

Algorithms of Oppression: How Search Engines Reinforce Racism,

By Safiya Umoja

New York: New York University Press, 2018, 256 pp.

ISBN 978-1-4798-3364-1

A Book Review by

Abakar Malloum

Université d'Ottawa, Canada

Big data et la numérisation de l'exclusion sociale

L'idée selon laquelle l'Internet serait une révolution sociale et cognitive, et qu'il permettrait une ère de démocratie généralisée (J. Barlow 1996; M. Serre 2015) laisse le plus souvent dans l'ombre la dynamique de transfert des représentations entre les cyberspaces et le monde réel. Il y a une relation de détermination mutuelle entre la technique (les applications) et les valeurs que portent ses concepteur.trice.s (ex. ingénieur.es) et ses contrôleur.euse.s (les géants du numérique). Telle est la thèse générale de Safiya Umoja Noble dans *Algorithms of Oppression: How Search Engines Reinforce Racism*. S. Noble est professeure adjointe en études de l'information à l'Université de Californie. Sa recherche, indique-t-elle, est concentrée sur « unveiling the many ways that African American people have been contained and constrained in classification systems, from Google's commercial search engine to library databases » (2018, p. 5). C'est donc dans le contexte de ce programme que l'auteure marie les perspectives issues du féminisme noir et les concepts fournis par les études bibliothéconomiques afin de mettre en lumière des pratiques d'exclusion et de marginalisation numériques («algorithmic redlining») dont sont victimes des personnes issues de certaines catégories sociales, notamment les femmes, les personnes noires et les personnes issues des communautés latino-américaines.

Selon Noble, le tournant du numérique, que certains appellent « le capitalisme de la surveillance », coïncidant avec le contrôle de l'information publique par des entités privées, au premier chef desquelles Google, s'est opéré conjointement à la capitalisation des préjugés sociaux qui sont devenus sources de profits commerciaux et par-là, des valeurs consommables. Ainsi par exemple quand on cherche «black girl» dans Google, des termes oppressifs comme «Big booty Black girls» (2018, p. 67) apparaissent *en premier* parce que des sites pornographiques payent Google pour que ces « réponses », qui sont en fait des annonces ('ads'), s'affichent en tête des résultats possibles d'une requête.

Partant d'ailleurs de cet usage commercial des préjugés, les deux premiers chapitres du livre démasquent le processus par lequel les catégories oppressives combattues par «women and people of color» (2018, p. 26) sont reproduites et renforcées par le biais des résultats de recherche Google (*Google Search*). Ainsi quand, entre 2009-2015, Noble cherchait dans Google Search les termes «women need to», elle recevait les suggestions suivantes : «be put in their places, know their place, be controlled, be disciplined» (2018, p. 15). Idem, quand Noble cherchait les termes «black girls», il en résultait ce qui suit : «black girls pussy, Free black girls chats, black girls big booty» (2018, p. 19). À la demi-question, «why are black people so», elle a reçu les réponses suivantes: « angry, loud, mean, attractive, lazy, annoying, confident, sassy» (2018, p. 21). La demi-question «why are white women so» provoquait les termes plus positifs tels que «pretty, beautiful, mean, easy, insecure, skinny, perfect» (*Ibid*). Quant à

Google Image Search, le terme “professor style” représentait quasi totalement des images d’hommes blancs en costume. Le problème, selon Noble, ne se trouve pas exactement dans le fait que ces résultats reflètent un transfert dans le numérique de l’oppression dont font objet ces groupes sociaux; il serait plutôt particulièrement dérangent parce que ces résultats sont considérés comme étant, pour une grande majorité d’internautes, objectifs et neutres – « a simple mirror of the collective » (2018, p. 36). Or, comme le soutient Noble, le miroir de la société n’est ni simple ni neutre, il est un lieu de pouvoir et de domination au sein duquel *Google Search* est un outil de normalisation, comme l’a été la Bibliothèque du Congrès américain (nous y reviendrons) ; de ceux qui sont déjà «overracialized and hypersexualized» (*Ibid*).

Cette oppression numérique, selon l’auteure, n’est pas le fait d’une machine « anomique » comme l’insinuaient la campagne de Nations unies en 2013 (2018, p. 15), encore moins le simple fruit d’une erreur (*glitch* ou *bug*) dans le système internet. Elle est plutôt le fait d’une triple rencontre entre l’idéologie néolibérale, une absence de protections légales et des pratiques oppressives profondément enracinées dans les institutions sociales et universitaires. À titre d’exemple, aux Etats-Unis, où les protections se font toujours attendre, on trouve de « greater encroachments on personal information privacy [...] vulnerable communities and individuals are less likely to find recourse » (2018, 22). Cette absence relative de l’État dans la protection des internautes américains permet à des sites tels que Mugshots.com et UnpublishArrest.com, de proposer aux plus offrants la possibilité d’être effacés «across all major search engines» (2018, 124). Comme l’a constaté Kandis, (la coiffeuse afro-américaine dont l’entrevue est au cœur de la conclusion du livre), l’internet est en train de redéfinir « who is valuable and where the value lies» (2018, 175). Par ailleurs, il y existe certains endroits où des exemples de protections semblent émerger : en France, par exemple, les lois interdisent la vente en ligne d’objets nazis, Google filtrant ses résultats conformément aux dispositifs légaux (2018, p. 42). De même, les législations relatives au « Right to Be Forgotten » mises en place dans l’espace de l’Union européenne limitent le contrôle de Google sur les données personnelles des internautes.

C'est en prenant le contre-pied de la thèse de la neutralité du Net que le chapitre trois de Noble expose comment cette fausse neutralité et objectivité peut mener à des conséquences tragiques comme le massacre raciste (2015) de neuf Afro-Américain.es lors d’une prière à l’église où l’assaillant, Dylann Roof, «allegedly used Google Search in the development of his racial attitudes» (Noble 2018, p. 11 et p. 133). Pour Noble, l’idée que l’Internet est un outil politiquement neutre est à la base de la conception que les internautes se font des résultats de leur recherche sur Google. Résultats, y compris les préjugés commercialisés, que ces mêmes internautes considèrent comme une représentation neutre, objective et donc naturelle de la société. Mais ces pratiques oppressives, dans lesquelles Google et consorts sont devenus maîtres numériques, sont-elles nouvelles? Cette question Noble la traite dans le chapitre cinq pour débouter une autre fausse neutralité, celle de *la science de l’information*, notamment la « Traditional library and information science (LIS)» (Noble 2018, p. 137). Tel que l’explique Noble, cette science de classification est non seulement impliquée dans la conception du *Google Search*, mais aussi et surtout dans la longue histoire de «how people are conceptualized and represented» (Noble 2018, p. 135) à travers les systèmes de classification bibliothécaire.

Les classifications - qu’elles soient numériques ou non - sont politiquement chargées, car elles sont à la fois les produits de leur temps et porteuses de la vision du monde de la classe dominante. Sur le plan historique, le cas de la classification est particulièrement pertinent en Amérique du Nord où la taxinomie scientifique a servi à la légitimation de l’exclusion sociale et politique des autochtones, des femmes et des afro-américain.es. C’est en ce sens que Noble expose les relations étroites entre les pratiques oppressives de Google et une tradition scientifique compromise dans les nominations (*Namings*) des Asiatiques comme “Yellow Peril”, des juifs comme “Jewish Question”, des Afro-Américain.es comme “Negroes” et des femmes comme “Women as Accountants” par la bibliothèque

nationale. Ainsi, si les violences numériques dans les catégorisations et les collectes de données nous renvoient aux pratiques incrustées dans les gestions de la connaissance et de l'information, Noble ne peut que réitérer l'importance des réglementations publiques du numérique qui manquent de façon criante aux États-Unis, un pays où le sujet est « hotly contested » (Noble 2018, p.160).

L'ouvrage de Noble est, sans conteste, un fort appel à nuancer les hypothèses selon lesquelles les cyberespaces sont un havre de liberté. Il nous faut, selon Noble, se résoudre à affronter la réalité: «social inequality will not be solved by an App» (Noble 2018, p.165). Les entités privées qui contrôlent l'Internet et nos diverses applications sont essentiellement orientées vers le profit, leurs algorithmes sont avant tout des machines à faire de l'argent, le tout informé par une vision néolibérale du monde. L'une des premières choses qui sera déplorée par certain.e.s lecteur.trice.s, peut-être, est le fait que, comme l'auteure le reconnaît dès l'introduction, la plupart des pratiques numériques oppressives analysées dans le livre n'étaient déjà plus d'actualité au moment même de la publication : *Google Search* ne suggère plus des réponses oppressives aux termes « femme », « personne noire », etc. En effet les algorithmes de *Google search* semblent avoir changé de stratégies ou « d'espace de recherche » selon le terme consacré en informatique. Ainsi en 2020, quand nous faisons la requête « woman's », Google se contente de citer des passages des livres, des films ou des chansons populaires en guise des suggestions. Par exemple « woman's heart is a deep ocean of secrets » renvoie à une scène du film *Titanic*. N'empêche, cette formule demeure teintée des préjugés combattus par les femmes, ce qui confirme l'actualité de la thèse générale de l'auteure : la réalité virtuelle américaine est originaire de la société américaine où les dispositifs culturels et scientifiques ont pour longtemps été complices des pratiques oppressives envers certaines catégories sociales. *Google PageRank* est un produit de ces dispositifs.

Il nous faut toutefois noter aussi certains raccourcis méthodologiques dans la manière dont l'auteure traite les résultats de *Google Search* dans la mesure où ceux-ci sont centraux à la thèse. Ainsi, pour pouvoir attribuer les résultats racistes et sexistes aux intentions humaines ou d'affirmer que « racism is the fundamental application program interface (API) of the Internet » (Noble 2018, p. 4), il nous semble indispensable d'expliquer comment cet algorithme spécifique de classification (*Google Search*) procède pour produire ces résultats. Car sans cela, on fait abstraction des dimensions importantes de classification algorithmique telle que la 'boite noire', ou par exemple le fait que, selon Jenna Burrel «rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs» (2016, p. 1). Non seulement il est souvent difficile de prévoir les sorties (*outputs*) comme le souligne Burrel, mais comme Gregory Chaitin l'a fameusement démontré : dans tout processus computationnel, il y a une dissymétrie entre les entrées (données utilisées par les algorithmes/*inputs*) et les sorties (résultats/*outputs*). Et de là, on peut spéculer qu'il se passe la même chose entre les intentions des concepteur.trices et les résultats entendus. Un autre aspect épistémologiquement important du constat de Chaitin est que dans le processus mathématique/algorithmique de la pure rationalisation même, il y a des aléas (*randomness*) «area in which mathematical truth has absolutely no structure, no structure that we will ever be able to appreciate in detail» (2004, p. 4).

Cet élément d'irrationalité - «maximally unknowable» - dans le fonctionnement de l'algorithme ne conduit nullement à une neutralité idéologique de la machine : un algorithme est avant tout un discours sur le monde quelle que soit l'intentionnalité derrière. Par exemple un algorithme d'une caméra de surveillance (ou de classification) des humains doit contenir une *définition* de ce qu'un être humain physiquement. Cette définition peut donc déboucher sur l'exclusion/inclusion sur des bases idéologiques (racistes et sexistes), économiques, sécuritaires, etc. Mais une évaluation éthique de cet algorithme nécessite de l'analyser en passant par un certain nombre de paliers distincts de sa constitution: l'algorithme (la partie purement théorique), le code (langage choisi pour l'exécution du programme), les données à la disposition de ce programme (les liens des sites dans le cas de Google), le contexte, etc.

En outre, un arrêt sur l'algorithme lui-même aurait utilement mené l'auteure à examiner le remplacement par Google Search (en 2011) de son algorithme par un nouvel algorithme («Panda») qui réduisit fortement les possibilités qu'avaient certains sites, ceux qui sont « de faible qualité » selon Google, à se placer en tête des résultats. Ce qui aurait du même coup projeté plus de clarté sur la classification, à savoir pourquoi et comment les termes comme «black girl's » et « woman's», etc., venaient avec ces résultats et ensuite peut être de proposer aux programmeur.euses chez Google et aux législateur. trice.s de *types d'intervention* que Noble appelle si justement de ses vœux.

Bibliographie

- Barlow, John. (1996). *A Declaration of the Independence of Cyberspace*, 'Hache' Une édition électronique.
- Burrell, Jenna. (2016). «How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms», dans *Big Data & Society*, 3(1), 1-12.
- Chaitin, Gregory. (2004). *Leibniz, Randomness & the Halting Probability*, IBM Watson Research Center: Yorktown Heights.
- Serres, Michel (2015, 2012). *Petite Poucette*, Paris : Le Pommier.

About the Reviewer

Abakar Malloum is a Doctoral candidate at the at the School of Political Studies, University of Ottawa, Ottawa, Canada

Citing this book review:

Malloum, Abakar (2019). [Review of the book *Algorithms of Oppression: How Search Engines Reinforce Racism*]. *Global Media Journal -- Canadian Edition*, 11(2), 117-120.