2, p. 33 », *Droit administratif*, 2020, n° n° 2, p. 33.

BENKLER Yochai, « Don't let industry write the rules for AI », *Nature*, 569, mai 2019, n° 7755, p. 161-161.

BENNETT MOSES Lyria, « How to Think About Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target », *Law, Innovation and Technology*, 5, 2013, n° 1, p. 1-20.

BENSAMOUN Alexandra, « Filtrage des contenus par les plateformes contributives et liberté d'expression », *Légipresse*, septembre 2021, n° 395, p. 399.

BEN-SHAHAR Omri, « The Myth of the 'Opportunity to Read' in Contract Law », 5, *De Gruyter*, mars 2009, n° 1, p. 1-28.

BILGE Sirma, « Théorisations féministes de l'intersectionnalité », *Diogene*, n° 225, Presses Universitaires de France, 2009, n° 1, p. 70-88.

BLANC Alain, « La preuve aux assises: entre formalisme et oralité, la formation de l'intime conviction », *AJ Pénal*, 2005, p. 271-274.

BOINE Claire, « La responsabilité civile liée à l'intelligence artificielle », dans CASTETS-RENNARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général. Perspectives comparées*, Bruylant, 2023, p. 679-708.

BOVENS Mark, « Two Concepts of Accountability : Accountability as a Virtue and as a Mechanism », *West European Politics*, 33, Routledge, septembre 2010, n° 5, p. 946-967.

BUCHANAN Bruce et FEIGENBAUM Edward, « Heuristic Dendral: a Program for Generating Explanatory Hypotheses in Organic Chemistry », *Edinburg University Press*, 1969, p. 209-254.

BUOLAMWINI Joy et GEBRU Timnit, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR, 2018, p. 77-91.

BURRELL Jenna, « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, 3, SAGE Publications Ltd, juin 2016, n° 1, p. 205.

CAILLÉ Catherine, « Responsabilité du fait des produits défectueux », *Répertoire de droit civil*, juin 2018.

CALO Ryan, « Artificial Intelligence Policy: A Primer and Roadmap », *U.C. Davis Law Review*, 51, décembre 2017, n° 2, p. 399-436.

CALO Ryan, « Robotics and the Lessons of Cyberlaw », *California Law Review*, 103, juin 2015, n° 3, p. 513-564

CASAROSA Federica, « Cybersecurity certification of Artificial Intelligence: a missed opportunity to coordinate between the Artificial Intelligence Act and the Cybersecurity Act », *International Cybersecurity Law Review*, 3, juin 2022, n° 1, p. 115-130.

CASTETS-RENARD Céline, « Proposition de règlement sur l'intelligence artificielle (derniers développements) », *Recueil Dalloz*, avril 2023, n° 13, p. 680.

CASTETS-RENARD Céline et BESSE Philippe, « Responsabilité ex ante de l'AI Act : Entre certification et normalisation, à la recherche des droits fondamentaux au pays de la conformité », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général; Perspectives comparées*, Bruylant, 2023, p. 597-629.

CASTETS-RENARD Céline, « Le Livre blanc de la Commission européenne sur l'intelligence artificielle : vers la confiance ? », *Recueil Dalloz Sirey*, 15 avril 2020, p. 837-842.

CASTETS-RENARD Céline, « Comment construire une intelligence artificielle responsable et inclusive ? », *Recueil Dalloz Sirey*, 4, février 2020, p. 225-230.

CASTETS-RENARD Céline, « The Intersection Between AI and IP: Conflict or Complementarity? », *IIC - International Review of Intellectual Property and Competition Law*, 51, février 2020, n° 2, p. 141-143.

CHESTERMAN Simon, « Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity », *The American Journal of Comparative Law*, 69, juin 2021, n° 2, p. 271-294.

CHIANDUSSI G., CODEGONE M., FERRERO S. *et al.*, « Comparison of multi-objective optimization methodologies for engineering applications », *Computers & Mathematics with Applications*, 63, mars 2012, n° 5, p. 912-942.

CHOULDECHOVA Alexandra, « Fair prediction with disparate impact: A study of bias in recidivism prediction instruments », *Big Data*, 5, octobre 2016, n° 2, p. 153-163.

CRAWFORD Kate, « The Hidden Biases in Big Data », *Harvard Business Review*, 1 avril 2013.

CRENSHAW Kimberlé, « Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics », *U. Chi. Legal F.*, 1989, janvier 1989, p. 139.

CUCCURU Pierluigi, « Interest Representation in European Standardisation: The Case of CEN and CENELEC », *Amsterdam Law School Research Paper*, décembre 2019, n° 52.

CUMMINGS M. L., *Automation Bias in Intelligent Time Critical Decision Support Systems*, *AIAA 3rd Intelligent Systems Conference*, AIAA, 2004.

DARBY Michael R. et KARNI Edi, « Free Competition and the Optimal Amount of Fraud », *The Journal of Law and Economics*, 16, The University of Chicago Press, avril 1973, n° 1, p. 67-88.

DARLING Kate, « Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects », dans CALO Ryan, FROOMKIN Michael A. et KERR Ian, *Robot Law*, Rochester, 2016, p. 213-232.

DATTA Amit, TSCHANTZ Michael et DATTA Anupam, « Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination », *Proceedings Priv. Enhancing Tech*, 2015, n° 1.

DE SCHUTTER Olivier, « Prouver la discrimination », *ERA Forum*, 4, 2003, n° 3, p. 45-54.

DE LAAT Paul B., « Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? », *Philosophy & Technology*, 31, décembre 2018, n° 4, p. 525-541.

DELAGE Pierre-Jerôme, « Vers un principe de fiabilité de la preuve pénale ? », *Revue de sciences criminelles et de droit comparé*, décembre 2019, n° 3, p. 653.

DELTORN Jean-Marc, « La protection des données personnelles face aux algorithmes prédictifs | Revue des droits et libertés fondamentaux », *Revue des Droits et Libertés Fondamentales*, 2017, n° 12.

DODGE Jonathan, LIAO Q. Vera, ZHANG Yunfeng *et al.*, Explaining models: an empirical study of how explanations impact fairness judgment, *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Marina del Ray, California, Association for Computing Machinery, 2019, p. 275-285.

DOUVILLE Thibault, « Du renforcement de l'effectivité du RGPD : contrôle juridictionnel plénier de la décision de rejet d'une réclamation prononcée par une autorité de contrôle compétente en matière de protection des données », *RTD Com.*, juin 2024, n° 2, p. 349-351.

DOUVILLE Thibault, « De l'approche extensive de la prise de décision exclusivement automatisée », *Recueil Dalloz*, mai 2024, n° 20, p. 1000

EASTERBROOK Frank, « Cyberspace and the Law of the Horse », *University of Chicago Legal Forum*, 1996, décembre 2015, n° 1.

EBERS Martin, « Standardizing AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act », dans DI MATTEO Larry A., PONCIBO Cristina et CANNARSA Michel, *The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics*, 2022, p. 321-346.

EDWARDS Lilian et VEALE Michael, « Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'? », *IEEE Security & Privacy*, 16, 2018, n° 3, p. 46-54.

EDWARDS Lilian et VEALE Michael, « Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for », *Duke Law Technology Review*, 16,

décembre 2017, n° 18, p. 18-84.

ELIANTONIO Mariolina et CUCCURU Pierluigi, « It is not all about judicial review: Internal appeal proceedings in the European standardisation process », *Le futur du droit administratif/The Future of Administrative Law*, Lexis Nexis, 2019, p. 475-488.

ELIANTONIO Mariolina, « Judicial Control of the EU Harmonized Standards: Entering a Black Hole? », *Legal Issues of Economic Integration*, 44, septembre 2017.

ENGSTROM Nora Freeman, « 3-D Printing and Product Liability: Identifying the Obstacles », *University of Pennsylvania Law Review*, 162, octobre 2013, n° 35, p. 35-41.

FABRE Alexandre, « Contrat de travail à durée indéterminée », *Répertoire du droit du travail*, juin 2023.

FERRAND Frédérique, « Preuve », *Répertoire de procédure civile*, Dalloz, mars 2024.

FERRAND Frédérique, « Le juge et la preuve écrite », *Répertoire de procédure civile*, Dalloz, mars 2023, p. 598-618.

FJELD Jessica, ACHTEN Nele, HILLIGOSS Hannah *et al.*, « Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI », *Berkman Klein Center of Research Publication*, janvier 2020, n° 1.

FLORIDI Luciano, « Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical », *Philosophy & Technology*, 32, juin 2019, n° 2, p. 185-193.

GALLAND Jean-Pierre, « Big Third-Party Certifiers and the Construction of Transnational Regulation », *The ANNALS of the American Academy of Political and Social Science*, 670, SAGE Publications Inc, mars 2017, n° 1, p. 263-279.

GAUTRAIS Vincent, « Preuve des reproductions : vues d'ailleurs ! », *Cahiers Droit, Sciences & Technologies*, PUP, décembre 2014, n° 4, p. 301-316.

GAUTRAIS Vincent et GINGRAS Patrick, « La preuve des documents technologiques », *Les Cahiers de propriété intellectuelle*, 22, Les cahiers de propriété intellectuelle Inc., 2017, n° 22-2.

GRANET-LAMBRECHTS Frédérique, « Preuves en droit de la famille : La preuve des liens de filiation », *AJ Famille*, décembre 2007, n° 12, p. 459-463.

GRAZ Jean-Christophe, « Quand les normes font loi : Topologie intégrée et processus différenciés de la normalisation internationale », *Études internationales*, 35, IQHEI, 2004, n° 2, p. 233-260.

GREEN Ben, « “Good” isn’t good enough », *Proceedings of the AI for Social Good workshop at NeurIPS*, 17, décembre 2019.

GREENE Daniel, HOFFMANN Anna Lauren et STARK Luke, « Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning », *Hawaii International Conference on System Sciences*, 2019.

GROFFE-CHARRIER Julie, « Vers un encadrement légal des deepfakes », *Dalloz actualités*, juillet 2023.

GROZDANOVSKI Ljupcho, « In search of effectiveness and fairness in proving algorithmic discrimination in EU law », *Common Market Law Review*, 58, janvier 2021, n° 1, p. 99-136.

GROZDANOVSKI Ljupcho, « Le probable, le plausible et le vrai. Contribution à la théorie générale de la présomption en droit », *Revue interdisciplinaire d'études juridiques*, 84, Université Saint-Louis - Bruxelles, 2020, n° 1, p. 39-72.

GRYNBAUM Luc, « Preuve », *Répertoire de droit commercial*, Dalloz, octobre 2020.

HACKER Philipp, « Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law », *55 Common Market Law Review* 1143-1186 (2018), avril 2018, p. 35.

HAGENDORFF Thilo, « The Ethics of AI Ethics -- An Evaluation of Guidelines », *Minds and Machines*, 30, mars 2020, n° 1, p. 99-120.

HARRISON Galen, HANSON Julia, JACINTO Christine *et al.*, An empirical study on the perceived fairness of realistic, imperfect machine learning models, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Association for Computing Machinery, 2020, p. 392-402.

HICKOK Merve, « Lessons learned from AI ethics principles for future actions », *AI and Ethics*, 1, février 2021, n° 1, p. 41-47.

HOFMANN Herwig C.H., ROWE Gerard C. et TÜRK Alexander H., « Judicial Supervision », da,s Hofmann Herwig C.H., Rowe Gerard C. et Türk Alexander H. (éd.), *Administrative Law and Policy of the European Union*, Oxford University Press, 2011, p. 797-903.

JOBIN Anna, IENCA Marcello et VAYENA Effy, « The global landscape of AI ethics guidelines », *Nature Machine Intelligence*, 1, Nature Publishing Group, septembre 2019, n° 9, p. 389-399.

JOURDAIN Patrice, « La responsabilité du producteur peut être engagée en cas de faute », *RTD Civ.*, mai 2024, n° 01, p. 130-132.

KLEINBERG Jon, MULLAINATHAN Sendhil et RAGHAVAN Manish, Inherent Trade-Offs in the Fair Determination of Risk Scores, *Proceedings of Innovations in Theoretical Computer Science*, 2017.

KOP Mauritz, « EU Artificial Intelligence Act: The European Approach to AI », *Transatlantic Antitrust and IPR Developments*, septembre 2021.

KRANZBERG Melvin, « Technology and History: “Kranzberg’s Laws” », *Technology and Culture*, 27, [The Johns Hopkins University Press, Society for the History of Technology], 1986, n° 3, p. 544-560.

KROLL Joshua A., HUEY Joanna, BAROCAS Solon *et al.*, « Accountable Algorithms », *University of Pennsylvania Law Review*, 165, 2017, p. 633-706.

KULLMANN Miriam et CEFALIELLO Aude, « The Draft Artificial Intelligence Act (AI Act): Offering False Security to Undermine Fundamental Workers' Rights », *European Labour Law Journal*, 13, 2022, n° 4, p. 542-562.

LARDEUX Gwendoline, « Preuve : modes de preuve », *Répertoire de droit civil*, février 2023.

LARDEUX Gwendoline, « Preuves : règles de preuves », *Répertoire de droit civil*, octobre 2018.

LASSERRE-CAPDEVILLE Jérôme, « Autorité de contrôle prudentiel et de résolution : supervision bancaire - Pouvoirs de l'ACPR », *Répertoire de droit commercial*, juillet 2023.

LAUTRETTE Laurence, « Telex, télécopie, télégramme : valeur juridique et force probante », *Les Petites Affiches*, mai 1996, n° 57, p. 5.

LAVELLE Sylvain, « Politiques des artefacts. Ce que les choses font et ne font pas », *Cités*, 39, Presses Universitaires de France, 2009, n° 3, p. 39-51.

LECLERC Olivier, « La distinction entre la “preuve en droit” et la “preuve en science” est-elle pertinente ? », dans TRUILHÉ-MARENGO Ève, *Preuve scientifique, preuve juridique*, Larcier, 2011, p. 55-77.

LE GUIDE Raymond et CHABOT Gérard, « Filiation : modes judiciaires d'établissement », *Répertoire de droit civil*, janvier 2023.

LELIEUR Juliette, « L'intelligence artificielle, une nouvelle technologie probatoire en émergence », *AJ Pénal*, mars 2022, n° 3, p. 112.

LEONHARD Julie, « La place de l'ADN dans le procès pénal », *Cahiers Droit, Sciences & Technologies*, PUP, octobre 2019, n° 9, p. 45-56.

LEPAGE Agathe, « Droits de la personnalité - Droit de la personnalité et droit à la preuve », *Répertoire de droit civil*, avril 2023.

LESSIG Lawrence, « The Law of the Horse: What Cyberlaw Might Teach », *Harvard Law Review*, 113, The Harvard Law Review Association, 1999, n° 2, p. 501-549.

LESSIG Lawrence, « The New Chicago School », *The Journal of Legal Studies*, 27, [The University of Chicago Press, The University of Chicago Law School], 1998, n° S2, p. 661-691.

LIU Hin-Yan, MAAS Matthijs M., DANAHER John *et al.*, « Artificial Intelligence and Legal Disruption: A New Model for Analysis », *Law, Innovation and Technology*, 12, septembre 2020, n° 2, p. 205-258.

LOCHAK Danièle, « La notion de discrimination », *Confluences Mediterranee*, 48, 2004, n° 1, p. 13-23.

LOCONTO Allison Marie, « Models of Assurance: Diversity and Standardization of Modes of Intermediation », *The Annals of the American Academy of Political and Social Science*, 670, [Sage Publications, Inc., American Academy of Political and Social Science], 2017, p. 112-132.

LOUREIRO Maria et MCCLUSKEY Jill, « Reputation and Production Standards », *Journal of Agricultural and Resource Economics*, 30, février 2005.

LYNDON Mary L., « Tort Law and Technology Response », *Yale Journal on Regulation*, 12, 1995, n° 1, p. 137-176.

MALGIERI Gianclaudio, The concept of fairness in the GDPR: a linguistic and contextual interpretation, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Association for Computing Machinery, 2020, p. 154-166.

MANGEMATIN Céline, « Produits défectueux - Responsabilité du fait des produits défectueux et intelligence artificielle : une proposition presque parfaite », *Responsabilité civile et assurances*, juin 2023, n° 6.

MARON Albert, « Enquête préliminaire, deux noms pour un acte innomé », *Droit pénal*, 69, 2018, n° 4.

MARON Albert, « Mais où sont les neiges d'antan ? », *Droit pénal*, 2007, n° 91.

MATUS Kira J.M. et VEALE Michael, « Certification systems for machine learning: Lessons from sustainability », *Regulation & Governance*, 16, 2022, n° 1, p. 177-196.

MAZEAU Laurène, « Responsabilité. Quel(s) droit(s) pour les systèmes d'intelligence artificielle ? », *Cahiers Droit, Sciences & Technologies*, PUP, mai 2020, n° 10, p. 223-229.

MAZEAU Laurène, « Responsabilité. Les enjeux de la normalisation technique dans le domaine de l'intelligence artificielle », *Cahiers Droit, Sciences & Technologies*, PUP, mars 2019, n° 8, p. 225-231.

MAZEAU Laurène, « L'influence des sciences sur le droit de la responsabilité civile professionnelle : l'exemple de la normalisation », in Larrieu Jacques (éd.), *Qu'en est-il du droit de la recherche ?*, Toulouse, Presses de l'Université Toulouse 1 Capitole, 2018, p. 239-278.

MCCULLOCH Warren S. et PITTS Walter, « A logical calculus of the ideas immanent in nervous activity », *The bulletin of mathematical biophysics*, 5, décembre 1943, n° 4, p. 115-133.

MENECEUR Yannick, « Les trois grands défis posés par la gouvernance de l'intelligence artificielle et de la transformation numérique », *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale*, 23, Éditions Nota bene, décembre 2021, n° 2.

MERRER Erwan Le et TRÉDAN Gilles, « Le problème du videur : la crédibilité des explications de l'IA en question », *Interstices*, juillet 2021.

NACHBAR Thomas, « Algorithmic Fairness, Algorithmic Discrimination », *Florida State Law Review*, 509, mars 2021, n° 48, p. 509-558.

NETTER Emmanuel, « Quand la force de conviction du scoring bancaire provoque sa chute. L'interprétation extensive, par la CJUE, de la prohibition des décisions entièrement automatisées », *RTD Com.*, septembre 2024, n° 2, p. 342-348.

NETTER Emmanuel, « La part de l'Homme et celle de la machine dans les décisions "automatisées" ». Propositions pour une réécriture de l'article 22 du RGPD », dans BARBÉ Vanessa et MAUCLAIR Stéphane, *Vers un droit de l'algorithme ?*, [s. l.], Mare&Martin, 2022, p. 31-52.

NISSENBAUM Helen, « Accountability in a computerized society », *Science and Engineering Ethics*, 2, mars 1996, n° 1, p. 25-42.

NOVELLI Claudio, TADDEO Mariarosaria et FLORIDI Luciano, « Accountability in artificial intelligence: what it is and how it works », *AI & Society*, 39, février 2023, p. 1871-1882.

OBAR Jonathan A. et OELDORF-HIRSCH Anne, « The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services », *Information, Communication & Society*, 23, Routledge, janvier 2020, n° 1, p. 128-147.

OEHLER Andreas et WENDT Stefan, « Good Consumer Information: The Information Paradigm at Its (Dead) End? », *Journal of Consumer Policy*, 40, 2017, n° 2, p. 179-192.

PARK Roger C. et SAKS Michael J., « Interdisciplinary Trends in Evidence Scholarship », 2005. <https://papers.ssrn.com/abstract=783824>

PERELMAN Ch., « La Distinction Du Fait Et Du Droit Le Point De Vue Du Logicien », *Dialectica*, 15, Wiley, 1961, n° 3/4, p. 601-610.

PLEDEL Chloé, GALBOIS-LEHALLE Diane et CASSAR Bertrand, « L'articulation du projet de règlement sur l'intelligence artificielle avec le droit du numérique européen », *Dalloz Actualité*, juillet 2023, p. 4-12.

PONCE DEL CASTILLO Aida, « The AI Regulation: Entering an AI Regulatory Winter? Why An Ad Hoc Directive on Ai in Employment Is Required », *ETUI Policy Brief*, 6, juin 2021, n° 9, p. 1-9.

PONS Ronan et RISSER Laurent, « Preuve juridique et systèmes d'intelligence artificielle », dans CASTETS-RENARD Céline et EYNARD Jessica, *Un droit de l'intelligence artificielle : Entre règles sectorielles et régime général ; Perspectives comparées* Bruylant Edition, 2023, p. 751-779.

QUIVIGER Pierre-Yves, « Remarques sur les preuves testimoniales en droit », *Cahiers philosophiques*, 142, Réseau Canopé, 2015, n° 3, p. 9-20.

RADÉ Christophe, « La responsabilité du producteur résultant d'un défaut de suivi de ses produits », *Responsabilité civile et assurances*, janvier 2024, n° 01, p. 35-37.

RAGHAVAN Manish, BAROCAS Solon, KLEINBERG Jon *et al.*, Mitigating bias in algorithmic hiring: evaluating claims and practices, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Association for Computing Machinery, 2020, p. 469-481.

RIEU Alain-Marc et MORIZET-MAHOUDEAUX Pierre, « Introduction: Les programmes globaux japonais : le projet Ordinateurs de cinquième génération et ses contextes », *Persée - Portail des revues scientifiques en SHS*, 1994.

ROCHFELD Judith et ZOLYNSKI Célia, « La « loyauté » des « plateformes ». Quelles plateformes ? Quelle loyauté ? », *Dalloz IP/IT*, novembre 2016, n° 11, p. 520-524.

ROCHFELD Judith, « Le projet de loi pour une République numérique : entre espoirs et regrets », *Dalloz IP/IT*, janvier 2016, n° 1, p. 4-5.

ROTH Andrea L., « Machine Testimony », *The Yale Law Journal*, 126, mai 2017, n° 7, p. 1972-2053.

ROUQUETTE Rémi, « Petit traité du procès administratif », *Praxis Dalloz*, 2021.

ROUVROY Antoinette et BERNIS Thomas, « Gouvernamentalité algorithmique et perspectives d'émancipation », *Rezeaux*, n° 177, La Découverte, mai 2013, n° 1, p. 163-196.

RUCKERT Jolina, « The new ontological category hypothesis in human-robot interaction », *Proceedings of the 6th international conference on Human-robot interaction - HRI*, janvier 2011.

SAAS Claire, « La valeur probante des procès verbaux », *AJ Pénal*, mai 2007, n° 5, p. 232.

SÁNCHEZ-MONEDERO Javier, DENCİK Lina et EDWARDS Lilian, What does it mean to « solve » the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Association for Computing Machinery, 2020, p. 458-468.

SCASSA Teresa, « Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making », *University of British Columbia Review*, 54, septembre 2021, n° 1, p. 521-298.

SCHAPPEL Harm, « The New Approach to the New Approach: The Juridification of Harmonized Standards in EU Law », *Maastricht Journal of European and Comparative Law*, 20, SAGE Publications Ltd, décembre 2013, n° 4, p. 521-533.

SCHERER Matthew U., « Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies », *Harvard Journal of Law & Technology*, 29, mai 2015, n° 2, p. 353-401.

SCHIEK Dagmar, « Private Rule-Making and European Governance - Issues of Legitimacy », *European Law Review*, 32, août 2007, p. 443-466.

SCOTT Joanne, « In legal limbo: Post-legislative guidance as a challenge for European administrative law », *Common Market Law Review*, 48, avril 2011, n° Issue 2, p. 329-355.

SÉNÉCHAL Juliette, « L'AI Act dans sa version finale – provisoire –, une hydre à trois têtes », *Dalloz Actualités*, mars 2024, n° 11.

SÉNÉCHAL Juliette, « Les dynamiques actuelles de la future régulation de l'IA, aux niveaux européen et français : entre complexité et angle mort », *Dalloz actualités*, octobre 2023.

SÉNÉCHAL Juliette, « Vote des parlementaires européens sur l'AI Act : vers une réglementation accrue des IA, des modèles de fondation et des IA génératives, s'inspirant du DSA, du Data Act et du RGPD ? », *Dalloz actualités*, juin 2023.

SIEGEL Reva B., « Equality Divided », *Harvard Law Review*, 127, novembre 2013, no 1, p. 1.

SIEGMANN Charlotte et ANDERLJUNG Markus, « The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market », arXiv, 2022.

SIMMONEAUX Antoine, « Le contentieux du Data Protection Officer devant le juge administratif », *AJDA*, octobre 2022, p. 335.

SOLOVE Daniel, « A Taxonomy of Privacy », *University of Pennsylvania Law Review*, 154, janvier 2006, n° 3, p. 477.

TADDEO Mariarosaria, « Trusting Digital Technologies Correctly », *Minds and Machines*, 27, décembre 2017, n° 4, p. 565-568.

TSAMADOS Andreas, FLORIDI Luciano et TADDEO Mariarosaria, « Human control of AI systems: from supervision to teaming », *AI and Ethics*, mai 2024, p. 11-14.

TURING Alan, « Computing Machinery and Intelligence », *Mind*, LIX, octobre 1950, n° 236, p. 433-460.

TWINING William (dir.), « Evidence as a multi-disciplinary subject », dans *Rethinking Evidence: Exploratory Essays*, 2^e éd., Cambridge, Cambridge University Press, 2006, p. 436-456.

VAN DER HEIJDEN Jeroen, « Brighter and Darker Sides of Intermediation: Target-Oriented and Self-Interested Intermediaries in the Regulatory Governance of Buildings », *The ANNALS of the American Academy of Political and Social Science*, 670, mars 2017, n° 1, p. 207-224.

VAN WAEYENBERGE Arnaud et RESTREPO David, « James Elliot construction: A “New(ish) approach” to judicial review of standardisation », *European Law Review*, 42, décembre 2017, p. 882-893.

VEALE Michael et ZUIDERVEEN BORGESIOUS Frederik, « Demystifying the Draft EU Artificial Intelligence Act », *Computer Law Review International*, 22, août 2021, n° 4, p. 97-112.

VERGES Etienne, « Le juge face à la boîte noire: l'intelligence artificielle au tribunal », *Recueil Dalloz*, novembre 2022, n° 38, p. 1920-1927.

VERGES Etienne, « Loyauté et licéité, deux apports majeurs à la théorie de la preuve », *Recueil Dalloz*, février 2014, n° 6, p. 407-411.

VERGÈS Etienne et KHOURY Lara, « Le traitement judiciaire de la preuve scientifique : une modélisation des attitudes du juge face à la connaissance scientifique en droit de la responsabilité civile », *Les Cahiers de droit*, 58, Faculté de droit de l'Université Laval, 2017, n° 3, p. 517-548.

VIAL Géraldine, « Prise en main d'un outil d'intelligence artificielle par des auditeurs de justice : l'office du juge sous l'influence des algorithmes », *Recueil Dalloz*, novembre 2022, n° 38, p. 1928.

VIAL Géraldine, VERGÈS Étienne et GAUTRAIS Vincent, « Preuves scientifiques et technologiques », *Cahiers Droit, Sciences & Technologies*, PUP, octobre 2019, n° 9, p. 179-197.

VORMS Marion, « La valeur probante du témoignage : perspectives épistémologique et juridique », *Cahiers philosophiques*, 142, Réseau Canopé, 2015, n° 3, p. 21-52.

WACHTER Sandra, MITTELSTADT Brent et FLORIDI Luciano, « Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation », *International Data Privacy Law*, 1, 2017, n° 2, p. 76-99.

WACHTER Sandra, MITTELSTADT Brent et RUSSELL Chris, « Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law », *West Virginia Law Review*, 123, 2021, n° 3, p. 753-790.

WACHTER Sandra, MITTELSTADT Brent et RUSSELL Chris, « Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI », *Computer Law & Security Review* 41, 41, juillet 2021, p. 105567.

WAEYENBERGE Arnaud Van, « La normalisation technique en Europe. L'Empire (du droit) contre-attaque », *Revue internationale de droit économique*, t. XXXII, 2018, n° 3, p. 305-317.

WAGNER Ben, « Ethics As An Escape From Regulation. From “Ethics-Washing” To Ethics-Shopping? », *Amsterdam University Press*, 2018, p. 84-89.

WALLERMAN Anna, « Pie in the Sky When You Die: Civil Liability of Notified Bodies under the Medical Devices Directive: Schmitt », *Common Market Law Review*, 55, 2018, p. 265.

WEXLER Rebecca, « Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System », 70, mai 2018, n° 5, p. 1343.

WHITTLESTONE Jess, NYRUP Rune, ALEXANDROVA Anna *et al.*, The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions, *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, New York, NY, USA, Association for Computing Machinery, 2019, p. 195-200. <https://doi.org/10.1145/3306618.3314289>

WINNER Langdon, « Do Artifacts Have Politics ? », 109, 1980, n° 1, p. 121-136.

WRÓBLEWSKI Jerzy, « Facts in Law », *ARSP: Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, 59, Franz Steiner Verlag, 1973, n° 2, p. 161-178.

YIN Ming, VAUGHAN Jennifer et WALLACH Hanna, Understanding the Effect of Accuracy on Trust in Machine Learning Models, 2019, p. 1-12.

ZEHLIKE Meike, LOOSLEY Alex, JONSSON Håkan *et al.*, « Beyond Incompatibility: Trade-offs between Mutually Exclusive Fairness Criteria in Machine Learning and Law », arXiv, 2022. <http://arxiv.org/abs/2212.00469>

ZHANG Yunfeng, LIAO Q. Vera et BELLAMY Rachel K. E., Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona, Spain, Association for Computing Machinery, 2020, p. 295-305. <https://doi.org/10.1145/3351095.3372852>

ŽLIOBAITĖ Indrė et CUSTERS Bart, « Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models », *Artificial Intelligence and Law*, 24, juin 2016, n° 2, p. 183-201.

ZUBOFF Shoshana, « Big other: surveillance capitalism and the prospects of an information civilization », *Journal of Information Technology*, 30, mars 2015, n° 1, p. 75-89.

Contributions aux mélanges et actes de colloques

ROCHFELD Judith et ZOLYNSKI Célia, « La valeur des émotions : Quel régime pour le “capitalisme mental” ? », *Entre art et technique : les dynamiques du droit. Mélanges en l'honneur de Pierre Sirinelli*, Dalloz, 2022, p. 749-770.

Notes, observations, avis et conclusions de jurisprudence

Jurisprudence française

Conseil constitutionnel

Conseil constitutionnel, 21 mars 2019, n° 2018-768, *Décision n° 2018-768 QPC du 21 mars 2019*.

Conseil constitutionnel, 23 janvier 1987, n° 86-224, *Décision n° 86-224 DC*. ☞

Conseil d'État

Conseil d'État, 21 octobre 2022, n° 459254, *Conseil d'État, 10ème - 9ème chambres réunies, 21/10/2022, 459254, Publié au recueil Lebon*, Publié au recueil Lebon. ☞

Conseil d'État, 26 avril 2022, n° 442364, *Conseil d'État, 10ème chambre, 26/04/2022, 442364, Inédit au recueil Lebon*, Inédit au recueil Lebon. ☞

Conseil d'État, 19 juin 2020, n° 434684, *Conseil d'État, 10ème - 9ème chambres réunies, 19/06/2020, 434684*. ☞

Conseil d'État, 16 octobre 2019, n° 433069, *Conseil d'État, 10ème - 9ème chambres réunies, 16/10/2019, 433069, Publié au recueil Lebon*, Publié au recueil Lebon. ☞

Conseil d'État, 12 juin 2019, n° 427916, *Conseil d'État, 4ème - 1ère chambres réunies, 12/06/2019, 427916*. ☞

Conseil d'État, 23 décembre 2016, n° 405791, *Conseil d'État, Juge des référés, 23/12/2016, 405791*. ☞

Conseil d'État, 6 avril 2016, n° 374224, *Conseil d'État, 6ème / 1ère SSR, 06/04/2016, 374224*, Publié au recueil Lebon, Publié au recueil Lebon. [↗](#)

Conseil d'État, 21 mars 2016, n° 368082, *Conseil d'État, Assemblée, 21/03/2016, 368082*, Publié au recueil Lebon, Publié au recueil Lebon. [↗](#)

Conseil d'État, 21 mars 2016, n° 390023, *Conseil d'État, Assemblée, 21/03/2016, 390023*, Publié au recueil Lebon, Publié au recueil Lebon. [↗](#)

Conseil d'État, 27 août 2014, n° 374473, *Conseil d'État, 7ème sous-section jugeant seule, 27/08/2014, 374473*, Inédit au recueil Lebon, Inédit au recueil Lebon. [↗](#)

Conseil d'État, 11 avril 2014, n° 348972, *Conseil d'État, 5ème / 4ème SSR, 11/04/2014, 348972*. [↗](#)

Conseil d'État, 23 octobre 2013, n° 360961, *Conseil d'État, 6ème et 1ère sous-sections réunies, 23/10/2013, 360961*. [↗](#)

Conseil d'État, 17 octobre 2012, n° 351833, *Conseil d'État, 3ème sous-section jugeant seule, 17/10/2012, 351833*, Inédit au recueil Lebon, Inédit au recueil Lebon. [↗](#)

Conseil d'État, 16 février 2009, n° 274000, *Conseil d'État, Assemblée, 16/02/2009, 274000*, Publié au recueil Lebon, Publié au recueil Lebon. [↗](#)

Conseil d'État, 12 décembre 2007, n° 298155, *Conseil d'État, 6ème et 1ère sous-sections réunies, 12/12/2007, 298155*, Inédit au recueil Lebon, Inédit au recueil Lebon. [↗](#)

Conseil d'État, Assemblée, 6 novembre 2002, n° 194295, *M. Sun Myung Moon*. [↗](#)

Cour de cassation

Cour de cassation, 15 novembre 2023, n° 22-21.174, *Cour de cassation, civile, Chambre civile 1, 15 novembre 2023, 22-21.174*, Publié au bulletin, Publié au bulletin. [↗](#)

Cour de cassation, 15 novembre 2023, n° 22-21.178, *Cour de cassation, civile, Chambre civile 1, 15 novembre 2023, 22-21.178*, Publié au bulletin, Publié au bulletin. [↗](#)

Cour de cassation, 15 novembre 2023, n° 22-21.179, *Cour de cassation, civile, Chambre civile 1, 15 novembre 2023, 22-21.179*, Publié au bulletin, Publié au bulletin. [↗](#)

Cour de cassation, 15 novembre 2023, n° 22-21.180, *Cour de cassation, civile, Chambre civile 1, 15 novembre 2023, 22-21.180*, Publié au bulletin, Publié au bulletin. [↗](#)

Cour de cassation, Chambre civile 3, 1 mars 2023, n° 21-25.497, *Cour de cassation, civile, Chambre civile 3, 1 mars 2023, 21-25.497*, Inédit, Inédit. [↗](#)

Cour de cassation, Chambre commerciale, 10 novembre 2021, n° 20-14.670, *Cour de cassation, civile, Chambre commerciale, 10 novembre 2021, 20-14.670*, Publié au bulletin, Publié au bulletin. [↗](#)

Cour de cassation, Chambre sociale, 25 novembre 2020, n° 17-19.523, *Cour de cassation, civile, Chambre sociale, 25 novembre 2020, 17-19.523, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre sociale, 30 septembre 2020, n° 19-12.058, *Cour de cassation, civile, Chambre sociale, 30 septembre 2020, 19-12.058, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, 8 juillet 2020, n° 18-20.961, *Cour de cassation, civile, Chambre civile 1, 8 juillet 2020, 18-20.961, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre criminelle, 14 décembre 2021, n° 20-86.969, *Cour de cassation, criminelle, Chambre criminelle, 14 décembre 2021, 20-86.969, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre civile 1, 26 février 2020, n° 19-13.423 19-14.240, *Cour de cassation, civile, Chambre civile 1, 26 février 2020, 19-13.423 19-14.240, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre civile 1, 27 novembre 2019, n° 18-16.537, *Cour de cassation, civile, Chambre civile 1, 27 novembre 2019, 18-16.537, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre criminelle, 18 juin 2019, n° 19-80.105, *Cour de cassation, criminelle, Chambre criminelle, 18 juin 2019, 19-80.105, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre commerciale, 5 juin 2019, n° 17-22.192, *Cour de cassation, civile, Chambre commerciale, 5 juin 2019, 17-22.192, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre criminelle, 14 mai 2019, n° 18-84.552, *Cour de cassation, Chambre criminelle, du 14 mai 2019, 18-84.552, Inédit*. [↗](#)

Cour de cassation, 10 octobre 2018, n° 15-26.093, *Cour de cassation, civile, Chambre civile 1, 10 octobre 2018, 15-26.093, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre civile 1, 11 juillet 2018, n° 17-20.154, *Cour de cassation, civile, Chambre civile 1, 11 juillet 2018, 17-20.154, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre civile 1, 11 juillet 2018, n° 17-17.441 17-19.581, *Cour de cassation, civile, Chambre civile 1, 11 juillet 2018, 17-17.441 17-19.581, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre commerciale, 4 juillet 2018, n° 17-10.158, *Cour de cassation, civile, Chambre commerciale, 4 juillet 2018, 17-10.158, Inédit*, Inédit. [↗](#)

Cour de cassation, Chambre sociale, 4 juillet 2018, n° 17-18.241, *Cour de cassation, civile, Chambre sociale, 4 juillet 2018, 17-18.241, Publié au bulletin*, Publié au bulletin. [↗](#)

Cour de cassation, Chambre sociale, 4 juillet 2018, n° 17-18.241, *Cour de cassation, civile, Chambre sociale, 4 juillet 2018, 17-18.241, Publié au bulletin*, Publié au bulletin. [↗](#)

Conseil d'État, 14 février 2018, n° 407914, *Conseil d'État, 5ème et 6ème chambres réunies, 14/02/2018, 407914.* ↗

Cour de cassation, Chambre civile 1, 22 juin 2017, n° 15-27.845, *Cour de cassation, civile, Chambre civile 1, 22 juin 2017, 15-27.845, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre commerciale, 20 avril 2017, n° 15-21.877, *Cour de cassation, civile, Chambre commerciale, 20 avril 2017, 15-21.877, Inédit,* Inédit. ↗

Cour de cassation, Chambre civile 1, 1 mars 2017, n° 15-29.413, *Cour de cassation, civile, Chambre civile 1, 1 mars 2017, 15-29.413, Inédit,* Inédit. ↗

Cour de cassation, Chambre sociale, 9 novembre 2016, n° 15-10.203, *Cour de cassation, civile, Chambre sociale, 9 novembre 2016, 15-10.203, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre civile 1, 25 février 2016, n° 15-12.403, *Cour de cassation, civile, Chambre civile 1, 25 février 2016, 15-12.403, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre criminelle, 17 mars 2015, n° 14-88.351, *Cour de cassation, criminelle, Chambre criminelle, 17 mars 2015, 14-88.351, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre civile 1, 4 février 2015, n° 13-19.781, *Cour de cassation, civile, Chambre civile 1, 4 février 2015, 13-19.781, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre civile 1, 10 septembre 2014, n° 13-22.612, *Cour de cassation, civile, Chambre civile 1, 10 septembre 2014, 13-22.612, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre civile 2, 20 mars 2014, n° 13-14.985, *Cour de cassation, civile, Chambre civile 2, 20 mars 2014, 13-14.985, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre civile 1, 5 février 2014, n° 12-23.467, *Cour de cassation, civile, Chambre civile 1, 5 février 2014, 12-23.467, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre criminelle, 7 janvier 2014, n° 13-85.246, *Cour de cassation, criminelle, Chambre criminelle, 7 janvier 2014, 13-85.246, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, 25 septembre 2013, n° 12-24.588, *Cour de cassation, civile, Chambre civile 1, 25 septembre 2013, 12-24.588, Inédit,* Inédit. ↗

Cour de cassation, Chambre criminelle, 23 janvier 2013, n° 12-85.059, *Cour de cassation, criminelle, Chambre criminelle, 23 janvier 2013, 12-85.059, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre mixte, 28 septembre 2012, n° 11-18.710, *Cour de cassation, Chambre mixte, 28 septembre 2012, 11-18.710, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre mixte, 28 septembre 2012, n° 11-18.710, *Cour de cassation, Chambre mixte, 28 septembre 2012, 11-18.710, Publié au bulletin,* Publié au bulletin. ↗

Cour de cassation, Chambre civile 2, 26 mai 2011, n° 10-20.048, *Cour de cassation, civile, Chambre civile 2, 26 mai 2011, 10-20.048, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre sociale, 22 mars 2011, n° 09-68.693, *Cour de cassation, civile, Chambre sociale, 22 mars 2011, 09-68.693, Inédit*, Inédit. ☞

Cour de cassation, Chambre sociale, 2 février 2011, n° 10-14.263, *Cour de cassation, civile, Chambre sociale, 2 février 2011, 10-14.263, Inédit*, Inédit. ☞

Cour de cassation, Chambre commerciale, 14 septembre 2010, n° 09-16.347, *Cour de cassation, civile, Chambre commerciale, 14 septembre 2010, 09-16.347, Inédit*, Inédit. ☞

Cour de cassation, Chambre civile 1, 9 juillet 2009, n° 08-11.073, *Cour de cassation, civile, Chambre civile 1, 9 juillet 2009, 08-11.073, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre criminelle, 24 juin 2009, n° 09-81.119, *Cour de cassation, criminelle, Chambre criminelle, 24 juin 2009, 09-81.119, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, 6 mai 2009, n° 08-10.936, *Cour de cassation, civile, Chambre civile 1, 6 mai 2009, 08-10.936, Inédit*, Inédit. ☞

Cour de cassation, Chambre civile 1, 16 octobre 2008, n° 07-15.778, *Cour de cassation, civile, Chambre civile 1, 16 octobre 2008, 07-15.778, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre civile 2, 26 juin 2008, n° 07-13.875, *Cour de cassation, civile, Chambre civile 2, 26 juin 2008, 07-13.875, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre civile 1, 22 mai 2008, n° 06-18.848, *Cour de cassation, civile, Chambre civile 1, 22 mai 2008, 06-18.848, Inédit*, Inédit. ☞

Cour de cassation, Chambre sociale, 18 mars 2008, n° 06-45.093, *Cour de cassation, civile, Chambre sociale, 18 mars 2008, 06-45.093, Publié au bulletin*, Publié au bulletin. ☞

Cour de Cassation, Chambre sociale, 9 octobre 2007, n° 06-42.959, *Cour de Cassation, Chambre sociale, du 9 octobre 2007, 06-42.959, Inédit*. ☞

Cour de cassation, Chambre civile 1, 12 juillet 2007, n° 06-10.362, *Cour de cassation, civile, Chambre civile 1, 12 juillet 2007, 06-10.362, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre criminelle, 13 juin 2007, n° 06-85.441, *Cour de cassation, criminelle, Chambre criminelle, 13 juin 2007, 06-85.441, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre commerciale, 15 mai 2007, n° 06-10.606, *Cour de cassation, civile, Chambre commerciale, 15 mai 2007, 06-10.606, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre criminelle, 24 avril 2007, n° 06-88.051, *Cour de cassation, criminelle, Chambre criminelle, 24 avril 2007, 06-88.051, Publié au bulletin*, Publié au bulletin. ☞

Cour de cassation, Chambre criminelle, 6 février 2007, n° 06-83.356, *Cour de cassation, criminelle, Chambre criminelle, 6 février 2007, 06-83.356, Publié au bulletin*, Publié au bulletin. ☐

Cour de cassation, Chambre civile 2, 21 septembre 2006, n° 05-21.978, *Cour de cassation, civile, Chambre civile 2, 21 septembre 2006, 05-21.978, Publié au bulletin*, Publié au bulletin. ☐

Cour de Cassation, Chambre civile 1, 21 février 2006, n° 04-13.51, *Cour de Cassation, Chambre civile 1, du 21 février 2006, 04-13.512, Publié au bulletin*. ☐

Cour de Cassation, Chambre civile 2, 8 février 2006, n° 05-14.198, *Cour de Cassation, Chambre civile 2, du 8 février 2006, 05-14.198, Publié au bulletin*. ☐

Cour de Cassation, Chambre civile 1, 10 mai 2005, n° 02-11.759, *Cour de Cassation, Chambre civile 1, du 10 mai 2005, 02-11.759, Publié au bulletin*. ☐

Cour de Cassation, Chambre sociale, 13 janvier 2004, n° 01-45.931; 01-45.932, *Cour de Cassation, Chambre sociale, du 13 janvier 2004, 01-45.931 01-45.932, Publié au bulletin*. ☐

Cour de Cassation, Chambre criminelle, 14 octobre 2003, n° 03-84.539, *Cour de Cassation, Chambre criminelle, du 14 octobre 2003, 03-84.539, Publié au bulletin*. ☐

Cour de Cassation, Chambre commerciale, 4 mars 2003, n° 00-18.668, *Cour de Cassation, Chambre commerciale, du 4 mars 2003, 00-18.668, Inédit*. ☐

Cour de Cassation, Chambre criminelle, 28 novembre 2001, n° 01-86.467, *Cour de Cassation, Chambre criminelle, du 28 novembre 2001, 01-86.467, Publié au bulletin*. ☐

Cour de Cassation, Chambre sociale, 2 octobre 2001, n° 99-42.942, *Cour de Cassation, Chambre sociale, du 2 octobre 2001, 99-42.942, Publié au bulletin*. ☐

Cour de Cassation, Chambre sociale, 11 juillet 2001, n° 99-42.927, *Cour de Cassation, Chambre sociale, du 11 juillet 2001, 99-42.927, Inédit*. ☐

Cour de Cassation, Chambre sociale, 27 mars 2001, *Cour de Cassation, Chambre sociale, du 27 mars 2001, 98-44.666, Publié au bulletin*. ☐

Cour de Cassation, Chambre criminelle, 12 décembre 2000, n° 00-83.852, *Cour de Cassation, Chambre criminelle, du 12 décembre 2000, 00-83.852, Publié au bulletin*. ☐

Cour de Cassation, Chambre criminelle, 18 octobre 2000, n° 00-82.68, *Cour de Cassation, Chambre criminelle, du 18 octobre 2000, 00-82.689, Inédit*. ☐

Cour de Cassation, Chambre civile 1, 28 mars 2000, n° 98-12.806, *Cour de Cassation, Chambre civile 1, du 28 mars 2000, 98-12.806, Publié au bulletin*. ☐

Cour de Cassation, Chambre sociale, 14 mars 2000, n° 97-45.899, *Cour de Cassation, Chambre sociale, du 14 mars 2000, 97-45.899, Inédit*. ☐

Cour de Cassation, Chambre civile 1, 15 juillet 1999, n° 97-16.783, *Cour de Cassation, Chambre civile 1, du 15 juillet 1999, 97-16.783, Inédit*. ☐

Cour de Cassation, Chambre criminelle, 5 mai 1999, n° 97-83.117, *Cour de Cassation, Chambre criminelle, du 5 mai 1999, 97-83.117, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 2, 7 janvier 1999, n° 97-10.831, *Cour de Cassation, Chambre civile 2, du 7 janvier 1999, 97-10.831, Publié au bulletin.* ☞

Cour de Cassation, Chambre sociale, 7 mai 1998, n° 95-44.382, *Cour de Cassation, Chambre sociale, du 7 mai 1998, 95-44.382, Inédit.* ☞

Cour de Cassation, Chambre civile 1, 5 novembre 1996, n° 93-21.762, *Cour de Cassation, Chambre civile 1, du 5 novembre 1996, 93-21.762, Publié au bulletin.* ☞

Cour de Cassation, Chambre commerciale, 8 octobre 1996, n° 94-17.96, *Cour de Cassation, Chambre commerciale, du 8 octobre 1996, 94-17.967, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 1, 2 juillet 1996, n° 94-14.943, *Cour de Cassation, Chambre civile 1, du 2 juillet 1996, 94-14.943, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 1, 25 juin 1996, n° 94-11.745, *Cour de Cassation, Chambre civile 1, du 25 juin 1996, 94-11.745, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 1, 9 mai 1996, n° 94-13.310, *Cour de Cassation, Chambre civile 1, du 9 mai 1996, 94-13.310, Publié au bulletin.* ☞

Cour de Cassation, Chambre criminelle, 24 janvier 1996, n° 95-83.574, *Cour de Cassation, Chambre criminelle, du 24 janvier 1996, 95-83.574, Inédit.* ☞

Cour de Cassation, Chambre civile 1, 7 juin 1995, n° 93-13.060, *Cour de Cassation, Chambre civile 1, du 7 juin 1995, 93-13.060, Publié au bulletin.* ☞

Cour de Cassation, Chambre sociale, 16 janvier 1991, n° 89-41.052, *Cour de Cassation, Chambre sociale, du 16 janvier 1991, 89-41.052, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 3, 14 juin 1989, n° 87-19.312, *Cour de Cassation, Chambre civile 3, du 14 juin 1989, 87-19.312, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 1, 2 mai 1989, n° 86-17.804, *Cour de cassation, Chambre civile 1, du 2 mai 1989, 86-17.804, Inédit, Inédit.* ☞

Cour de Cassation, Chambre civile 1, 21 juillet 1987, n° 85-16.436, *Cour de Cassation, Chambre civile 1, du 21 juillet 1987, 85-16.436, Publié au bulletin.* ☞

Cour de Cassation, Chambre commerciale, 18 janvier 1984, n° 82-14.977, *Cour de cassation, Chambre commerciale, du 18 janvier 1984, 82-14.977, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 1, 26 novembre 1980, n° 79-13870, *Cour de cassation, Chambre civile 1, du 26 novembre 1980, 79-13870, Bulletin des arrêts Cour de Cassation Chambre civile 1 N. 308, AHJUCAF.* ☞

Cour de Cassation, Chambre civile 1, 21 octobre 1975, n° 74-12.73, *Cour de Cassation, Chambre civile 1, du 21 octobre 1975, 74-12.739, Publié au bulletin.* ☞

Cour de Cassation, Chambre civile 1, 15 juillet 1975, n° 74-12.965, *Cour de Cassation, Chambre civile 1, du 15 juillet 1975, 74-12.965, Publié au bulletin.* ↗

Cour de Cassation, Chambre commerciale, 28 mai 1974, n° 73-10.022, *Cour de Cassation, Chambre commerciale, du 28 mai 1974, 73-10.022, Publié au bulletin.* ↗

Cour de Cassation, Chambre civile 1, 27 janvier 1971, n° 69-13.273, *Cour de Cassation, Chambre civile 1, du 27 janvier 1971, 69-13.273, Publié au bulletin.* ↗

Cour de Cassation, Chambre civile 2, 12 mars 1970, n° 69-12.291, *Cour de Cassation, Chambre civile 2, du 12 mars 1970, 69-12.291, Publié au bulletin.* ↗

Cour de Cassation, Chambre civile 1, 24 novembre 1969, *Cour de Cassation, Chambre civile 1, du 24 novembre 1969, Publié au bulletin.* ↗

Cour de Cassation, Chambre civile 2, 20 juin 1968, *Cour de Cassation, Chambre civile 2, du 20 juin 1968, Publié au bulletin.* ↗

Cours d'appel

Cour d'appel de Rouen, Chambre des Etrangers, 22 octobre 2022, n° 22/03388, *Cour d'appel de Rouen, Chambre des Etrangers, 18 octobre 2022, 22/03388.* ↗

Jurisprudence européenne

Jurisprudence de la CEDH

Cour européenne des droits de l'Homme, 7 août 2021, n° 27516-14, *M.P. c/ Portugal.*

Cour européenne des droits de l'Homme, 16 novembre 2020, n° 10795/14, *Vladimir Kharitonov c. Russie.*

Cour européenne des droits de l'Homme, 8 février 2018, n° 31446/12, *Ben Faiza c/ France.*

Cour européenne des droits de l'Homme, 5 octobre 2017, n° 39014/12, *Kormev c. Bulgarie.*

Cour européenne des droits de l'Homme, 22 juin 2017, n° 8806/12, *Aycaguer c/ France.*

Cour européenne des droits de l'Homme, 2 juin 2015, n° 22037/13, *Canonne c. France.*

Cour européenne des droits de l'Homme, 18 septembre 2014, n° 21010/10, *Brunet c/ France.*

Cour européenne des droits de l'Homme, 18 avril 2013, n° 19522/09, *M.K. c/ France.*

Cour européenne des droits de l'Homme, 21 juillet 2011, n° 42310/04, *Nechiporuk et Yonkalo c. Ukraine.*

Cour européenne des droits de l'Homme, 1 juin 2010, n° 22978/05, *GÄFGEN c. ALLEMAGNE.*

Cour européenne des droits de l'Homme, 10 mars 2009, n° 4378/02, *Bykov c. Russie.*

Cour européenne des droits de l'Homme, 13 novembre 2007, n° 57325/00, *D.H. et Autres contre République Tchèque.*

Cour européenne des droits de l'Homme, 11 juillet 2006, n° 54810/00, *Jalloh c. Allemagne*.

Cour européenne des droits de l'Homme, 31 mai 2005, n° 59842, *Vetter c/ France*.

Cour européenne des droits de l'Homme, 5 novembre 2002, n° 48539/99, *Allan c. Royaume-Uni*.

Cour européenne des droits de l'Homme, 12 mai 2000, n° 35394/97, *Khan c. Royaume-Uni*.

Cour européenne des droits de l'Homme, 16 février 2000, n° 27052/95, *Jasper c. Royaume-Uni*.

Cour européenne des droits de l'Homme, 18 mars 1997, n° 29266/08, 32422/08, ..., 27813/13, *Mantovanelli c. France*.

Cour européenne des droits de l'Homme, 26 mars 1996, n° 20524/92, *Doorson c. Pays-Bas*.

Cour européenne des droits de l'Homme, 20 février 1996, n° 19075/91, *Vermeulen c. Belgique*.

Cour européenne des droits de l'Homme, 23 juin 1993, n° 12952/87, *Ruiz-Mateos c. Espagne*.

Cour européenne des droits de l'Homme, 14 avril 1990, n° 11105/84, *Huvig c. France*.

Cour européenne des droits de l'Homme, 20 novembre 1989, n° 62171/00, *Kostovski c. Pays-Bas*.

Cour européenne des droits de l'Homme, 29 mai 1986, n° 53268/08, 53273/08, ..., 25125/12, *Feldbrugge c. Pays-Bas*.

Jurisprudence de l'Union européenne

Cour de justice de l'Union européenne, 5 mars 2024, n° C-588/21 P, *Public.Resource.Org Inc et Right to Know CLG c. Commission européenne*.

Cour de justice de l'Union européenne, 7 décembre 2023, n° C-634/21, *SCHUFA Holding (Scoring)*.

Cour de justice de l'Union européenne, 22 juin 2023, n° C-588/21 P, *Conclusions de l'avocate générale Mme Laila Medina - Public.Resource.Org, Inc., Right to Know CLG contre Commission européenne*.

Cour de justice de l'Union européenne, 26 avril 2022, n° C-401/19, *République de Pologne contre Parlement européen et Conseil de l'Union européenne*. ☞

Cour de justice, 22 février 2022, n° C-160/20, *Stichting Rookpreventie Jeugd ea contre Staatssecretaris van Volksgezondheid, Welzijn en Sport*. ☞

Cour de justice de l'Union européenne, 15 juillet 2021, n° C-401/19, *Conclusions de l'avocat général M. H. Saugmandsgaard Øe, présentées le 15 juillet 2021. République de Pologne contre Parlement européen et Conseil de l'Union européenne. Recours en annulation*. ☞

Cour de justice de l'Union européenne, 16 janvier 2019, n° C-265/17, *Commission européenne c/ UPS et Fedex*.

Cour de justice, 14 décembre 2017, no C-630/16, *Anstar*. ↗

Cour de justice, 21 juin 2017, n° C-621/15, *N W ea contre Sanofi Pasteur MSD SNC ea*. ↗

Tribunal de l'Union européenne, 7 mars 2017, n° T-194/13, *UPS c/ Commission européenne*.

Cour de justice de l'Union européenne, 16 février 2017, n° C-219/15, *Elisabeth Schmitt contre TÜV Rheinland LGA Products GmbH*. ↗

Tribunal de l'Union européenne (Deuxième chambre), 26 janvier 2017, n° T-474/15, *Global Garden Products Italy SpA (GGP Italy) contre Commission européenne*. ↗

Cour de justice de l'Union européenne, 3ème chambre, 27 octobre 2016, n° C-613/14, *James Elliott Construction Limited contre Irish Asphalt Limited*. ↗

Cour de justice de l'Union européenne, 15 septembre 2016, n° C-219/15, *Conclusions de l'avocat général Mme E. Sharpston, présentées le 15 septembre 2016. Elisabeth Schmitt contre TÜV Rheinland LGA Products GmbH*. ↗

Cour de justice, 15 septembre 2016, n° C-484/14, *Tobias Mc Fadden contre Sony Music Entertainment Germany GmbH*. ↗

28 janvier 2016, n° C-613/14, *Conclusions de l'avocat général M. Manuel Campos Sanchez-Bordona. Affaire James Elliott Construction Limited contre Irish Asphalt Limited*. ↗

Cour de justice de l'Union européenne, 27 mars 2014, n° C-314/12, *UPC Telekabel Wien GmbH contre Constantin Film Verleih GmbH et Wega Filmproduktionsgesellschaft mbH*. ↗

Cour de justice de l'Union européenne, 22 janvier 2014, n° C-270/12, *Royaume-Uni de Grande-Bretagne et d'Irlande du Nord contre Parlement européen et Conseil de l'Union européenne*. ↗

Cour de justice de l'Union européenne, 12 juillet 2012, n° C-171/11, *Frabo SpA contre Deutsche Vereinigung des Gas- und Wasserfaches eV (DVGW) — Technisch-Wissenschaftlicher Verein*. ↗

Cour de justice de l'Union européenne, 16 février 2012, n° C-360/10, *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) contre Netlog NV*.

Cour de justice, 22 décembre 2010, n° C-279/09, *DEB Deutsche Energiehandels- und Beratungsgesellschaft mbH contre Bundesrepublik Deutschland*. ↗

Cour de justice de l'Union européenne, 16 juillet 2009, n° C-12/08, *Mono Car Styling SA, en liquidation contre Dervis Odemis e.a*. ↗

Tribunal de l'Union européenne, 11 septembre 2002, n° T-13/99, *Pfizer Animal Health*.

Tribunal de l'Union européenne, 11 septembre 2002, n° T-70/99, *Alpharma c. Conseil de l'Union européenne*.

Cour de justice de l'Union européenne, 25 juillet 2002, n° C-50/00 P, *Unión de Pequeños Agricultores contre Conseil de l'Union européenne*. ↗

Cour de justice des Communautés européennes, 25 avril 2002, n° C-52/00, *Commission des Communautés européennes contre République française*.

Cour de justice des Communautés européennes, 25 avril 2002, n° C-154/00, *Commission des Communautés européennes contre République hellénique*.

Cour de justice de l'Union européenne, 21 mars 2002, n° C-50/00 P., *Conclusions de l'avocat général M. Jacobs présentées le 21 mars 2002 - Affaire C-50/00 P - Unión de Pequeños Agricultores contre Conseil de l'Union européenne*. ↗

Cour de justice de l'Union européenne, 9 février 1999, n° C-167/97, *Regina contre Secretary of State for Employment, ex parte Nicole Seymour-Smith et Laura Perez*. ↗

Cour de justice des Communautés européennes, 31 mai 1989, n° 109/88, *Conclusions de l'avocat général M. Carl Otto Lenz - Handels- og Kontorfunktionærernes Forbund I Danmark contre Dansk Arbejdsgiverforening, agissant pour Danfoss*.

Cour de justice de l'Union européenne, 23 avril 1986, n° 294/83, *Parti écologiste « Les Verts » contre Parlement européen*. ↗

Cour de justice de l'Union européenne, 13 Juin 1958, n° 9-56, *Meroni & Co., Industrie Metallurgiche, SpA contre Haute Autorité de la Communauté européenne du charbon et de l'acier*. ↗

Jurisprudence étrangère

Cour suprême du Wisconsin, 13 juillet 2016, n° 881 N.W.2d 749, *State v. Loomis*. ↗

Federal Court, 7 mai 2002, n° T-1119-01, *Byer v. Canada*. ↗

Textes législatifs, réglementaires et textes internationaux

Textes français

Lois

« LOI n° 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique (1) », 2024-449, 2024.

« LOI n° 2023-380 du 19 mai 2023 relative aux jeux Olympiques et Paralympiques de 2024 et portant diverses autres dispositions (1) », 2023-380, 2023.

« LOI n° 2021-1017 du 2 août 2021 relative à la bioéthique (1) », 2021-1017, 2021.

« LOI n° 2019-1479 du 28 décembre 2019 de finances pour 2020 », 2019-1479, 2019.

« LOI n° 2019-1428 du 24 décembre 2019 d'orientation des mobilités », 2019-1428, 2019.

« Délibération n° 2019-093 du 4 juillet 2019 portant adoption de lignes directrices relatives à l'application de l'article 82 de la loi du 6 janvier 1978 modifiée aux opérations de lecture et écriture dans le terminal d'un utilisateur (notamment aux cookies et autres traceurs). », 2019.

« LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice », 2019-222, 2019.

« LOI n° 2018-670 du 30 juillet 2018 relative à la protection du secret des affaires (1) », 2018-670, 2018.

« LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles (1) », 2018-493, 2018.

« LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique », 2016-1321, 2016.

« LOI n° 2008-496 du 27 mai 2008 portant diverses dispositions d'adaptation au droit communautaire dans le domaine de la lutte contre les discriminations (1) », 2008-496, 2008.

« LOI no 94-653 du 29 juillet 1994 relative au respect du corps humain (1) », 94-653, 1994.

« Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ».

« Loi n° 89-469 du 10 juillet 1989 relative à diverses dispositions en matière de sécurité routière et en matière de contraventions (1) ».

« Loi n° 92-644 du 13 juillet 1992 modifiant la loi n° 91-650 du 9 juillet 1991 portant réforme des procédures civiles d'exécution et l'ordonnance n° 45-2592 du 2 novembre 1945 relative au statut des huissiers et comportant diverses dispositions relatives aux procédures civiles d'exécution ».

« Loi n° 94-665 du 4 août 1994 relative à l'emploi de la langue française. ».

« Loi n° 2001-1062 du 15 novembre 2001 relative à la sécurité quotidienne ».

« Loi n° 2003-495 du 12 juin 2003 renforçant la lutte contre la violence routière (1). ».

Ordonnances

« Ordonnance n° 2018-1125 du 12 décembre 2018 prise en application de l'article 32 de la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles et portant modification de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel », 2018.

« Ordonnance n° 2016-131 du 10 février 2016 portant réforme du droit des contrats, du régime général et de la preuve des obligations ».

Décrets

« Décret n° 2019-871 du 21 août 2019 relatif au droit de conduire limité aux seuls véhicules équipés d'un éthylotest anti-démarrage », 2019-871, 2019.

« Décret n° 2018-1126 du 11 décembre 2018 relatif à la protection du secret des affaires », 2018-1126, 2018.

« Décret n° 2017-1434 du 29 septembre 2017 relatif aux obligations d'information des

opérateurs de plateformes numériques », 2017-1434, 2017.

« Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique », 2017-330, 2017.

« Décret n° 2016-1673 du 5 décembre 2016 relatif à la fiabilité des copies et pris pour l'application de l'article 1379 du code civil », 2016-1673, 2016.

« Décret n° 74-74 du 30 janvier 1974 réglementant la catégorie d'instruments de mesurage : cinémomètres de contrôle routier. », 74-74, 1974.

« Décret n°71-941 du 26 novembre 1971 relatif aux actes établis par les notaires », 71-941, 1971.

« Décret n°47-1047 du 12 juin 1947 relatif à la formule exécutoire. », 47-1047, 1947.

« Décret n° 97-109 du 6 février 1997 relatif aux conditions d'agrément des personnes habilitées à procéder à des identifications par empreintes génétiques dans le cadre d'une procédure judiciaire ou de la procédure extrajudiciaire d'identification des personnes décédées. ».

« Décret n° 2021-148 du 11 février 2021 portant modalités de mise en œuvre par la direction générale des finances publiques et la direction générale des douanes et droits indirects de traitements informatisés et automatisés permettant la collecte et l'exploitation de données rendues publiques sur les sites internet des opérateurs de plateforme en ligne. ».

Arrêtés

« Arrêté du 4 juin 2009 relatif aux cinémomètres de contrôle routier ».

« Arrêté du 8 juillet 2003 relatif au contrôle des éthylomètres. ».

« Arrêté du 31 décembre 1985 relatif à la construction, la vérification et l'utilisation des éthylomètres. ».

Circulaires

« Circulaire relative aux nouvelles dispositions en matière de circulation et de sécurité routières applicables à compter du 1er décembre 1992 par les services de police », 1992. ↗

Autres

« Vocabulaire des transports et de l'urbanisme (liste de termes, expressions et définitions adoptés) ».

Textes européens

Règlements, directives, conventions, traités, chartes

« Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit », 2024.

« Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828 (règlement sur l'intelligence artificielle) Texte présentant de l'intérêt pour l'EEE. », 2024.

« Règlement (UE) 2023/1230 du Parlement européen et du Conseil du 14 juin 2023 sur les machines, abrogeant la directive 2006/42/CE du Parlement européen et du Conseil et la directive 73/361/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2023. [↗](#)

« Règlement (UE) 2022/2554 du Parlement européen et du Conseil du 14 décembre 2022 sur la résilience opérationnelle numérique du secteur financier et modifiant les règlements (CE) no 1060/2009, (UE) no 648/2012, (UE) no 600/2014, (UE) no 909/2014 et (UE) 2016/1011 (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. [↗](#)

« Directive (UE) 2022/2555 du Parlement européen et du Conseil du 14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement (UE) no 910/2014 et la directive (UE) 2018/1972, et abrogeant la directive (UE) 2016/1148 (directive SRI 2) (Texte présentant de l'intérêt pour l'EEE) », 2022.

« Règlement (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. [↗](#)

« Règlement (UE) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le règlement (UE) 2018/1724 (règlement sur la gouvernance des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2022. [↗](#)

Proposition de directive du Parlement européen et du Conseil à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle (Directive sur la responsabilité en matière d'IA), 2022. [↗](#)

Proposition de Parlement européen et du Conseil relative à la responsabilité du fait des produits défectueux, 2022. [↗](#)

« Directive (UE) 2019/1937 du Parlement européen et du Conseil du 23 octobre 2019 sur la protection des personnes qui signalent des violations du droit de l'Union », OJ L, 2019. [↗](#)

« Règlement (UE) 2019/1020 du Parlement européen et du Conseil du 20 juin 2019 sur la surveillance du marché et la conformité des produits, et modifiant la directive 2004/42/CE et les règlements (CE) n° 765/2008 et (UE) n° 305/2011 (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. [↗](#)

« Règlement (UE) 2019/881 du Parlement européen et du Conseil du 17 avril 2019 relatif à l'ENISA (Agence de l'Union européenne pour la cybersécurité) et à la certification de cybersécurité des technologies de l'information et des communications, et abrogeant le règlement (UE) no 526/2013 (règlement sur la cybersécurité) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2019. [↗](#)

« Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives

96/9/CE et 2001/29/CE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2019. ↗

« Règlement (UE) 2018/1807 du Parlement européen et du Conseil du 14 novembre 2018 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2018. ↗

« Règlement (UE) 2018/1725 du Parlement européen et du Conseil du 23 octobre 2018 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les institutions, organes et organismes de l'Union et à la libre circulation de ces données, et abrogeant le règlement (CE) n° 45/2001 et la décision n° 1247/2002/CE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2018. ↗

« Règlement (UE) 2017/745 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) n° 178/2002 et le règlement (CE) n° 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017. ↗

« Règlement (UE) 2017/746 du Parlement européen et du Conseil du 5 avril 2017 relatif aux dispositifs médicaux de diagnostic in vitro et abrogeant la directive 98/79/CE et la décision 2010/227/UE de la Commission (Texte présentant de l'intérêt pour l'EEE.) », OJ L, 2017. ↗

« Directive (UE) 2016/943 du Parlement européen et du Conseil du 8 juin 2016 sur la protection des savoir-faire et des informations commerciales non divulgués (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. ↗

« Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. ↗

« Directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel par les autorités compétentes à des fins de prévention et de détection des infractions pénales, d'enquêtes et de poursuites en la matière ou d'exécution de sanctions pénales, et à la libre circulation de ces données, et abrogeant la décision-cadre 2008/977/JAI du Conseil », OJ L, 2016. ↗

« Règlement (UE) 2016/425 du Parlement européen et du Conseil du 9 mars 2016 relatif aux équipements de protection individuelle et abrogeant la directive 89/686/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2016. ↗

« Règlement (UE) n ° 910/2014 du Parlement européen et du Conseil du 23 juillet 2014 sur l'identification électronique et les services de confiance pour les transactions électroniques au sein du marché intérieur et abrogeant la directive 1999/93/CE », OJ L, 2014. ↗

« Directive 2014/68/UE du Parlement européen et du Conseil du 15 mai 2014 relative à l'harmonisation des législations des États membres concernant la mise à disposition sur le marché des équipements sous pression (refonte) Texte présentant de l'intérêt pour l'EEE », OJ L, 2014. ↗

« Règlement (UE) n ° 167/2013 du Parlement européen et du Conseil du 5 février 2013 relatif

à la réception et à la surveillance du marché des véhicules agricoles et forestiers. Texte présentant de l'intérêt pour l'EEE », OJ L, 2013. ☞

« Règlement (UE) n ° 168/2013 du Parlement européen et du Conseil du 15 janvier 2013 relatif à la réception et à la surveillance du marché des véhicules à deux ou trois roues et des quadricycles. Texte présentant de l'intérêt pour l'EEE », OJ L, 2013. ☞

« Règlement (UE) n ° 1025/2012 du Parlement européen et du Conseil du 25 octobre 2012 relatif à la normalisation européenne, modifiant les directives 89/686/CEE et 93/15/CEE du Conseil ainsi que les directives 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE et 2009/105/CE du Parlement européen et du Conseil et abrogeant la décision 87/95/CEE du Conseil et la décision n ° 1673/2006/CE du Parlement européen et du Conseil Texte présentant de l'intérêt pour l'EEE », OJ L, 2012. ☞

« Charte des droits fondamentaux de l'Union européenne », OJ C, 2012. ☞

« Traité sur l'Union européenne (version consolidée) », OJ C, 2012. ☞

« Traité sur le fonctionnement de l'Union européenne (version consolidée) », 2012. ☞

« Règlement (UE) no 182/2011 du Parlement Européen et du Conseil du 16 février 2011 établissant les règles et principes généraux relatifs aux modalités de contrôle par les États membres de l'exercice des compétences d'exécution par la Commission », OJ L, 2011. ☞

« Règlement (CE) n o 765/2008 du Parlement européen et du Conseil du 9 juillet 2008 fixant les prescriptions relatives à l'accréditation et à la surveillance du marché pour la commercialisation des produits et abrogeant le règlement (CEE) n o 339/93 du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008. ☞

« Règlement (CE) n° 765/2008 du Parlement européen et du Conseil du 9 juillet 2008 fixant les prescriptions relatives à l'accréditation et à la surveillance du marché pour la commercialisation des produits et abrogeant le règlement (CEE) n° 339/93 du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008. ☞

« Règlement (CE) n° 300/2008 du Parlement européen et du Conseil du 11 mars 2008 relatif à l'instauration de règles communes dans le domaine de la sûreté de l'aviation civile et abrogeant le règlement (CE) n° 2320/2002 (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008. ☞

« Directive 2006/54/CE du Parlement européen et du Conseil du 5 juillet 2006 relative à la mise en œuvre du principe de l'égalité des chances et de l'égalité de traitement entre hommes et femmes en matière d'emploi et de travail (refonte) », 204, 2006. ☞

« Directive 2006/42/CE du Parlement européen et du Conseil du 17 mai 2006 relative aux machines et modifiant la directive 95/16/CE (refonte) (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2006. ☞

« Directive 2004/113/CE du Conseil du 13 décembre 2004 mettant en œuvre le principe de l'égalité de traitement entre les femmes et les hommes dans l'accès ` des biens et services et la fourniture de biens et services », 373, 2004. ☞

« Règlement (CE) n° 1049/2001 du Parlement européen et du Conseil du 30 mai 2001 relatif à l'accès du public aux documents du Parlement européen, du Conseil et de la Commission », OJ

L, 2001. ↗

« Directive 2000/78/CE du Conseil du 27 novembre 2000 portant création d'un cadre général en faveur de l'égalité de traitement en matière d'emploi et de travail », 303, 2000. ↗

« Directive 2000/43/CE du Conseil du 29 juin 2000 relative à la mise en oeuvre du principe de l'égalité de traitement entre les personnes sans distinction de race ou d'origine ethnique », 180, 2000. ↗

« Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données », OJ L, 1995. ↗

« Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux », 210, 1985. ↗

CONSEIL DE L'EUROPE, « Convention européenne des droits de l'Homme », 1950.

Résolutions, requêtes, communications, décisions

COMMISSION EUROPÉENNE, « Draft standardisation request to the European Standardisation Organisations in support of safe and trustworthy artificial intelligence », 5 octobre 2022.

COMMISSION EUROPÉENNE, Proposition de règlement du Parlement européen et du Conseil modifiant le règlement (UE) n° 910/2014 en ce qui concerne l'établissement d'un cadre européen relatif à une identité numérique, 3 juin 2021, 2021/0136.

PARLEMENT EUROPÉEN, « Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes », 2020. ↗

PARLEMENT EUROPÉEN, « Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle », 2020. ↗

PARLEMENT EUROPÉEN, « Résolution du Parlement européen du 20 octobre 2020 sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle », 2020. ↗

Communication de la Commission européenne au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions - Renforcer la confiance dans l'intelligence artificielle axée sur le facteur humain, 2019. ↗

Communication de la Commission européenne au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions - L'intelligence artificielle pour l'Europe, 2018. ↗

Avenir de la normalisation européenne Résolution du Parlement européen du 21 octobre 2010

sur l'avenir de la normalisation européenne (2010/2051(INI)), 2010. [↗](#)

« Décision n° 768/2008/CE du Parlement européen et du Conseil du 9 juillet 2008 relative à un cadre commun pour la commercialisation des produits et abrogeant la décision 93/465/CEE du Conseil (Texte présentant de l'intérêt pour l'EEE) », OJ L, 2008. [↗](#)

Textes internationaux

« Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence », 2023. [↗](#)

Loi sur l'intelligence artificielle (Corée du sud), 14 février 2023.

Loi sur l'intelligence artificielle et les données, 16 juin 2022, C-27. [↗](#)

Algorithmic Accountability Act, 3 février 2022, H.R.6580. [↗](#)

« The Toronto Declaration : Protecting the right to equality and non-discrimination in machine learning systems », 2018. [↗](#)

« Déclaration de Montréal pour un développement responsable de l'IA ». [↗](#)

« Directive sur la prise de décision automatisée (Canada) », 1^{er} avril 2019. [↗](#)

Projeto de lei n°2338 dispõe sobre o uso da Inteligência Artificial (Projet de loi relatif à l'utilisation de l'IA), 3 mai 2023, 2338.

Projeto de lei n° 21/2020 estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências (Projet de loi n° 21/2020 établissant des principes, des droits et des devoirs pour l'utilisation de l'intelligence artificielle au Brésil et prévoit d'autres dispositions), 11 février 2020, 21/2020.

Rapports, études, travaux et avis

ADMINISTRATIVE COOPERATION GROUPS, « Bonnes pratiques pour la surveillance du marché », Administrative Cooperation Groups, 2017.

AUTOMATED DECISION SYSTEMS TASK FORCE, « New York City Automated Decision Systems Task Force Report », New York, NY, 2019. [↗](#)

BERTOLINI Andrea, « Artificial Intelligence and Civil Liability », Parlement Européen, 2020.

BITAR Omar, DESHAIES Benoit et HALL Dawn, « 3rd Review of the Treasury Board Directive on Automated Decision-Making », 2022. [↗](#)

COMMISSION DES AFFAIRES JURIDIQUES, « Rapport contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle », 2020. [↗](#)

COMMISSION EUROPÉENNE, « Le « Guide bleu » relatif à la mise en œuvre de la réglementation de l'UE sur les produits 2022 », 2022. [↗](#)

COMMISSION EUROPÉENNE, « Le « Guide bleu » relatif à la mise en œuvre de la réglementation de l'UE sur les produits 2016 (2016/C 272/01) », 2016.

COMMISSION EUROPÉENNE, « Livre blanc sur l'intelligence artificielle : Une approche européenne axée sur l'excellence et la confiance », Bruxelles, Commission européenne, 2020.

COMMISSION EUROPÉENNE, « Impact Assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative Acts », Brussels, 2021.

COMMISSION EUROPÉENNE POUR L'EFFICACITÉ DE LA JUSTICE (CEPEJ), « Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement », Strasbourg, Conseil de l'Europe, 2018. ↗

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Le guide du délégué à la protection des données | CNIL », 2022. ↗

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Comment permettre à l'Homme de garder la main ? Les enjeux éthiques de l'intelligence artificielle. », CNIL, 2017.

CONSEIL D'ETAT, « S'engager dans l'intelligence artificielle pour un meilleur service public », sur *Conseil d'État*, publié le 30 août 2022. ↗

CONSEIL D'ETAT, « Le juge administratif et les autorités de régulation économique - Dossier thématique », 2016. ↗

CONSEIL D'ETAT, « Le juge administratif et le droit des médicaments - Dossier thématique », Conseil d'Etat, 2014.

CONSEIL D'ETAT, « Les autorités administratives indépendantes : évaluation d'un objet juridique non identifié », Conseil d'Etat, 2006.

CONTROLEUR EUROPEEN DE LA PROTECTION DES DONNEES (CEPD), « Lignes directrices relatives à la prise de décision individuelle automatisée et au profilage aux fins du règlement (UE) 2016/679 », 2018. ↗

COUR DE CASSATION, *La preuve dans la jurisprudence de la Cour de cassation*, Paris, la Documentation française, 2013, 614 p.

DÉFENSEUR DES DROITS et COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS, « Algorithmes : prévenir l'automatisation des discriminations », Défenseur des Droits, 2020.

DINGEMANN Kathrin et KOTTMANN Matthias, « Legal Opinion On the European System of Harmonised Standards », Berlin, German Federal Ministry for Economic Affairs and Energy, 2020.

DUBNICK Melvin et FREDERICKSON H. George, « Public Accountability: Performance Measurement, the Extended State, and the Search for Trust », Rochester, NY, National Academy of Public Administration et Kettering Foundation, 2011.

EIM BUSINESS & POLICY RESEARCH, « Access to Standardisation - Study for the European Commission », Enterprise and Industry [DG], 2010.

ENA, « Ethique et responsabilité des algorithmes publics », Etalab, 2019.

EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES, « Vision of the European Forensic Science Area - 2030 ».

FOOD AND DRUG ADMINISTRATION, « Glossary of Computer System Software Development Terminology (8/95) », Office of regulatory affairs - FDA, 2019. [↗](#)

FREDMAN Sandra, « Intersectional discrimination in EU gender equality and non-discrimination law », Bruxelles, Commission européenne, 2016.

GROUPE D'EXPERTS INDÉPENDANTS DE HAUT NIVEAU SUR L'INTELLIGENCE ARTIFICIELLE, « Lignes directrices en matière d'éthique pour une IA digne de confiance », Commission européenne, 2019.

IEEE SA, « The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems », sur *IEEE Standards Association*. [↗](#)

PY Bruno, « L'utilisation des caractéristiques génétiques dans les procédures judiciaires - Etude de dix années de pratiques en Meurthe-et-Moselle (2003-2013) », Mission de recherche Droit et Justice, 2017.

REISMAN Dillon, SCHULTZ Jason, CRAWFORD Kate *et al.*, « Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability », 2018.

Sites internet

Articles de presse

BERTUZZI Luca, « Commission leaves European standardisation body out of AI standard-setting », sur *Www.euractiv.com*, publié le 7 décembre 2022. [↗](#)

BERTUZZI Luca, « European Commission sets out a plan to regain clout in standard-setting », sur *Www.euractiv.com*, publié le 3 février 2022. [↗](#)

COMMISSION EUROPEENNE, « High-level expert group on artificial intelligence | Shaping Europe's digital future », publié le 6 février 2023. [↗](#)

CROCHET-DAMAIS Antoine, « Algorithme génétique : définition et apport en machine learning », sur *Journal du net*, publié le 19 juillet 2022. [↗](#)

COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, « Création d'un service de l'intelligence artificielle à la CNIL et lancement des travaux sur les bases de données d'apprentissage | CNIL ». [↗](#)

CONTEXTE, « Document - Les corapporteurs revoient la classification des systèmes à haut risque - Contexte », publié le 26 janvier 2023. [↗](#)

DILMEGANI Cem, « Handwriting Recognition in 2023: In-depth Guide ». [↗](#)

EMAIL et PRINT, « Traque à la piscine non déclarée : quand l'IA prend l'eau », sur *ZDNet France*, publié le 5 avril 2022. [↗](#)

FABIENSOYEZ, « Le porno deepfake explose, et c'est un vrai problème », sur *CNET France*, publié le 5 novembre 2022. [↗](#)

GIEBLER Sebastian, SPIELKAMP Matthias, « 'Trustworthy AI' is not an appropriate framework », sur *AlgorithmWatch*. [↗](#)

GUILLAUD Hubert, « La reconnaissance faciale en ses limites », sur *Hubert Guillaud*, publié le 21 novembre 2023. [↗](#)

JULIA ANGWIN Jeff Larson, « Machine Bias », sur *ProPublica*, ProPublica, publié le 23 mai 2016. [↗https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing](https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing)

KUN Jeremy, « Big data algorithms can discriminate, and it's not clear what to do about it », sur *The Conversation*. [↗](#)

L Bastien, « Test de Turing - Un test pour mesurer l'intelligence artificielle », sur *INTELLIGENCE-ARTIFICIELLE.COM*, publié le 5 juin 2024. [↗](#)

L'ETUDIANT, « Classement des meilleures villes étudiantes 2023-2024 - L'Etudiant ». [↗](#)

LAUSSON Julien, « Fraude fiscale : ça y est, la surveillance des réseaux sociaux par le fisc se met en place », sur *Numerama*, publié le 18 février 2021. [↗](#)

LECHER Colin, « New York City's algorithm task force is fracturing », sur *The Verge*, publié le 15 avril 2019. [↗](#)

LÉGIPRESSE, « Dieudonné condamné pour injure à caractère racial : les propos lui sont imputables, bien qu'il invoque avoir été victime d'un « deepfake » », octobre 2022, n° 407, p. 529.

LES ECHOS, « Le fisc généralise la détection des piscines non déclarées par intelligence artificielle », sur *Les Echos*, publié le 29 août 2022. [↗](#)

MATTU Jeff Larson Julia Angwin, Lauren Kirchner, Surya, « How We Analyzed the COMPAS Recidivism Algorithm », sur *ProPublica*. [↗](#)

METZ Cade et WAKABAYASHI Daisuke, « Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I. », *The New York Times*, 3 décembre 2020. [↗](#)

O'FLAHERTY Kate, « YouTube keeps deleting evidence of Syrian chemical weapon attacks », *Wired*. [↗](#)

PETITJEAN Sophie, « ChatGPT, l'invité surprise des négociations sur le règlement pour l'IA

- Contexte », publié le 10 février 2023. [↗](#)

PINAULT Laure, « TOP 10 des villes où il fait bon vivre en France », sur *Paradisimmobilier*. [↗](#)

PIPER Kelsey, « Exclusive: Google cancels AI ethics board in response to outcry », sur *Vox*, publié le 4 avril 2019. [↗](#)

PREVENTION INTERNET ET RESEAUX SOCIAUX, « Le cyberharcèlement », sur *Prévention internet et réseaux sociaux*. [↗](#)

RUSSEL Stuart, PERSET Karine, GROBELNIK Marco, « Updates to the OECD's definition of an AI system explained ». [↗](#)

SCASSA Teresa, « Canada's Proposed AI & Data Act - Purpose and Application », publié le 8 août 2022. [↗](#)

SHEAD Sam, « Facebook Reportedly Has A Dedicated AI Ethics Team », sur *Forbes*. [↗](#)

THE INTERNET SOCIETY, « Artificial Intelligence & Machine Learning: Policy Paper », sur *Internet Society*, publié le 18 avril 2017. [↗](#)

WALSH Philip, « Do Algorithms Have Politics? », sur *Blog Thinking Tech*, publié le 12 février 2021. [↗](#)

Autres sites internet

AFFAIRS Office of Regulatory, « Glossary of Computer System Software Development Terminology (8/95) », *FDA*, FDA, juin 2019. [↗](#)

ALGORITHM WATCH, « AI Ethics Guidelines Global Inventory », sur *AI Ethics Guidelines Global Inventory*. [↗](#)

AMAZON WEB SERVICES, « Vue d'ensemble de la détection des visages et de la comparaison faciale - Amazon Rekognition ». [↗](#)

CEN-CENELEC, « CEN and CENELEC launched a new Joint TC on Artificial Intelligence », sur *CEN-CENELEC*, publié le 3 mars 2021. [↗](#)

CEN-CENELEC, « The CEN and CENELEC Annual Report 2021 is out! », sur *CEN-CENELEC*. [↗](#)

CEN-CENELEC, « Reference documents », sur *BOSS CEN*. [↗](#)

COMMISSION EUROPENNE, « -ARTIFICIAL INTELLIGENCE | Joinup », 2024. [↗](#)

COMMISSION EUROPÉENNE, « List of national market surveillance authorities by country », publié le 3 février 2023. [↗](#)

COMMISSION EUROPENNE, « Commission appoints expert group on AI and launches the European AI Alliance | Shaping Europe's digital future », publié le 14 juin 2018. [↗](#)

IEC, « IEC - CENELEC Agreement on Common planning of new work and parallel voting

(Dresden Agreement) - 1996 | IEC ». [↗](#)

INRIA, « Le projet-pilote REGALIA au service de la régulation des algorithmes | Inria », publié le 8 décembre 2020. [↗](#)

LABORATOIRE D'INNOVATION NUMERIQUE DE LA CNIL, « [Algaudit] 1 - Choisir une solution d'audit algorithmique | Linc ». [↗](#)

LAROUSSE Éditions, « Définitions : représentativité - Dictionnaire de français Larousse ». [↗](#)

LAROUSSE Éditions, « Définitions : métrologie - Dictionnaire de français Larousse ». [↗](#)

LAROUSSE Éditions, « Définitions : discrétisation - Dictionnaire de français Larousse ». [↗](#)

MINISTERE DE L'ECONOMIE, DES FINANCES ET DE LA SOUVERAINETE INDUSTRIELLE ET NUMERIQUE, « Le marquage « CE » | entreprises.gouv.fr ». [↗](#)

ORGANISATION MONDIALE DE LA SANTE, « Obésité et surpoids ». [↗](#)

SHANGHAI RANKING, « ShanghaiRanking's Academic Ranking of World Universities ». [↗](#)

« Loi + preuve dans 10, 20, 30 ans.... », 2020. [↗](#)

« Toolbox : Dynamics of AI Principles », sur *AI ETHICS LAB*, publié le 25 février 2020. [↗](#)

« Log : définition simple et utilisation », publié le 9 janvier 2019. [↗](#)

« RÉSULTAT : Définition de RÉSULTAT ». [↗](#)

« PROBANT : Définition de PROBANT ». [↗](#)

« PRINCIPE : Définition de PRINCIPE ». [↗](#)

« Preuve - Définitions, synonymes, prononciation, exemples | Dico en ligne Le Robert ». [↗](#)

« Outil PIA : téléchargez et installez le logiciel de la CNIL | CNIL ». [↗](#)

« MÉTROLOGIE : Définition de MÉTROLOGIE ». [↗](#)

« Log_1 noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com ». [↗](#)

« Inférence - Définitions, synonymes, conjugaison, exemples | Dico en ligne Le Robert ». [↗](#)

« INDICATEUR : Définition de INDICATEUR ». [↗](#)

« Home | Starlight ». [↗](#)

« Fichier LOG : définition, ouverture et utilisation | Adobe ». [↗](#)

« European Standards », sur *CEN-CENELEC*. [↗](#)

« ÉTALONNAGE : Définition de ÉTALONNAGE ». [↗](#)

« Apprentissage par renforcement », sur *CNIL*. [↗](#)

« Apprentissage automatique », sur *Cnil*. [↗](#)

« Algorithme ». [↗](#)

« Planetoscope - Statistiques : Informations publiées dans le monde sur le net (en Gigaoctets) ». [↗](#)

« DÉTERMINER : Définition de DÉTERMINER ». [↗](#)

INDEX

Les numéros renvoient aux numéros des pages

A

Algorithme, 13, 38, 55, 64, 78, 86-87, 174

Analyse d'impact, 47

Protection des données, 38, 118, 412

Règlement sur l'IA, 113, 315

Apprentissage automatique, 5, 7, 12, 38, 54, 129, 300, 333, 473, 476

Approche par les risques, 111, 118

Autonomie, 48, 104, 106, 334

Autorité de protection des droits fondamentaux, 261, 265, 267

Autorité de surveillance du marché, 256, 258-260, 271

B

Biais discriminatoires, 128, 164, 172-176, 179, 185, 187, 230, 443, 462, 487

C

Certification, 239

Externe, 231, 235-236

Interne, 226, 228

CNIL, 272, 273-279

Code source, 69, 249

COMPAS, 178, 183, 417, 443

Contrôle humain, 48-49, 76, 78

D

Définition du système d'IA

Commission européenne, 107

Droit canadien, 10, 108

Droit étatsunien, 10, 109

OCDE, 10, 109

Règlement sur l'IA, 9, 108, 110

Demi-preuve, 470-474, 477

Déploieurs

Obligations, 132, 143, 145, 255, 315, 421

Directive européenne sur la responsabilité en matière d'IA (proposition), 303, 305, 355, 357, 359, 361, 364, 391

Directive européenne sur les produits défectueux (proposition), 21, 25, 343, 358-359, 364

Directive sur la prise de
décision automatisée
(Canada), 67-69, 74, 80-81,
83, 93, 118-119

Discrimination
Algorithmique, 21, 164-165, 321,
326
Notion, 176, 180, 321, 325

Disruptivité juridique
Intelligence artificielle, 104-105
Théorie, 99-100, 103

Divulgaration, 51, 53, 355, 359,
361, 365

Droit à la liberté d'expression,
319

Droit à la non-discrimination,
16, 59, 180, 327, 366, 438,
443

Droits fondamentaux, 147, 208,
261, 268, 313, 315, 316, 320,
326

E

Équité, 37, 176-179

Exactitude, 41, 55, 134, 161,
172, 179, 189, 296, 311, 323,
325, 327, 369, 395, 423, 435-
436, 438, 484, 486-487

Expert
De partie, 451-452
Judiciaire, 448-449, 451-456

Explicabilité, 45, 90

F

Fiabilité

Garantie, 432-433
Preuve, 385-386, 394, 422
Système d'IA, 40-41, 436-437, 454,
473

Force probante, 385, 428, 429,
431-432, 434-435, 438, 447,
468-470, 473-477

Fournisseurs

Obligations, 113, 127, 130-131,
133, 137, 143-144, 151, 175,
226, 228, 245-246, 248-249,
337, 341, 344, 347
Responsabilité, 293-294, 297-298,
303, 305, 308, 319, 323-324,
327

I

Indicateur, 12, 151, 163, 169-
170, 172-174, 177, 179, 181,
223, 230, 252, 276, 278, 296-
297, 339, 476, 487

Intervention humaine, 78, 92

J

Jeux de données, 26, 128-130,
149, 151, 323

Juge

Administratif, 273-274, 276-278
Judiciaire, 204, 209, 212, 216

L

Légitimité démocratique, 195,
197, 199
Licéité de la preuve, 398, 401,
403, 411-412
Loi informatique et libertés, 76,
413

N

Normes techniques
Normes européennes
harmonisées, 140-141, 194,
203-204, 210, 212-213, 240,
242, 339, 347
Spécifications communes, 208,
217-218, 242
Nouvelle approche, 137, 139,
150, 162, 196, 207

O

Opacité, 43-44, 72, 105, 417,
419, 474-475
Organisme notifié, 232, 234-
235, 241, 243, 246, 249-250,
269
Organismes de normalisation,
199-201, 224, 242, 295, 461
Européens, 138, 192, 196, 223,
246, 456
Internationaux, 197
Nationaux, 138

P

Parcoursup, 68
Présomption, 357, 358, 388,
389, 390, 391, 392
Preuve
Charge, 117, 142, 241, 324
Mode, 378, 388, 411
Objet, 169, 177, 179, 230, 301,
389, 396, 440, 442, 445, 446,
447

R

Reconnaissance faciale, 399,
409, 432, 476
Règlement européen sur les
services numériques, 69, 71,
210, 316, 350
Règlement général à la
protection des données
(RGPD), 38, 66-67, 77, 81,
86-88, 92, 165, 349, 350
Représentativité
Organisme, 193
Données, 151, 249, 323
Responsabilité
Civile, 289, 291
Contractuelle, 293, 295-296, 302
Délictuelle, 359
Ex ante, 142, 313
Ex post, 321, 337, 339, 352, 496
Extracontractuelle, 303-304
Principe éthique, 43, 46, 60

Produits défectueux, 102, 306-308, 310, 347, 358

Risques systémiques, 123-124, 210-211, 317

Robustesse, 41, 134, 150, 369

S

Seuil statistique, 182-184, 188, 190, 231, 236, 276, 278, 301, 485, 486, 487

Systèmes d'IA à usage général, 121-124

T

TAJ, 412

Traitement des antécédents judiciaires (TAJ), 399, 409

Transparence, 43-45, 51, 53, 55-56, 59, 64, 68-71, 88, 90, 91, 132, 150, 189, 341, 419, 420

U

Utilité probatoire, 363, 369-370, 457

TABLE DES MATIERES

REMERCIEMENTS.....	V
SOMMAIRE.....	VII
LISTE DES ABREVIATIONS.....	IX
INTRODUCTION.....	1
I. L'explication de l'IA.....	4
II. Les enjeux normatifs de l'encadrement des systèmes d'IA.....	11
III. La preuve juridique comme outil d'analyse du droit des systèmes d'IA	17
PARTIE 1 : LES PREUVES RELATIVES AUX EXIGENCES DE CONCEPTION DES SYSTEMES D'IA.....	29
Titre I : L'apparition d'un droit spécifique aux systèmes d'IA.....	31
Chapitre 1 : Les normes applicables aux technologies d'IA	33
Section I : Les principes éthiques : sources d'inspiration du droit de l'IA	34
Sous-section 1. L'apport d'un consensus de principes de l'IA.....	34
§1. Les principes-objectifs	36
A– Le principe de fairness	36
B– Le principe de vie privée.....	38
C– Le principe de fiabilité technique	40
§2. Les principes-outils	42
A– Le principe de transparence	43
B– Le principe de responsabilité.....	46
C– Le principe de contrôle humain	48
Sous-section 2. Les limites à l'apport des textes éthiques sur l'IA.....	50
§1. Les limites sur le fond des principes éthiques d'IA.....	50
A– Le manque de clarté des principes.....	51
B– La présence de conflits entre les principes.....	54
§2. Les limites liées à la nature des principes éthiques.....	57
A– L'appropriation des principes éthiques par les acteurs.....	57
B– Le faible niveau de contrainte des textes éthiques.....	61
Conclusion de la section I :.....	63

Section II : Les besoins du droit positif relatif à l'utilisation d'outils algorithmiques ..	63
Sous-section 1 : Un besoin d'articulation avec le droit positif	64
§1. Les mécanismes juridiques de la transparence	64
A – Les formes de la transparence	65
1) L'obligation d'information individuelle.	65
2) La communication d'informations sur demande	67
3) La publication d'informations	68
B – Le contenu de la transparence	70
1) L'information d'utilisation d'un algorithme	70
2) L'information sur le fonctionnement de l'algorithme	72
§2. Les mécanismes juridiques de contrôle humain sur la décision produite	76
A – Le contrôle humain pendant la prise de décision	76
B – Le contrôle humain après la prise de décision.....	78
Sous-section 2 : Un besoin de dépasser les limites du droit positif	79
§1. Les limites relatives à la portée des dispositions juridiques existantes.....	79
A –La définition de la décision impliquant un algorithme	80
B – Le champ d'application restreint des textes existants.....	82
§2. Les difficultés d'application : l'inadéquation entre le droit et la technique	85
A – L'imprécision technique des termes juridiques	85
1) La qualification juridique complexe des décisions concernées	86
2) La portée incertaine des exigences juridiques.	88
B – L'aggravation des difficultés d'exécutions des obligations dans le contexte des technologies d'IA	90
1) L'opacité problématique des technologies d'IA	90
2) Le contrôle difficile des technologies d'IA.....	91
Conclusion de la section II :	94
Conclusion du chapitre 1 :	95
Chapitre 2 : L'adoption de normes juridiques spécifiques aux systèmes d'IA	97
Section I : Des règles juridiques centrées sur les systèmes d'IA à haut risque	98
Sous-section 1 : Le système d'IA comme objet de réglementation	98
§1. Le choix de réglementer spécifiquement les technologies d'IA.....	99
A- La justification de l'encadrement spécifique d'une technologie	99
B- La justification de l'encadrement spécifique de l'IA.....	104
§2. La définition du système d'IA comme objet de droit	106

A- Une définition juridique bipartite des systèmes d'IA.	107
B- Le changement pour une définition générale encore à préciser	108
Sous-section 2 : Le niveau de risque comme critère majeur d'application	111
§1. La qualification du système d'IA à risque élevé.....	112
A- Les critères de qualification du système d'IA à risque élevé	112
B- L'exception de l'absence de risques significatifs	115
§2. La place du risque élevé en droit des décisions impliquant un algorithme	117
Conclusion de la section I :	125
Section II : Une conformité juridique ex ante pour les systèmes d'IA à haut risque ..	125
Sous-section 1 : Les exigences essentielles relatives aux systèmes d'IA.....	126
§1. Les obligations juridiques posées pour les systèmes d'IA à haut risque	126
A - Les exigences substantielles pour un système d'IA conforme.	127
B - Les exigences processuelles pour un système d'IA conforme.	135
§2. La concentration sur les exigences essentielles : le recours à la nouvelle approche	137
Sous-section 2 : La responsabilisation des acteurs lors de la phase de conception .	142
§1. La mise en place d'une logique de redevabilité.....	142
A- Les obligations détaillées de documentation à la charge du fournisseur ..	142
B- Le passage à une redevabilité juridique ex ante.	145
§2. La délégation des objectifs de qualité aux normes techniques	147
A- Des nouvelles exigences de qualité des systèmes d'IA imprécises.	147
B- La place des normes techniques dans le règlement européen sur l'IA	150
Conclusion de la section II :	152
Conclusion du chapitre 2 :	153
Conclusion du titre 1 :	155
Titre II : La place de la preuve dans les exigences de conception d'un système d'IA	157
Chapitre 1 : L'élaboration des preuves de conception	159
Section I : L'importance des preuves au sein des normes techniques	160
Sous-section 1 : La place accordée aux preuves techniques dans le règlement.....	161
§1. La reconnaissance partielle de l'importance de preuves techniques adéquates	161
§2. La contrepartie juridique à l'obtention des preuves techniques	163
Sous-section 2 : La place concrète des preuves techniques au sein du règlement ...	168
§1. La détermination de l'objet de la preuve : l'indicateur	169

A- Le choix de l'indicateur	169
B- Le choix d'un indicateur adéquat à l'exigence légale.....	173
§2. Le niveau de qualité acceptable : le seuil statistique	181
A- Le rôle du seuil statistique dans le cadre juridique des systèmes d'IA.....	181
1) Le seuil statistique en droit	182
2) Le seuil statistique dans le règlement sur l'IA	184
B- L'équilibre entre les seuils ou la hiérarchisation des exigences de qualité 185	
1) La hiérarchisation des seuils intra-exigences.....	185
2) La hiérarchisation des seuils inter-exigences.....	188
Conclusion de la section I :.....	190
Section II : L'influence des acteurs dans la définition des normes techniques	191
Sous-section 1 : Les acteurs en charge de l'élaboration des normes techniques	192
§1. Les organismes de normalisation officiels de l'Union européenne	192
§2. Les organismes de normalisation non officiels de l'Union européenne	197
Sous-section 2 : Les acteurs en charge du contrôle des normes techniques.....	202
§1. L'importance juridique des contrôles.....	202
A- Le contrôle de la Commission avant la publication.....	203
B- Le contrôle judiciaire après la publication.	204
§2. Les modalités d'exercice des contrôles	205
A- Le contrôle par la Commission européenne.	206
B - Le contrôle judiciaire par la CJUE	209
1) Le recours direct en annulation auprès de la CJUE	209
2) Le renvoi préjudiciel par les juridictions nationales.....	214
§3. Les conséquences des contrôles.....	215
A - Les conséquences limitées du contrôle judiciaire.	216
B - Les conséquences d'un contrôle par la Commission européenne	217
Conclusion de la section II :.....	218
Conclusion du chapitre 1 :.....	220
Chapitre 2 : Le contrôle des preuves de conception.....	223
Section I : Le contrôle ex ante de la conception du système d'IA	224
Sous-section 1 : Les acteurs de la certification	225
§1. Les acteurs impliqués dans le contrôle <i>ex ante</i>	226
A- Le fournisseur certificateur	226

B- Le tiers certificateur	231
§2. L’auteur choisi pour le contrôle <i>ex ante</i>	238
A- La désignation de l’acteur par la réglementation sectorielle européenne .	239
B- La désignation de l’acteur par le règlement européen sur l’IA	240
Sous-section 2 : Le processus de contrôle <i>ex ante</i> de conformité	244
§1. Le périmètre des procédures de contrôle <i>ex ante</i>	244
§2. L’influence substantielle du contrôle <i>ex ante</i> sur le système probatoire	247
A- Le contrôle des preuves documentaires	248
B- Le contrôle des éléments techniques de preuve	249
Conclusion de la section I :	251
Section II : Le contrôle <i>ex post</i> de la conception du système d’IA	252
Sous-section 1 : Le contrôle administratif de la conception	253
§1. Les acteurs du contrôle administratif	253
A- Les initiateurs du contrôle	254
1. Les opérateurs économiques	254
2. L’autorité administrative	256
B- Les auteurs du contrôle	258
1. Le contrôle par les autorités de surveillance du marché	258
2. L’intervention des autorités de protection des droits fondamentaux	261
§2. La réalisation du contrôle administratif	262
A- La place du contrôle <i>ex post</i> dans les choix des preuves techniques pertinentes	262
B- Les prérogatives octroyées aux acteurs du contrôle <i>ex post</i>	264
1. L’accès aux preuves du fournisseur	264
2. L’obtention proactive de preuves techniques.	266
Sous-section 2 : Le contrôle juridictionnel <i>ex post</i>	268
§1. Les activités faisant l’objet d’un recours juridictionnel	268
A- La décision de l’organisme notifié	269
B- La décision de l’autorité de surveillance du marché	270
§2. Le contrôle juridictionnel de la conformité des systèmes d’IA	272
A- Les acteurs du contrôle juridictionnel	272
B- Le contenu du contrôle juridictionnel	273
1. Les décisions susceptibles de faire l’objet d’un contrôle par le juge administratif	274

2. Le contenu du contrôle exercé par le juge administratif	276
Conclusion de la section II :	279
Conclusion du chapitre 2 :	281
Conclusion du titre 2 :	283
Conclusion de la partie 1 :	285
PARTIE 2 : LES CONSEQUENCES DES PREUVES DE CONCEPTION LORS DE L'UTILISATION DU SYSTEME D'IA.....	287
Titre I : La preuve réutilisée contre un système d'IA : la responsabilité ex post du fait de l'IA	289
Chapitre 1 : L'articulation entre le règlement européen sur l'IA et la responsabilité ex post du fournisseur de systèmes d'IA.....	291
Section I : Les effets de la conformité ex ante du système d'IA sur la responsabilité ex post de son fournisseur.....	292
Sous-section 1 : La défaillance contractuelle du fournisseur du système d'IA	293
§1. La délivrance non-conforme du système d'IA	293
A- La conception conforme du système d'IA délivré.....	293
B- La contractualisation des preuves techniques de conception	295
§2. La présence de vice cachés.....	297
A- Un vice caché.....	298
B- Un vice rendant le système d'IA impropre à l'usage.....	300
Sous-section 2 : Le fait délictuel du fournisseur du système d'IA	302
§1. La faute du fournisseur.....	303
§2. Le défaut du système d'IA	306
A- La notion de défaut	307
B- L'appréciation du défaut	308
Conclusion de la section I :	312
Section II : L'articulation inhabituelle avec les droits et libertés fondamentaux	312
Sous-section 1 : Le lien particulier entre le fonctionnement du système d'IA et les droits et libertés fondamentaux	313
§1. Un lien nouveau reconnu entre les systèmes d'IA et les droits fondamentaux	313
§2. Un lien entre le fonctionnement du système d'IA et l'atteinte aux droits fondamentaux	316
Sous-section 2 : Les interactions entre le fonctionnement du système d'IA à haut risque et les droits et libertés fondamentaux.....	320

§1. La réutilisation des exigences réglementaires de conception du système d'IA à haut risque : l'exemple de la discrimination algorithmique	321
A- La caractérisation d'une différence de traitement	321
B- La justification de la discrimination	324
§2. De l'atteinte aux droits fondamentaux à la conformité du système d'IA.....	326
Conclusion de la section II :.....	328
Conclusion du chapitre 1 :.....	329
Chapitre 2 : La réutilisation des preuves de conception lors d'un dommage causé par un système d'IA	331
Section I : L'apparition de nouvelles preuves issues de la conformité du système d'IA	332
Sous-section 1 : Le besoin d'accès la preuve en cas de dommage causé par un système d'IA	332
§1. Les particularités des systèmes d'IA dans la réalisation du dommage	332
§2. Le besoin d'informations relatives au résultat individuel dommageable.....	335
Sous-section 2 : Les nouveaux éléments de preuve issus du règlement sur l'IA.....	336
§1. Les éléments de preuves relatifs au fonctionnement global du système d'IA.....	337
A- Les informations indirectes sur le fonctionnement d'un système d'IA	337
B- La documentation technique inaccessible au demandeur	340
§2. Les éléments de preuves relatifs au résultat individuel dommageable	343
A- Les journaux d'enregistrement des systèmes d'IA.....	344
B- L'explication de la décision dommageable	348
Conclusion de la section I :.....	353
Section II : Les mécanismes de réutilisation des preuves de conception du système d'IA	354
Sous-section 1 : La mise en avant des preuves de conception par le législateur européen.....	354
§1. Les incitations à collecter préalablement des preuves auprès du défendeur ..	355
A- Une condition à la sollicitation du juge.....	355
B- Une condition pour bénéficier des présomptions probatoires	357
§2. La priorisation de la divulgation des preuves détenues par les acteurs de la conformité	358
A- Les destinataires des mesures prononcées par le juge	359
B- La nature des prérogatives accordées au juge	361
Sous-section 2 : L'évaluation incertaine de l'utilité probatoire par le juge.....	363

§1. La théorie de l'utilité probatoire	364
§2. L'utilité du procédé probatoire	367
Conclusion de la section II :	370
Conclusion du chapitre 2 :	372
Conclusion du titre 1 :	373
Titre II : La preuve produite par un système d'IA	375
Chapitre 1 : L'admissibilité de la preuve produite par un système d'IA	377
Section I : La qualification de la preuve produite par un système d'IA	378
Sous-section 1 : Les catégories de modes de preuves écartées	378
§1. La qualification comme preuve par déclaration logiquement exclue	378
A- La définition des principaux modes de preuves par déclaration	379
B- L'absence rédhibitoire de personnalité juridique du système d'IA	380
§2. La catégorie de l'écrit électronique difficilement envisageable	380
A- L'écrit produit par un système d'IA	381
B- La copie électronique produite par un système d'IA	384
Sous-section 2 : La qualification du résultat d'un système d'IA comme preuve indiciaire	388
§1. Le refus de l'expression « présomption judiciaire »	388
§2. Les caractéristiques matérielles de la présomption de fait	391
§3. Le résultat d'un système d'IA comme indice au service d'une présomption de fait	393
A- La qualification du résultat en tant qu'élément existant	393
B- La qualification du résultat en tant qu'information signifiante	395
Conclusion de la section I :	396
Section II : La licéité de la preuve produite par un système d'IA	397
Sous-section 1 : Le principe de licéité de la preuve	398
§1. Un principe constitué de deux blocs	398
§2. Le champ d'application du principe	402
Sous-section 2 : L'articulation avec les principes généraux du droit de la preuve ..	404
§1. Le droit au respect de la vie privée	407
A- La conciliation entre le droit à la preuve et le droit au respect de la vie privée	408
B- La licéité de la preuve par un système d'IA au regard de l'atteinte au droit au respect de la vie privée	411

§2. Le respect des droits de la défense	415
A- L'opacité des systèmes d'IA face au principe du contradictoire	416
B- La contestation de la fiabilité des preuves par un système d'IA	420
Conclusion de la section II :	424
Conclusion du chapitre 1 :	426
Chapitre 2 : L'appréciation des preuves produites par un système d'IA	427
Section I : L'appréciation libre par le juge	428
Sous-section 1 : L'évaluation de la force probante de la preuve produite par un système d'IA	428
§1. La force probante de l'élément de preuve	429
A- La préexistence du résultat	429
B- La crédibilité de l'élément de preuve	431
§2. La force probante du procédé de preuve	434
A- La conformité aux normes	435
B- La fiabilité du fonctionnement du procédé	436
Sous-section 2 : L'analyse de la pertinence de la preuve par un système d'IA	439
§1. Le rôle de la pertinence dans les preuves statistiques en droit positif	439
A- Le résultat statistique à haute pertinence : le test de paternité en droit de la filiation	439
B- Le résultat statistique à faible pertinence : l'empreinte ADN en droit pénal	441
§2. L'examen de la pertinence du résultat de systèmes d'IA	442
A- L'objet de la preuve statistique	442
B- L'objet de la preuve juridique	445
Conclusion de la section I :	447
Section II : L'expertise et la preuve par un système d'IA	448
Sous-section 1 : Les missions de l'expert	448
§1. La production de la preuve par un système d'IA	449
§2. L'évaluation de la preuve produite par un système d'IA par l'expert	450
A- La contestation de la preuve produite par un système d'IA	451
B- Les rôles de l'expert judiciaire dans l'utilisation des preuves de conception devant le juge	455
Sous-section 2 : La place de l'expert dans l'appréciation du juge	457
§1. Le choix du juge de recourir à un expert judiciaire	457

§2. L'appréciation du rapport d'expertise par le juge.....	462
A- L'influence concrète du rapport d'expertise sur l'appréciation du juge. ..	463
B- Le comportement du juge face à l'expertise relative à la preuve par un système d'IA	464
Conclusion de la section II :	466
Section III : La détermination préalable de règles d'appréciation de la preuve par un système d'IA.....	467
Sous-section 1 : L'exclusion du fonctionnement du système d'IA dans l'appréciation de son résultat.....	468
§1. La fixation de la force probante des modes de preuve en droit positif	468
A- L'augmentation de la force probante	468
B- La réduction de la force probante : la demi-preuve	470
§2. L'élaboration de règles spécifiques d'appréciation des preuves par un système d'IA	472
A- Le résultat généré par un système d'IA réduit à une demi-preuve	473
B- Les conséquences du statut de demi-preuve pour le résultat d'un système d'IA	476
Sous-section 2 : L'imposition du fonctionnement du système d'IA dans l'appréciation de son résultat	478
§1. La qualité de fonctionnement comme critère de détermination de la force probante en droit pénal.....	478
A- L'interprétation des résultats des cinémomètres	479
B- L'interprétation des résultats produits par les éthylomètres.....	481
§2. La qualité de fonctionnement des systèmes d'IA comme critère d'appréciation de la force probante de leurs résultats.....	484
A- L'application aux preuves produites par un système d'IA.....	484
B- Les spécificités des preuves produites par un système d'IA.	487
Conclusion de la section III :	489
Conclusion du chapitre 2 :	490
Conclusion du titre 2 :	491
Conclusion de la partie 2 :	493
CONCLUSION GENERALE	495
BIBLIOGRAPHIE	501
Dictionnaires et encyclopédies.....	501
Ouvrages généraux, manuels et traités	501

Ouvrages généraux.....	501
Thèses.....	503
Ouvrages spécialisés, thèses et mémoires	503
Ouvrages spécialisés	503
Mémoires.....	505
Articles.....	505
Contributions aux mélanges et actes de colloques	517
Notes, observations, avis et conclusions de jurisprudence	517
Jurisprudence française	517
Conseil constitutionnel	517
Conseil d'État.....	517
Cour de cassation	518
Cours d'appel	524
Jurisprudence européenne	524
Jurisprudence de la CEDH.....	524
Jurisprudence de l'Union européenne	525
Jurisprudence étrangère.....	527
Textes législatifs, réglementaires et textes internationaux	527
Textes français.....	527
Lois	527
Ordonnances	528
Décrets.....	528
Arrêtés	529
Circulaires	529
Autres.....	529
Textes européens.....	529
Règlements, directives, conventions, traités, chartes	529
Résolutions, requêtes, communications, décisions	533
Textes internationaux.....	534
Rapports, études, travaux et avis.....	534
Sites internet.....	536
Articles de presse	536
Autres sites internet.....	538

INDEX	541
TABLE DES MATIERES	545

Résumé en français

En 2021, la Commission européenne proposait d'encadrer spécifiquement les systèmes d'IA au sein du marché européen. Adopté et publié en 2024, le règlement européen sur l'IA a pour objectif de protéger la santé, la sécurité et les droits fondamentaux des individus face aux risques de l'IA. Pour cela, le texte énonce plusieurs obligations relatives à la conception de certains systèmes d'IA. Toutefois, l'application d'exigences juridiques vagues à des outils quantitatifs et précis comme les systèmes d'IA représente un défi de taille. Comment s'assurer que les futurs systèmes d'IA respecteront fidèlement les exigences prévues par le droit ? Au cœur de cet enjeu d'application se trouve la preuve juridique. Ainsi, nous nous intéresserons à la preuve afin de comprendre le(s) rôle(s) qu'elle joue au sein du droit des systèmes d'IA. Par le prisme de la preuve, nous proposons une perspective d'analyse originale sur le droit des systèmes d'IA et ses conséquences sur notre système juridique.

Résumé en anglais

In 2021, the European Commission proposed to regulate AI systems within the European market. Adopted and published in 2024, the European AI Act aims to protect individuals' health, safety and fundamental rights in the face of AI risks. To this end, the text sets out several obligations relating to the design of so-called high-risk AI systems. However, applying vague legal requirements to quantitative and precise tools such as AI systems represents a major challenge. How can we ensure that future AI systems will faithfully comply with legal requirements? At the heart of this application challenge lies legal evidence. We will therefore look at evidence to understand the role(s) it plays in the law of AI systems. Through the prism of evidence, we provide an original analytical perspective on the regulation of AI systems and its consequences for our legal system.

Mots-clés : Systèmes d'IA – Preuves – Certification - Règlement sur l'IA