# «Les algorithmes créent leur propre réalité»

**La mathématicienne Cathy O'Neil met en garde contre l'absence de transparence entourant les logiciels utilisant l'IA. Selon elle, un regard critique s'impose pour comprendre le fonctionnement de ces boîtes noires.**
Par Erwan Cario. Photo: Fred Kihn.

**Entretien.** Cathy O'Neil est une lanceuse d'alerte. Mais le scandale qu'elle dénonce est global et presque invisible tout en se déroulant sous nos yeux. Aujourd'hui, les modèles mathématiques et les algorithmes prennent des décisions majeures, servent à classer et catégoriser les personnes et les institutions, influent en profondeur sur le fonctionnement des États sans le moindre contrôle extérieur. Cathy O'Neil connaît bien le sujet. Après des études en mathématiques, elle travaille pour la finance jusqu'à la crise de 2008, puis se réoriente vers la science des données avant de se rendre compte que les mêmes mécaniques y sont à l'œuvre. Elle a publié en 2016 *Weapons of Math Destruction* (jeu de mots jouant sur la prononciation en anglais de « maths » et de « mass ») qui vient de sortir en français sous le titre *Algorithmes : la bombe à retardement*.

**Les mathématiques font peur, et en même temps on leur fait une confiance aveugle. C'est paradoxal, non ?**
Oui, c'est pratiquement la façon dont nous abordons Dieu. Quand une de mes amies, par exemple, a demandé des détails sur le modèle appelé « plus-value » utilisé pour la notation des enseignants, on lui a répondu : « *Ce sont des maths, vous ne pourrez pas comprendre.* » On le lui a dit quatre fois, quatre personnes différentes… Pourquoi ces quatre personnes lui ont-elles dit exactement la même chose ? Parce que, la plupart du temps, ça marche. C'est un bouclier très puissant pour se protéger de la curiosité de la population. Ayez confiance, ne posez pas de questions, et surtout sentez-vous honteux car vous n'êtes pas à la hauteur pour poser des questions.

**Du coup, comment expliquer la manière dont fonctionne un modèle mathématique ?**
Il faut faire exactement le contraire de ce que je viens de décrire et ne pas se dissimuler derrière une entité divine intimidante. J'explique aux gens que c'est quelque chose que nous faisons tous les jours. Comment est-ce que je m'habille ? Qu'est-ce



qu'on fait à manger ce soir ? Quel film est-ce qu'on regarde ? On utilise tous des algorithmes dans nos têtes pour prédire la réussite. La différence, c'est qu'on optimise notre propre algorithme. On décide si le film qu'on a voulu voir a été intéressant ou non. Si on l'a aimé, on va faire un peu plus confiance à notre intuition, sinon, on va se demander ce qui n'a pas marché et on va modifier notre algorithme pour l'avenir.
On contrôle les conditions du succès. La différence entre les modèles qu'on utilise dans nos têtes et ceux dont je parle dans le livre, c'est que des entreprises privées avec des intérêts commerciaux définissent leurs conditions de succès en nous ciblant. Et elles vont nous refuser des opportunités selon leur propre définition secrète de la réussite.

**Vous dites que les algorithmes sont des opinions intégrées à du code…**

Oui, car il y a toujours une définition des conditions de succès pour la personne à qui appartient l'algorithme. Et la question qu'on doit se poser, c'est : est-ce que ça correspond aussi à un succès pour moi, qui suis ciblée par ce programme ? Et nous avons des perspectives différentes, il n'y a pas de définition objective du succès.

**Ces algorithmes toxiques sont aujourd'hui omniprésents…**

Regardez le *news feed* de Facebook. C'est optimisé pour les profits de Facebook, mais il est probablement en train de détruire la démocratie. Et il est impossible de le mesurer précisément. Nous n'avons aucun contrôle.
Nous n'avons aucun pouvoir. C'est ridicule. Et il est difficile de mesurer la démocratie, par ailleurs. « *La démocratie a baissé de trois points hier.* » Qu'est-ce que ça voudrait bien dire ? Le sujet n'est pas de savoir si ce sont des gens malfaisants ; le sujet, c'est que c'est en train d'éroder notre concept de la vérité. C'est pour ça que nous avons besoin de plus de transparence.

**Mais avec les modèles fondés sur les réseaux de neurones qui ingèrent des grosses quantités de données, on est réellement face à des boîtes noires… Peut-on vraiment les auditer ?**

Bien sûr que oui ! Je ne dis pas que c'est facile, mais c'est possible si on a accès aux profils qui sont ciblés par un algorithme de ce type. Il faut observer comment cette population est traitée. Si par exemple, c'est un algorithme de recrutement, on va regarder comment il va se comporter avec des profils de femmes qualifiées. Est-ce qu'elles vont passer le filtre aussi souvent que les hommes qualifiés ? Il faut bien sûr définir « qualifié », et c'est compliqué. Je ne m'intéresse pas au fonctionnement interne de la boîte noire, je ne m'occupe que du résultat. Et il faut faire le même test pour toutes les boîtes noires qui opèrent des filtres de ce genre.
Ça n'a rien à voir avec la complexité mathématique qui est en jeu dans le fonctionnement même du processus. C'est ce que les experts en données voudraient vous faire croire, que c'est si compliqué que vous n'êtes même pas en mesure de poser des questions.

**Vous évoquez la nécessité de mettre en place l'équivalent d'un serment d'Hippocrate pour la science des données…**

Tous les experts en données devraient avoir conscience de l'importance de l'éthique. Mais, à ce jour, je n'ai pas lu de texte de ce genre assez fort pour que je le signe. Tout le monde propose sa liste, mais aucune ne se réfère spécifiquement aux droits de l'homme ou aux lois constitutionnelles. On devrait se concentrer là-dessus. Mais je pense que ça devrait exister. Avant de le signer, les gens seraient obligés de prendre en compte leur responsabilité éthique. Mais ce ne sera pas suffisant. Ce que je ne veux surtout pas voir, c'est la continuation de l'approche actuelle des *Big Data*, où les experts des données deviennent *de facto* des experts de l'éthique.

**C'est-à-dire ?**

Il existe par exemple un algorithme qui aide à la décision concernant le risque de maltraitance des enfants. Il doit établir le risque pour un enfant. Il y a plein de données en jeu, plein de choses qui peuvent mal tourner, plein de particularités à prendre en compte… Et il faut que l'enfant soit au centre des préoccupations. Mais au final, l'algorithme va aboutir à un score, un nombre qui va déterminer l'intervention ou non. Considérons alors ce que peut être un faux positif pour cet algorithme : il n'y a pas de risque pour l'enfant, mais il y a intervention et il est séparé de sa famille. Il ne va donc plus vivre avec ses parents. C'est une tragédie pour cet enfant, et pour la famille. Maintenant, un faux négatif : il y a maltraitance, mais personne ne va sauver l'enfant.
C'est aussi une tragédie. Pire que la précédente, bien sûr. Mais pire dans quelle mesure ? Que vaut ce « pire » ? Avec quelle valeur l'intégrer au score ? 3 ? 7 ? 12 ? C'est une question à laquelle personne ne veut répondre. C'est cette conversation compliquée que les algorithmes sont censés éviter. Aujourd'hui, l'expert des données construit l'algorithme et va y répondre, sans même le savoir. Il n'y pense pas, il n'est pas formé à l'éthique. Mais je ne veux pas qu'on se contente de donner une petite formation d'éthique aux experts et considérer qu'ils sont en charge du problème pour le reste de la société. Nous devons définir ce que ça veut dire pour un algorithme d'être responsable. Et, pour cet exemple, il faut avoir une conversation publique sur ce que devrait être le ratio. Et la responsabilité de l'expert des données sera de traduire dans le code, avec exactitude, la décision prise.

**Cet exemple peut se généraliser…**

Il existe plein de conversations que nous refusons d'avoir. C'est compliqué. On ne sait pas comment expliciter ce qu'est un bon professeur, on ne peut définir ce qui va faire d'un candidat un bon salarié, ni trouver des critères objectifs pour déterminer la qualité d'une université. Et on ne veut pas vraiment y réfléchir. On va donc appliquer des algorithmes qui vont se contenter de reproduire des pratiques passées en y intégrant des données multiples. Et on va affirmer de manière unilatérale que ça marche parfaitement. Ces algorithmes créent finalement leur propre réalité et les données utilisées en deviennent le socle. C'est de fait une opération de blanchiment des données. ○

(Paru dans *Libération* le 17 novembre 2018)