

Retour sur le livre de Dominique Cardon

Maxime Amblard

À quoi rêvent les algorithmes, Nos vies à l'heure des big data, Dominique Cardon, La république des idées, Seuil, Octobre 2015.

Dominique Cardon est un sociologue spécialisé dans le numérique et Internet qui propose de revenir sur les *algorithmes*¹. La place prise par l'informatique et le calcul dans l'organisation de nos sociétés est une question importante et l'auteur montre la nécessité de réfléchir aux nouveaux usages et possibilités liés aux *algorithmes*.

Le travail de Dominique Cardon apparaît à la fois d'une grande nécessité, et souffrant de plusieurs défauts. L'auteur focalise son discours et notre attention sur les *algorithmes* sans chercher à donner de définition de ce qu'il considère comme tel. Cette imprécision dans les concepts le conduit à faire des raccourcis qui fragilisent son argumentation.

Ce livre ne parle pas des algorithmes mais d'outils informatiques. Il est vrai que les outils informatiques sont constitués de programmes, qui implémentent des algorithmes. Mais la partie ne fait pas le tout. Un algorithme est d'abord une façon de décrire le calcul. Le livre "À quoi rêvent les algorithmes" devraient plutôt se placer au cœur du calcul et non seulement analyser les objets sur lequel porte le calcul. C'est une différence importante qui laisse croire que l'analyse proposée vaut pour toute l'informatique, ce qui n'est pas le cas. Ici il s'agit de discuter d'une partie très spécifique relative aux usages de l'informatique, de plus en la réduisant aux interactions sur Internet. Un livre avec une telle ambition ne devrait pas alimenter cette ambiguïté.

L'informatique n'est pas Internet. C'est une science. Et elle ne se réduit pas à la science des usages ou de la donnée. C'est une science qui s'intéresse à la résolution de problèmes par le calcul, avec des aspects aussi divers que la modélisation mathématique², la logique³ ou encore la complexité des problèmes⁴ sur un versant théorique, ou des réalisations pratiques comme les réseaux⁵, l'impression 3D⁶ et la cryptographie, qui sont basés sur des modèles mathématiques élaborés.

À l'heure où d'importantes discussions sont en cours à propos des programmes à introduire dans l'enseignement secondaire sur l'informatique, il est important de faire entendre cette différence. On retrouvera sur⁷ la présentation de Gilles Kahn à l'académie des sciences⁷ sur cette question : il est souvent aisé d'avoir une intuition sur la technologie alors qu'il est difficile de percevoir la question scientifique qui est derrière.

On retrouve ces approximations dans les définitions à propos de *machine learning*, d'intelligence artificielle, ou de traitement automatique des langues. S'il est vrai qu'il est délicat de proposer des définitions simples de ces concepts, il ne faut pas minimiser les enjeux politiques qui sont derrière. Entretenir la confusion obscurcit le débat plutôt que de le libérer. Le *machine learning* n'est pas l'intelligence artificielle qui n'est pas le traitement automatique de la langue.

1. Le terme d'algorithme utilisé par l'auteur est à la fois plus large et plus étroit que celui utilisé par les informaticiens. Pour distinguer les usages, algorithme sera marqué en italique lorsqu'il s'entend selon la définition de Dominique Cardon.

2. https://interstices.info/jcms/p_86521/mieux-modeliser-le-climat-grace-aux-statistiques

3. https://interstices.info/jcms/c_23115/preuves-formelles-preuves-calculatoires

4. https://interstices.info/jcms/p_81141/p-np-elementaire-ma-chere-watson

5. https://interstices.info/jcms/ni_75926/les-reseaux-pour-le-calcul-haute-performance-facteur-livre

6. https://interstices.info/jcms/ni_77930/des-logiciels-pour-l-impression-3d

7. https://interstices.info/jcms/c_7125/sciences-et-technologies-de-l-information-et-de-la-communication

- Le *machine learning* est un champ d'étude de l'informatique qui s'intéresse aux algorithmes de reconnaissance de motifs et d'apprentissage. Il s'agit de déterminer la suite d'un algorithme sans qu'elle n'ait été explicitement donnée, en fonction de résultats numériques obtenus à partir d'un motif.
- L'intelligence artificielle est le champ qui définit des algorithmes pour résoudre des tâches au moins aussi bien, voire mieux que ne le font actuellement les humains. Dans ce cas, l'intelligence artificielle peut mettre en oeuvre des techniques de *machine learning*, mais pas seulement.
- Le traitement automatique des langues se concentre sur la capacité de langage des humains. Il s'agit de comprendre ou de produire automatiquement des textes ou des sons en langue naturelle. À nouveau il est possible de répondre à ces défis par le *machine learning* (mais pas seulement).

Le *machine learning* trouve des cas d'applications à la fois en intelligence artificielle et en traitement automatique des langues. Il peut être vu comme un outil technique pour l'IA tout en restant un objet théorique. Bien sûr que des relations existent entre ces concepts, mais ils n'ont ni les mêmes objets d'études, ni les mêmes paradigmes.

Un autre aspect du livre, très symptomatique de la perception de l'informatique par la société actuelle, est l'abus de vocabulaire anthropocentré. Les algorithmes ne rêvent pas. Les algorithmes ne peuvent pas rêver. Les algorithmes n'ont pas de vie propre. Si l'analogie est sympathique, elle a autant de sens que de se demander à quoi rêve le plus de l'addition. Il ne rêve pas. Derrière les algorithmes il y a des programmeurs, des concepteurs, des théories, des entreprises, des centres de recherche, et eux peuvent avoir des desseins ou des rêves pour les programmes. Et c'est eux qu'il faut questionner, d'autant qu'ils ont la capacité de répondre. C'est aussi l'une des ambiguïtés que l'on retrouve dans le concept d'apprentissage du *machine learning* dont j'ai parlé précédemment. L'algorithme n'apprend pas, ne décide pas, ne se comporte pas. Il fournit un résultat en fonction de son contexte d'utilisation. Au mieux, il nous laisse croire qu'il a appris à réagir dans une situation. D'aucuns vont argumenter sur le fait que cela reste un apprentissage. Je ne le crois pas car l'algorithme n'a pas de modèle conceptuel expliquant les relations qu'il construit entre l'ensemble de ses "apprentissages" supposés. Un enfant qui apprend à marcher construit sa perception ontologique par son expérience. Un programme exécuté sur un ordinateur collectionne les décisions, les laissant s'influencer avec plus ou moins de hasard. Mais l'algorithme n'a pas appris ou n'a pas transformé son expérience en connaissance. Il reste au niveau de l'information brute.

Ce qui est inquiétant dans cette personnalisation des algorithmes (des outils informatiques) c'est qu'il serait possible de les désigner comme responsables des évolutions de la société. Mais un algorithme ne fait rien d'autre que ce qu'on lui demande de faire. Individuellement nous faisons le choix de suivre les prescriptions des algorithmes (même de manière inconsciente) et collectivement nous acceptons qu'ils aient une influence sur notre organisation en société. Un phénomène intéressant est que les *algorithmes* prédisent à partir de ce qu'ils ont rencontré (ou le passé), donc ce qui est considéré comme significatif au départ reste influent (au centre), par une forme d'inertie du système. Ainsi les inégalités intégrées à l'initialisation du processus ont une possibilité de devenir la norme.

Les pratiques actuelles sur Internet, malgré leur manque de structuration permettent aussi à des cultures très spécifiques, souvent à la marge, d'avoir une existence qu'elles ne trouveraient pas en dehors. Cette usine à possible ainsi proposée est tout à fait enthousiasmante et met en avant un changement profond de mode de fonctionnement, plus bottom-up (du bas vers le haut). Ces usages permettent à une activité à la marge d'exister grâce à un effet de seuil, mais en la laissant à cette marge.

Une idée intéressante développée par l'auteur est celle introduite autours du concept de loyauté : la nécessité pour un outil informatique de faire bien, et ne faire que, ce qu'il est censé faire. C'est probablement le seul rempart face à une modification de la perception par la société de la science informatique qui conduirait à une diabolisation, du type construction d'un Big Brother capable de tout voir et tout contrôler.

Par ailleurs, ce travail est sous-tendu par un important travail bibliographique provenant

de disciplines différentes (science politique, sociologie, économie, etc) qui donne un éclairage singulier sur ces questions. En particulier sur le rapport que nous entretenons avec les *big data*. Il est vrai que s'attaquer à ces informations va permettre de rapidement faire émerger des résultats, mais sans contextualisation, ces résultats pourraient rapidement devenir inutilisables car non interprétables, d'autant qu'une grande partie de l'information n'est que du bruit dans lequel nous risquons de nous perdre.

Malgré les réserves que j'ai pu avancer, je crois que l'auteur fait ici une alerte importante. Nous avons urgence à développer une réflexion plus précise tant sur les possibles, les usages et les influences de la science du numérique. Il serait dommage de ne pas entendre cette prescription à cause des imprécisions précédentes et plus encore de ne pas s'en saisir. Comment ne pas porter simultanément une réflexion sur les principes du vote quand on programme un système de vote numérique?

Le point de vue qui est présenté dans ce livre reste celui d'un non-spécialiste de l'informatique, ce qui me conduit à comprendre les ambiguïtés précédemment pointées. L'informatique est aujourd'hui très diverse dans ses aspirations et ses tensions et il est complexe de l'appréhender dans sa globalité. Dans son livre, Dominique Cardon esquisse des interprétations épistémologiques et sociologiques de l'informatique (de l'Internet) et c'est certainement pour cela qu'il est nécessaire de lire ce livre : mieux comprendre une partie des enjeux actuels de l'informatique, et mesurer la nécessité de mieux définir l'informatique comme science inscrite dans une histoire et une société. Entre les deux, le chemin sera long et ce livre est un pas.

Ch1. Les *algorithmes* du Web

Une classification des *algorithmes* du web^a, les *algorithmes* :

- à côté du web (Google Analytic, Médiamétrie, ...)
qui s'intéressent à la popularité des sites
- au dessus du web (PageRank de Google)^{i(b)}
qui analysent la structure du web pour inférer de la "méritocratie"
- à l'intérieur du web (FaceBook, Twitter)^{i(c)}
qui évaluent l'activité à l'intérieur des réseaux (likes, retweets, ...)
- au dessous du web (Recommandation Amazon, ...)
qui calculent à partir des traces laissées par les utilisateurs

Chacune de ces positions relatives au web permet de comprendre les enjeux sociétaux et commerciaux des outils proposés. En creux se dessine une perspective épistémologique, de la simple consultation d'une page à une analyse plus complexe basée sur la navigation.

^a. L'auteur utilise le terme *algorithme* de manière générale, mais il focalise son argumentation sur les *algorithmes* du web.

^b. https://interstices.info/jcms/c_47076/comment-google-classe-les-pages-web

^c. https://interstices.info/jcms/ni_79095/1-algorithme-edge-rank-ou-le-filtrage-selon-facebook

Ch2. La révolution du Big Data

La montée en puissance du traitement et l'explosion de la production de données (les *big data*) modifient la relation société / numérique qui a vécu trois changements :

1. Le choix des descripteurs d'une activité entraîne une modification du comportement des individus qui cherchent à s'adapter à la mesure. Un exemple est le site *nosdéputés.fr* qui relaie l'activité parlementaire. Son apparition a (naturellement) poussé les élus à améliorer leurs scores, sans que nous puissions constater une augmentation qualitative de l'activité du parlement.
2. L'augmentation de la capacité de mesure permet de se concentrer sur des descripteurs de plus en plus fins, jusqu'à ne contenir qu'un nombre très réduit d'individus. Il devient alors difficile de définir l'appartenance de l'individu à des catégories particulières, ou de les hiérarchiser entre-elles. Le risque est alors que chaque expression vaille celles des autres.
3. L'application massive des modèles mathématiques (numériques) permet de découvrir (aléatoirement) des corrélations. Mais il ne suffit pas de les interpréter comme des "causes probables", il faut aussi leur donner du sens, en mobilisant des théories et des modèles venus des sciences humaines.

C3. Signaux et traces

Ce chapitre se concentre sur la définition des signaux (informations explicitement délivrées par l'utilisateur : tweet, commentaire, ...) et des traces (temps de consultation d'une page, chemin pour parvenir à une information, ...). Actuellement, les *algorithmes* considérés comme les plus performants sont capables de profiter à la fois des traces et des signaux.

Ces systèmes prédisent un comportement par similarité de régularités précédemment observées, soit parce qu'ils influent sur les comportements en produisant une régularité. La mesure de leur efficacité est difficile. Il faut retenir l'exemple de la publicité où une campagne qui augmente de 100% sa productivité fait passer de 0,01 à 0,02% de clic par page. Mais ici, les *algorithmes* se focalisent sur le passé.

L'auteur explique aussi que la peur des *algorithmes* est légitime, mais qu'il est bien plus important que nous, en tant que citoyens, demandions aux programmes d'être loyaux. Si un programme est présenté comme réalisant une tâche, il ne doit faire que cela. C'est le seul moyen de ne pas laisser croire que l'informatique est biaisée et assujettie à des intérêts fantasmatiques.

Ch4. La société des calculs

La dernière partie du livre ouvre vers une lecture plus politique de la relation aux *algorithmes*. L'auteur revient sur la viralité de l'attention sur Internet et aussi sur l'auto reproduction du centre du réseau (plus un utilisateur est au centre d'un réseau, moins il a de chance d'en sortir). Mais le salut vient de la dernière partie de ce chapitre où l'auteur rappelle que les résultats des programmes informatiques ne sont généralement pas considérés comme de très grande qualité, et que les individus aiment à jouer avec de multiples plateformes en même temps, ce qui complexifie d'autant l'étude automatique de leur comportement, voire qu'ils continuent de considérer que leur vie se déploie en dehors des univers virtuels.