

Censure, plateformes numériques et algorithmes

Guillaume SIRE

Dans les années 1990, et encore dans les années 2000, il était assez rare d'entendre prononcer le mot « algorithme ». C'était un mot quasi inconnu du grand public. Pas une seule fois on ne le trouve prononcé dans le film « Matrix », sorti en salles en 1999, alors qu'il y a fort à parier que si le même film était réalisé en 2021, il y serait répété chaque minute, au point que le film ne s'appellerait peut-être pas « Matrix » mais « Algorithm ». Il y eut un temps où les mots « logiciel », « programmes » et « code » battaient le haut du pavé. C'était d'eux dont on avait peur. Ils investissaient les fantasmes science-fictionnels. Aujourd'hui, ce sont les algorithmes qui sont chargés, avec l'intelligence artificielle, d'assurer à la place des programmes et des logiciels de nourrir dans les imaginaires la part d'ombre du Progrès.

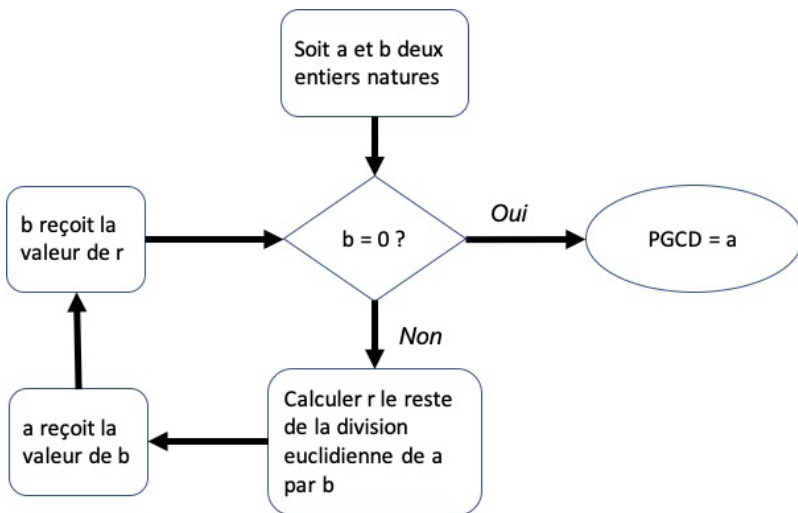
Le mot « algorithme » fait d'autant plus peur qu'il est désormais associé comme épithète à des mots qui l'ont précédé dans l'inquiétude, notamment « censure ». La « censure algorithmique », dit-on. Un ancien mot assorti d'un nouveau, une ancienne menace réinventée. Voilà de quoi nous terrifier. Dans ce chapitre, j'essaierai d'expliquer dans quelle mesure, pourquoi et comment, cette notion peut être comprise. Il ne s'agira pas de prétendre qu'il ne faut pas craindre la censure algorithmique, mais plutôt d'expliquer ce qu'il y a à craindre, et de suggérer le rôle que peuvent jouer les ingénieurs d'une part, et les autorités de régulation d'autre part, dès lors que l'on souhaiterait éviter que des algorithmes puissent censurer.

I. QU'EST-CE QU'UN ALGORITHME ?

Selon une formule célèbre de Robert Kowalski, un algorithme est fait à la fois de logique et de contrôle¹. Très concrètement, un algorithme est une manière d'agencer des variables. « $X + Y =$ », par exemple, est un algorithme. Donnez-le à n'importe quel être humain, et à n'importe quelle machine, et renseignez les variables (*inputs*) de telle sorte que « $X=5$ » et « $Y=4$ » ; et ces êtres humains et ces machines devront logiquement vous donner le même signal de sortie (*output*) : « 9 ». L'algorithme vous aura ainsi permis en vous appuyant sur la logique (et en supposant que les relais humains ou machiniques s'y appuieront également) d'exercer un contrôle sur le signal de sortie, et de réduire à zéro la part d'interprétation par les êtres humains et les machines, si bien que malgré leurs différences, différences de forme, de caractère, malgré les émotions qu'ils ressentent ou non, malgré leurs valeurs, leurs intérêts économiques et leurs projets, ils vous auront tous donné le même résultat.

Un algorithme peut être exprimé littéralement, algébriquement, schématiquement, ou bien grâce à un langage informatique qui permettra aux machines de l'exécuter. Voici par exemple trois expressions absolument équivalentes du célèbre algorithme d'Euclide, l'une en langage schématique, l'autre en langage naturel et la dernière en langage informatique :

Schéma



¹ R. KOWALSKI, « Algorithm = logic + control », *Commun. ACM* 22, 1979, p. 424-436.

Langage naturel

1. Pour trouver le plus grand diviseur commun (PGCD) de deux nombres entiers positifs a et b avec $a > b \geq 0$, on effectue la division euclidienne de a par b et on note r le reste.
2. Si r est strictement supérieur à zéro, alors b devient a et r devient b , puis on recommence : on effectue la division euclidienne de a par b et on note r le reste.
3. On continue ainsi de suite jusqu'à ce qu'une division donne un reste égal à 0. On peut alors en conclure que le PGCD est le dernier reste non nul.

Langage informatique (Javascript)

```
<script>
function pgcd (a,b) {
    document.write(« PGCD(« + a + « , « + b + « ) = « );
    if (b == 0) {
        return a
    }
else {
        r = a % b ;
        return pgcd(b,r) ;
    }
}
</script>
```

À bien y regarder, dans un algorithme, il y a quatre choses : des variables ou « critères », des opérations permettant d'articuler ces critères les uns aux autres, des pondérations desdits critères, et, enfin, des constantes. Observer, et *a fortiori* « critiquer » un algorithme de traitement de l'information comme on critiquerait par exemple la ligne éditoriale d'un journal, revient à identifier les critères, les opérations, les pondérations et les constantes, et à essayer de comprendre ce que ceux-ci impliquent en matière de traitement de l'information, et peut-être de favoritisme ou de censure. Cela revient également à comprendre ce que telle ou telle décision prise par les ingénieurs qui paramètrent l'algorithme, peut avoir comme effet concret, de manière à chercher comment provoquer des effets souhaitables, et comment éviter, ou atténuer, des effets qui le seraient moins.

Que se passe-t-il, par exemple, lorsque je tape sur Google « Rembrandt » ou « art contemporain » ? Un algorithme adjoint environ trois cents variables,

pondérées différemment, pour apparier ma requête à une liste de liens hypertextuels classés par ordre de pertinence supposée. Il s'agit bien d'une éditorialisation ou « méta-éditorialisation », dont le résultat dépend de ce que les ingénieurs ayant écrit l'algorithme considéraient eux-mêmes, ce jour-là, comme étant une définition acceptable de la pertinence – considération qui les aura conduits à effectuer certains arbitrages, lesquels auraient pu être réglés différemment, de même que ces arbitrages ont pu être influencés par les valeurs, les projets, les intérêts économiques, les rapports de pouvoir et les représentations propres aux ingénieurs de Google. Aujourd'hui, l'algorithme du moteur de recherche est incroyablement complexe et opaque, mais en 1998, lorsque le moteur de recherche a vu le jour, il était plus simple à comprendre, et transparent. Voilà comment algébriquement l'algorithme « PageRank » noté « r » est calculé quand une page « A » se présente :

$$r(A) = \frac{\alpha}{N} + (1 - \alpha) \left(\frac{r(B_1)}{|B_1|} + \dots + \frac{r(B_n)}{|B_n|} \right)$$

où $r(B_1)$, $r(B_2)$, $r(B_3)$... $r(B_n)$: les valeurs PageRank des pages B_1 , B_2 ... B_n pointant vers A ;

$|B_1|$ $|B_2|$... $|B_n|$: le nombre de lien présents sur chaque page B_1 , B_2 ,... B_n ;

N : le nombre de pages B_1 , B_2 ,... B_n ;

et α : une constante comprise entre 0 et 1, appelée « facteur d'amortissement ».

Cette formule nous apprend que plus une page effectue de liens, moins chacun des liens sera porteur de pertinence pour la page pointée. D'autre part, plus les liens entrants sont nombreux, plus la pertinence supposée de la page qui les reçoit augmente. Il y a ainsi deux effets principaux : le nombre de liens entrants et l'autorité des sources où se trouvent ces liens. La constante (intitulée « facteur d'amortissement ») a elle aussi un impact significatif sur le PageRank et peut par conséquent modifier de façon substantielle le classement des documents dans le cas où elle viendrait à être modifiée. Elle été fixée à 0,85 par les ingénieurs de Google. En gros, cela signifie que les ingénieurs de Google pensent qu'il y a 85 % de chance que l'internaute, une fois qu'il a cliqué sur un résultat et atterri sur une page web, clique sur un des liens hypertextes qui figureront sur celle-ci. Les ingénieurs pensent autrement dit qu'il faut butiner de page en page pour avoir une information complète, et qu'il est peu probable que toute l'information recherchée par un internaute se trouve sur une page en particulier, ce qui est une vision très discutable, d'autant que le fait de fixer le paramètre alpha à 0,85 défavorise

les contenus visant à l'exhaustivité, et peut agir par conséquent comme une prophétie auto-réalisatrice : parce que les ingénieurs ont pensé que les internautes *voudraient* butiner pour trouver l'information, les internautes *doivent* en effet butiner.

Des exégèses très poussées ont été réalisées du PageRank², et ont montré que l'algorithme prenait racine à la fois dans les traditions de la sociométrie et de la scientométrie. Ces études ont permis de mettre à jour les représentations du monde et les théories « pliées » à l'intérieur de l'algorithme, et participant, de fait, de sa politique éditoriale, en véhiculant certaines conceptions de ce que sont (ou plus exactement « de ce que devraient être ») un document, l'autorité, la pertinence, Internet et le Web.

II. DE LA NÉCESSITÉ D'UNE APPROCHE CRITIQUE ET INTERDISCIPLINAIRE

Laurence Lessig, dans son célèbre article « Code is Law », prévenait déjà en 2000 : « *Si c'est le code qui détermine nos valeurs, ne devons-nous pas intervenir dans le choix de ce code ? Devons-nous nous préoccuper de la manière dont les valeurs émergent ici ?* »³. Si le code charrie en effet certaines valeurs, certains principes, certaines représentations du monde, alors il est nécessaire de pouvoir comprendre comment il agit, et ce même si on n'est pas un informaticien, de même qu'il est nécessaire de nous interroger collectivement sur son rôle, d'une part, et, d'autre part, sur la manière dont il faut, ou non, le réguler. Treize ans après avoir écrit « Code is Law », Lawrence Lessig ne parlait plus de « code » mais d'algorithmes, en prévenant que ceux-ci étaient des « *entités puissantes qui contrôlent, gouvernent, trient, réglementent et façonnent tout, des transactions financières aux médias d'information* »⁴.

Afin d'expliquer comment, du point de vue théorique et méthodologique, les sciences sociales peuvent se pencher sur le rôle joué par les algorithmes vis-à-vis des contenus culturels, je ferai ici rapidement l'exposé des trois postulats fondamentaux de la sociologie des techniques, en partant du principe qu'un algorithme est un artefact technique au même titre qu'une voiture ou qu'un tournevis, et qu'il est possible, dès lors, de trouver dans la sociologie des techniques, très développée depuis quarante ans en France,

² Voir B. RIEDER « What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index », *Computational Culture*, 2012. Voir aussi, D. CARDON, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », *Réseaux*, vol. 1, n° 177, 2013, p. 63-95.

³ L. LESSIG, « Code is law. On Liberty in Cyberspace », *Harvard Magazine*, 2000 (traduction française sur Framablog).

⁴ L. LESSIG, « Governing algorithms NY Conference », 2013 (on line).

des instruments conceptuels permettant d'ouvrir ce qui sans elle demeurerait une boîte noire, et de faire dialoguer différentes disciplines.

Le premier postulat de la sociologie des techniques consiste à prétendre que rien, jamais, n'est purement technique. La technique est *aussi*, et *toujours*, un événement social. Les dispositifs que nous utilisons, quelle que soit leur complexité, sont *sociotechniques*, c'est-à-dire qu'ils sont enchâssés dans un monde politique, économique, idéologique, juridique. Dire cela, c'est éviter le piège du technodéterminisme, *i.e.* la tendance à faire de la technique un Tout extérieur au monde humain, et qui, depuis l'extérieur, l'influencerait. Quand on entend dire par exemple que « le Web nous aidera à faire ceci ou cela », ou bien que « la télévision rendra les enfants idiots » ou bien encore que « la génétique nous sauvera ou nous détruira », il semble que la technique provienne par au-dessus, et agissent sur nous depuis un ciel des idées dont la nature ne dépend pas de nous. Pourtant, la technique est élaborée par des individus. Elle n'existe ni n'agit en dehors d'eux. Dire cela, c'est redonner sa chair à la technique, et refuser d'oublier que toute machine suppose qu'il y ait des machinistes et des machinations. Concernant les algorithmes, il n'est pas rare d'entendre dire que leur fonctionnement est neutre sous prétexte qu'il est automatisé, ce qui reviendrait à gommer les rôles pourtant décisifs joués par celui qui a écrit l'algorithme, et par celui qui y a recours. Si nous reprenons l'exemple de Google, on entend régulièrement les ingénieurs de l'entreprise de Mountain View prétendre que les résultats produits par leur moteur de recherche sont « naturels » (en anglais « *organic results* »), comme si le signal de sortie de l'algorithme avait poussé sur la branche d'un arbre, par l'opération de Mère Nature et du Saint-Esprit. Cela sous-entend que les résultats n'auraient pas pu être différents, le fonctionnement du moteur étant prétendument neutre ; les ingénieurs ne pourraient par conséquent en aucun cas être tenus pour responsables de ce qui s'affiche à l'écran, et éviteraient de cette façon de se trouver confrontés aux mêmes problèmes que ceux qu'affrontent les journalistes qui, eux, sont tenus pour responsables de ce qui est publié dans leurs pages, et condamnés en cas de diffamation.

Le deuxième postulat de la sociologie des techniques, qui découle du précédent, consiste à prétendre que la technique est une forme de discours. En produisant une voiture comme ceci, et non comme cela, en écrivant un algorithme de cette façon, et pas d'une autre, bref, en « concevant » une technique, ses « concepteurs » ont exprimé leur « conception » de *ce qu'il fallait qu'elle soit*. Cette conception dépend des contraintes économiques et techniques, d'une part, et, d'autre part, de la façon dont le technicien sait en jouer, de la même manière que le discours d'un individu dépend des contraintes propres à la langue qu'il utilise, et de son niveau de maîtrise de la langue, des procédés d'argumentation et des moyens de communication. Les caractéristiques d'une voiture ou d'un algorithme ne dépendent pas seulement

des contraintes auxquelles sont soumis ceux qui les produisent, mais aussi de leurs valeurs, de leurs relations de pouvoir (comment prennent-ils les décisions ? quelqu'un a-t-il le dernier mot ?), des lois qui encadrent la production et l'utilisation, des modalités de financement (y a-t-il des actionnaires ? des objectifs à atteindre ? des emprunts à rembourser ?), de l'expérience des techniciens, des retours des consommateurs, des capacités des machines chargées de produire la voiture ou d'exécuter l'algorithme, *etc.* La sociologie des techniques essaye d'identifier ces tenants dont le fonctionnement technique est l'aboutissant, de comprendre ce qui détermine quoi, comment, et de les sonder. Pour un algorithme, et pour comprendre, *a fortiori*, comment d'éventuels phénomènes de censure algorithmique peuvent se mettre en place, il s'agira d'identifier à quels arbitrages les ingénieurs ont procédé au moment de définir les opérations, les critères, les pondérations et les constantes, et comment ils justifient d'avoir procédé à tel ou tel choix décisif.

Le troisième postulat de la sociologie des techniques consiste à prétendre que la technique est inter-action, c'est-à-dire littéralement qu'elle se situe « entre » plusieurs cours d'action, et que par son intermédiaire différents acteurs interagissent. On dit aussi qu'elle est le lieu d'une « médiation »⁵. Ainsi, la technique crée-t-elle un lien entre des acteurs qui certes ne se rencontrent jamais face à face, et n'agissent pas de façon synchronique, mais qui malgré tout interagissent, et dépendent les uns des autres. Quand j'utilise un algorithme, je « rencontre » celui qui l'a écrit, même si cette personne est loin de moi dans l'espace et dans le temps. Nous interagissons par la médiation de l'algorithme. Je peux agir d'une manière qui n'a pas été prévue par l'auteur de l'algorithme, je peux m'en servir pour autre chose. En reprenant le vocabulaire de Michel De Certeau, on dira dans ce cas que j'agis comme « un tacticien » par rapport à l'auteur de l'algorithme : le « *stratège* »⁶. Il fixe les règles, j'en joue.

On ne peut interroger l'idée d'une censure algorithmique qu'à condition d'avoir à l'esprit ce triple postulat, et de construire un appareil interdisciplinaire permettant d'approcher l'objet sans le réduire à une de ses dimensions, et de voir si oui ou non il écarte – et pourquoi, comment, de quel droit... – certains contenus, et s'il les écarte *systématiquement*.

5 M. AKRICH, « La description des objets techniques », in M. AKRICH, M. CALLON et B. LATOUR (dir.), *Sociologie de la traduction. Textes fondateurs*, Presses des Mines de Paris, 2006, p. 159-178.

6 Sur la différence entre le stratège et le tacticien, voir M. DE CERTEAU, *L'invention du quotidien*, t. 1, *Arts de faire*, UGE, coll. « 10-18 », 1980, 348 p.

III. UNE OPACITÉ PROBLÉMATIQUE

Parfois c'est l'algorithme lui-même qui est censuré quand au nom du secret industriel, les critères, les pondérations, les opérations et les constantes ne sont pas communiqués. Dans ce cas la critique dont nous vantions les mérites à l'instant est considérablement plus difficile, mais n'est pas impossible. À défaut d'en connaître la teneur, il s'agira de scruter le comportement des algorithmes. C'est ce que font un certain nombre de chercheurs en sciences de l'information et de la communication. Par exemple, imprégnés par le courant des « *software studies* »⁷, Rieder, Matamoros-Fernández et Coromina ont étudié comment l'algorithme de Youtube se comportait et de quelle manière son fonctionnement induisait, et nourrissait, la « *ranking culture* » des utilisateurs⁸.

Même si d'autres voies peuvent être trouvées, il n'en reste pas moins que l'opacité des algorithmes, qui elle-même peut être assimilée à une forme de censure (on ne censure plus le message dans ce cas mais les conditions de son acheminement) est problématique. En effet, comme l'a souligné le philosophe Paul Mathias, il s'agit du point de vue épistémologique d'un renversement de l'adage socratique : au lieu de savoir que je ne sais rien, je sais tout, mais à la seule condition de ne rien savoir de la manière dont c'est possible⁹.

Pourquoi les algorithmes sont-ils opaques ? D'une part, il s'agit d'éviter pour leurs auteurs d'être imités. À la fin des années 1990, les algorithmes œuvrant sur Internet pour traiter l'information étaient encore largement transparents, produits selon les canons d'un paradigme académique, dans un écosystème du Web encore jeune, où l'imitation était monnaie courante. Ainsi le PageRank de Google et l'algorithme HITS de Jon Kleinberg reposaient-ils sur des postulats semblables¹⁰. Après l'éclatement de la bulle Internet, lorsque le marché s'est concentré et que l'activité est devenue rentable¹¹, le paradigme académique a cédé le pas au paradigme industriel. Et après quelques publications de brevets, les acteurs impliqués ont opté pour le « secret industriel », à leur avis mieux à même de compliquer sinon d'empêcher l'imitation par les concurrents.

7 C. MÉADEL et G. SIRE, « Les sciences sociales orientées programme. État des lieux et perspectives », *Réseaux*, n° 206, 2017, p. 9-36.

8 B. RIEDER, A. MATAMOROS-FERNÁNDEZ et Ò. COROMINA, « From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results », *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 2018, p. 50-68.

9 P. MATHIAS, *Qu'est-ce que l'Internet ?*, Vrin, coll. « Chemins Philosophiques », 2009, 128 p.

10 G. SIRE, *Les moteurs de recherche*, La Découverte, coll. « Repères », 2016.

11 E. VAN COUVERING, « The History of the Internet Search Engine: Navigational Media and the Traffic Commodity », in M. ZIMMER et A. SPINK (dir), *Web Search. Multidisciplinary Perspectives*, Springer, 2008, p. 177-206.

L'opacité algorithmique permet également de décourager les stratégies consistant à endosser le rôle de tacticien au point de défaire le stratège sur son propre terrain. En effet, de nombreux éditeurs façonnent leurs contenus en fonction de ce qu'ils savent des algorithmes dont il leur faudra franchir le filtre. Ils n'hésitent pas à choisir les sujets qu'ils traitent, mais aussi les mots qu'ils emploient, la structure de leurs sites, la destination des liens hypertextes et bien d'autres choses, dans le seul but d'être visibles sur les plateformes, hissés en haut des classements effectués par les algorithmes¹². Cela s'appelle la « Search Engine Optimization » (SEO). Si jamais des plateformes comme Google ou Facebook publiaient en détail leurs algorithmes, leurs concepteurs risqueraient que ce ne soit pas forcément le plus pertinent qui grimpe en haut des classements, mais le plus malin, le plus roublard, le plus bricoleur, « le plus braconnier » aurait dit Michel de Certeau, c'est-à-dire celui qui aurait réussi à retourner les règles du jeu contre le jeu lui-même.

Enfin, l'opacité algorithmique permet de rendre invisibles les éventuels biais susceptibles d'être introduits dans l'algorithme, et, donc, la censure systématique dont celui-ci pourrait se faire l'instrument. Larry Page et Sergey Brin eux-mêmes, au moment de fonder Google, alors qu'ils n'étaient encore que des doctorants idéalistes loin de se douter qu'un jour ils deviendraient les plus importants publicitaires du monde, écrivirent dans l'annexe de l'article destiné à présenter leur invention, que jamais ils ne pactiseraient avec des annonceurs publicitaires pour la simple (et excellente) raison qu'un « *moteur de recherche pourrait ajouter un 'petit' facteur visant à avantager les entreprises amies et soustraire le facteur équivalent des résultats liés à leurs concurrents. Ce type de biais est très difficile à détecter, alors qu'il est susceptible d'avoir un impact significatif* »¹³.

IV. LA CENSURE DE FAIT

À présent que nous avons expliqué d'une part pourquoi il était nécessaire de procéder à une analyse critique des algorithmes en général, et en particulier de ceux des plateformes d'accès aux contenus culturels, et que nous avons expliqué d'autre part pourquoi les algorithmes étaient opaques, et comment cette opacité pouvait donner lieu à un phénomène de censure, j'aimerais expliquer comment la censure peut être induite par les algorithmes sans être forcément systématisée en leur sein. En effet, étant donné leur « faire faire »

12 G. SIRE, « Le pouvoir normatif de Google. Analyse de l'influence du moteur sur les pratiques des éditeurs », *Communication & Langages*, n° 188, 2016, p. 91-105.

13 S. BRIN et L. PAGE, « The Anatomy of a Large-Scale Hypertextual Web Search Engine », *Computer Networks and ISDN Systems*, Stanford University, 1998 (Annexe A, notre traduction).

ou « *gouvernementalité algorithmique* »¹⁴, les plateformes induisent certains comportements aboutissant, de fait, à écarter certains contenus, ou en tout cas à les rendre si peu visibles que l'effet sera le même que si on les avait censurés, et ce alors même que ce n'est pas ce que les auteurs desdits algorithmes ont souhaité.

Le premier de ces effets a été pour ainsi dire « transmis » à Internet par le monde de l'évaluation scientifique. Il a été mis à jour dans les années 1960 par Robert Merton, lorsque celui-ci a constaté que si l'on considère qu'un contenu cité à de nombreuses reprises est probablement plus pertinent que les autres, et si pour cette raison on le propose à la lecture avant les autres, alors celui-ci sera davantage lu, et pour cette raison il sera davantage cité, ce qui renforcera son autorité, de sorte qu'il sera plus souvent lu, et d'autant plus souvent cité, *etc.* Merton a nommé cet effet « effet Matthieu », à cause d'une phrase de l'Évangile : « *on donnera à celui qui a, et il sera dans l'abondance, mais à celui qui n'a pas on ôtera même ce qu'il a* » (Mt 13 :12). Cela nous intéresse dans la mesure où cela nous apprend que de fait un algorithme qui reposerait sur les mêmes principes que le PageRank (ce qui est le cas pour tous les algorithmes de plateformes d'accès aux contenus culturels) finirait par avantager de manière systématique certains contenus, ne laissant aucune chance aux autres d'être visibles. Un contenu qui existerait depuis plus longtemps qu'un autre, toutes choses étant égales par ailleurs, se trouverait avantagé. En matière algorithmique en effet, la valeur attend le nombre des années, au point que les plateformes doivent introduire des critères liés à la fraîcheur (c'est ce que fait Facebook) ou bien créer des sous-ensembles dédiés aux contenus les plus récents (c'est ce que fait Google Actualités), de manière à contrebalancer l'effet Matthieu. Mais malgré ces dispositifs, les plateformes ont tendance à mettre toujours les mêmes contenus en avant. Cela a été montré dans le cas de Google dès 2003 : « *Plutôt que de démocratiser la dissémination de l'information, l'étude de la googlearchy suggère que les citoyens pourraient continuer à obtenir des informations concernant la politique provenant de seulement quelques sources, et ce même s'il semble que les points de vue soient en nombre illimité dans le cyberspace [...] L'architecture ouverte n'est pas une garantie en termes de démocratie. Le nouveau média, comme les anciens, concentre l'attention sur un petit nombre de sources divulguant des informations politiques* »¹⁵.

La chercheuse Susan Gerhart a montré quant à elle comment, à cause des effets Matthieu, les algorithmes pouvaient avoir tendance à atténuer

14 R. BADOUARD et C. MABI et G. SIRE, « Beyond 'Points of Control': logics of digital governmentality », *Internet Policy Review*, 5(3), 2016.

15 M. HINDMAN, K. TSIOUTSIOULIKLIS et J. A. JOHNSON, « 'Googlearchy': How a Few Heavily Linked Sites Dominate Politics on the Web », Annual meeting of the Midwest Political Science Association, 2003.

les controverses en privilégiant ce qui fait consensus dans les débats. Pour affirmer cela, elle a effectué une batterie de requêtes sur des portails d'accès à propos de cinq sujets controversés et a constaté que les portails avaient tendance à privilégier les informations consensuelles. Par exemple, pour une requête concernant l'État du Belize, *Google* fait remonter des sites institutionnels et touristiques plus volontiers que des informations concernant les conflits frontaliers du Belize avec le Guatemala. De tels résultats ne sont pas imputables qu'aux seuls ingénieurs de Google, mais également au fait que les liens effectués vers les sites institutionnels et touristiques sont plus nombreux que les liens vers les pages consacrées aux conflits, et au fait que les internautes sont plus nombreux à chercher une information factuelle et touristique qu'à chercher des informations concernant les conflits. Ainsi, selon Susan Gerhart, ce n'est pas « *la technologie qui supprime les controverses, mais le social* »¹⁶. Des éléments à propos des controverses existent sur le Web, mais ils sont relayés après les 50 premiers résultats du moteur de recherche. La solution, selon la chercheuse, ne devrait donc pas tant venir des concepteurs des moteurs que des internautes qui devraient prendre conscience du biais produit par les moteurs et de leur prédisposition à présenter le bon côté des choses (« *sunny side* »). Ils devraient alors faire en sorte de diversifier eux-mêmes les résultats qu'ils obtiennent, pour accéder au « *dark side* » en multipliant les requêtes et en utilisant plusieurs moteurs.

Une autre censure de fait peut-être induite par l'attention prêtée dans les critères algorithmiques à la performance. Les plateformes peuvent en effet considérer que la vitesse de chargement d'une page est un signe positif. Cela revient à faire de la performance du contenant un signal de pertinence de contenu, ce qui évidemment est discutable. On imagine mal deux professeurs d'université, en cas de désaccord, courir un cent mètres et décréter que celui qui sera arrivé le premier détiendra la vérité. Pour les contenus numériques, la question est d'autant plus délicate si l'on considère qu'il existe une solution payante permettant d'accélérer le chargement de la page, en recourant aux services d'un « *content delivery network* » (CDN), afin de multiplier et de répartir les serveurs d'hébergement. Il existe également des systèmes de « *peering* » payant, permettant à un hébergeur de contenus particulièrement lourds, comme par exemple Netflix ou Youtube, de voir son contenu systématiquement séparé du reste pour être acheminé plus vite vers les ordinateurs des utilisateurs finaux. De tels contenus pourraient donc être privilégiés par des algorithmes prêtant attention à la performance, au détriment d'autres contenus hébergés sur des serveurs moins puissants, ou bien privés de solution de CDN ou de « *peering* », qui eux se trouveraient relégués en queue de peloton, et, de fait, censurés, ou en tout cas discriminés.

¹⁶ S. GERHART, « Do Web search engines suppress controversy? », *First Monday*, vol. 9, n° 1, 2004.

Le dernier point pouvant conduire à une censure de fait est celui de la personnalisation. Cette question en particulier a été popularisée notamment par un ouvrage du journaliste américain Eli Pariser¹⁷ accusant les plateformes d'accès de créer, avec leurs algorithmes, des « bulles filtres » à cause desquelles le Web ne serait plus tant un outil d'information qu'un outil de confirmation, enfermant l'internaute dans ce qu'il pense déjà, en ne lui proposant que des contenus qui « lui ressemblent » en censurant de fait les opinions avec lesquelles il risquerait de ne pas être d'accord. Comme l'ont montré de nombreux auteurs, la réalité est moins évidente que le propos d'Eli Pariser ne semble le supposer¹⁸. En réalité, les critères algorithmiques liés à l'historique de navigation, la position géographique, la langue, les contacts, *etc.*, ne conduisent jamais à une personnalisation telle que l'internaute serait « enfermé ». Ainsi selon Dominique Cardon, l'idée que la personnalisation « réduit l'univers de l'utilisateur à ses propres centres d'intérêt ne se vérifie guère. Internet contribue moins à isoler les internautes dans un espace informationnel restreint qu'à renforcer la coupure entre ceux qui s'exposent massivement aux informations et ceux qui s'éloignent de plus en plus de l'espace public »¹⁹. Même s'il ne faut pas exagérer les effets de « bulle filtre », la question des critères algorithmiques liés à la personnalisation, et de leur pondération, n'en mérite pas moins d'être posée. Notamment, les ingénieurs, mais aussi les utilisateurs et le régulateur, doivent chercher autant que possible, comme le préconise Engin Bozdag, à « minimiser les effets négatifs et [à] maximiser les effets positifs [des technologies de personnalisation] plutôt que d'essayer de nous en débarrasser totalement. La question n'est pas de savoir s'il doit y avoir personnalisation ou non, mais de comprendre comment mettre au point une technologie de personnalisation moralement acceptable »²⁰.

Un algorithme, nous l'avons vu, ce sont des opérateurs logiques, des pondérations et des constantes permettant d'agencer des variables. La machine exécute l'algorithme, ce qui permet à son auteur d'exercer un contrôle *a priori* sur le signal de sortie. Nous avons vu pourquoi il convenait de ne pas tomber dans le panneau de la neutralité, cela car un algorithme,

17 R. PARISER, *The Filter Bubble: What the Internet Is Hiding from You*, New York, Penguin Press, 2011.

18 Voir par exemple : D. J. VON SCHOULTZ et J. VAN NIEKERK, « You according to Google: the effects of filter bubbling », in *Proceedings of the 14th annual Conference on World Wide Web Applications*, Cape Peninsula University of Technology, 2012.

19 D. CARDON, « L'ordre du Web », *Médium*, vol. 4, n° 29, 2011, p. 191-202.

20 E. BOZDAG, « Bias in algorithmic filtering and personalization », *Ethics and Information Technology*, vol. 15, n° 3, 2013, p. 209-227.

comme toute technique, est le fruit d'un certain nombre d'arbitrages, eux-mêmes résultant de contraintes, de principes, de prises d'intérêts et de lois. Lui redonner sa chair, c'est lui rendre, analytiquement, le rôle qu'il joue en pratique, à la fois comme discours et comme médiateur. Et c'est s'interroger sur la censure possible, non pas comme si elle était produite par l'algorithme, ce qui ne veut rien dire, mais comme si le fonctionnement d'un système numérique de traitement de l'information aboutissait à une situation de censure, soit parce que ses concepteurs l'auront souhaité, soit parce qu'ils n'ont pas pu l'empêcher.

Nous avons ensuite vu comment il pouvait y avoir une censure de deuxième niveau, opérée non par le truchement de l'algorithme lui-même, mais par ses auteurs qui s'autocensurent pour éviter que des concurrents ne les imitent, que des éditeurs de contenus ne détournent l'algorithme à leur avantage ou encore pour ne pas risquer d'être accusés de favoritisme.

Nous ne pouvons nous poser la question de la censure algorithmique des contenus culturels sur les plateformes qu'à condition d'avoir compris à la fois ce qu'est un algorithme, pourquoi les principaux algorithmes sont opaques et quels sont les effets de censure qui peuvent se produire indépendamment de la volonté des ingénieurs (même s'ils y sont directement pour quelque chose). *A fortiori*, toute réflexion concernant la régulation ne pourra avoir lieu qu'à condition d'avoir d'abord déployé la logique algorithmique et les modalités du contrôle qu'elle sous-tend. C'est ainsi qu'il sera possible par exemple, pour éviter l'effet Matthieu, d'imposer un biais temporel, ou bien de demander aux concepteurs des plateformes de prévoir un dispositif grâce auquel les utilisateurs pourront en paramétrer un. De même, c'est en comprenant l'enjeu des « *content delivery networks* » et du « *peering* » qu'il sera possible ensuite d'interdire de prendre en compte la performance des contenants pendant des périodes déterminées, par exemple au moment d'une campagne électorale, de manière à éviter qu'un site soit défavorisé sous prétexte que son propriétaire n'a pas de quoi s'offrir un hébergeur multi-localisé. C'est à cette condition également qu'on pourra sortir du « pour ou contre la personnalisation », et débattre des critères un à un, de leurs pondérations respectives, et de la nécessité d'une part d'informer les utilisateurs quant à la personnalisation dont ils font l'objet, et d'autre part de leur donner la main sur certains critères.