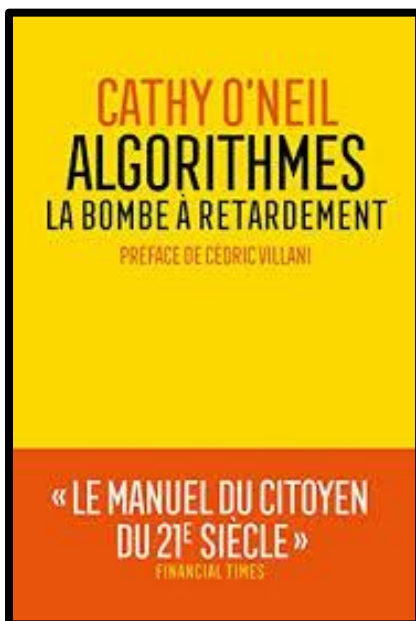


ALGORITHMES LA BOMBE A RETARDEMENT

O'Neil, Cathy. 2018. Les Arènes, 340 p.

Anne-Sarah Amuah, Charlotte Gaumer, Margot Fleick, Victoire Torres



Cathy O'Neil est une passionnée de mathématiques depuis son enfance et est une ancienne mathématicienne et analyste à Wall Street. Elle est devenue une figure importante de la lutte contre les combinaisons des algorithmes et leur puissance de calcul. En effet, elle est militante du mouvement citoyen Occupy Wall Street et met en garde contre les dangers et les impacts des algorithmes comme outils de domination sociale. Elle rend compte, de part son travail, que les mathématiques sont devenues complices des pouvoirs publics, de la justice, de l'éducation, de l'accès à l'emploi ou au crédit ainsi que des grandes firmes, à travers la collecte voire même le commerce des données personnelles.

L'économie du Big Data¹ est composée de ce qu'elle appelle des "Armes de Destruction Mathématique" (ADM), une métaphore des armes de destruction massive car elles ont le même effet sur la société, d'après l'auteure. Ces ADM sont un véritable jeu de profit. Les modèles mathématiques utilisés encodent les partis pris humains, les préjugés et ne sont pas utilisés avec la meilleure intention. Les ADM représentent un véritable danger car ils

¹ Mégadonnées désignent une grande quantité de données générées par tous les utilisateurs du monde entier.

creusent le fossé entre les privilégiés et les exclus, ce sont des boucles de rétroactions qui accentuent la fracture sociale dans notre société.

LA THESE DE CATHY O'NEIL

Les mathématiciens et statisticiens étudient nos désirs, nos déplacements, notre pouvoir d'achat, le niveau de confiance que l'on nous accorde et notre potentiel en tant qu'étudiant, travailleur, amant, acheteur ou criminel. En bref les formules mathématiques choisissent pour nous notre université, notre assurance, nos professeurs, notre vote lors des élections et décident ou non de nous accorder un crédit. A cause des ADM, beaucoup d'hypothèses dommageables et discriminantes sont camouflées par les mathématiques, c'est l'économie des données effrénées. Les modèles mathématiques sont mal conçus et contrôlent aujourd'hui les moindres aspects de l'économie : publicité, gestion des établissements publics, gestion des prisons, par exemple.

Les probabilités basées sur des calculs mathématiques consistent à créer un modèle de représentation arbitraire d'un processus quelconque, par exemple :

- Un modèle informel : une mère de famille qui prépare le repas en fonction du régime alimentaire de sa famille.
- Au niveau du sport : calculer les probabilités de gagner un match de baseball en mesurant les trajectoires des balles et les déplacements des joueurs, c'est la stratégie « moneyball ».

Les ADM définissent leur propre réalité et l'utilisent ensuite pour justifier leurs résultats. Ce type de modèle s'auto justifie et s'auto perpétue. Le traitement des données accroît les inégalités et menace la démocratie. Pour Cathy O'Neil, les coupables sont les acteurs de systèmes pervers dans des secteurs divers tels que la justice, l'éducation, l'enseignement supérieur, les banques, le crédit ou la planification des horaires de travail.

Les ADM sont utilisés à partir du moment où un phénomène est étudié à grande échelle : nationale ou internationale. Or les phénomènes sociaux sont difficiles à mesurer parfaitement car ce ne sont pas des données scientifiques, autrement dit, quantifiables. De plus, les données des algorithmes sont souvent elles-mêmes erronées. On bascule alors dans le "côté obscur du Big Data" (p. 30).

Les modèles mathématiques sont programmés en fonction des profits recherchés, de l'objectif définit par une entreprise. Le problème est que l'objectif recherché finit par générer sa propre vérité en s'auto justifiant à travers des modèles mathématiques. Les programmes traitant de grandes quantités d'informations sont condamnés à se méprendre sur un certain pourcentage d'individus, à les classer dans les mauvais groupes et ainsi leur refuser un emploi ou la chance d'acquérir la maison de leurs rêves. Mais en

règle générale, les personnes qui manient les ADM ne s'attardent pas sur leurs erreurs. Leurs systèmes sont conçus pour absorber toujours plus de données, affiner leurs outils d'analyse et rapporter ainsi toujours plus d'argent. Les investisseurs se nourrissent de ces profits et financent généreusement les entreprises qui conçoivent les ADM.

Les experts qui travaillent dans ces entreprises de Big Data estiment qu'aucun système statistique n'est parfait mais que les erreurs sont simplement des dommages collatéraux. Or les dommages collatéraux de ces systèmes imparfaits forment des dégâts importants et des injustices sociales, qu'ils perpétuent. Tous les aspects de la vie sont de plus en plus contrôlés par des modèles mathématiques, tenus secrets, qui attribuent des punitions arbitraires. En effet, il existe un choix stratégique voire même politique derrière ces algorithmes.

Nous allons citer quelques exemples qui illustrent ces injustices : entrer à l'université, emprunter de l'argent, purger une peine de prison ou trouver un emploi.

DISCRIMINATION DES ADM POUR L'ACCES AUX CREDITS, AUX PRETS ET A L'EMPLOI

Dans son livre, l'auteure nous montre un autre aspect discriminant des ADM avec les crédit-scores, qui rendent difficile l'accès à des prêts immobiliers (ou bien à des emplois).

Tout d'abord, elle rappelle que les préjugés concernant l'attribution des prêts ne sont pas récents ; en effet, à l'époque où seuls les banquiers contrôlaient l'argent (dans les années 1950), ils se basaient sur les habitudes des citoyens pour leur en accorder un, ou non. Par exemple, si vous alliez à l'église ou si vos relations avec votre patron étaient bonnes, il était plus facile d'obtenir un prêt que si vous faisiez partie d'une minorité ethnique ou si un membre de votre famille avait des démêlés avec la justice. Afin d'améliorer cette situation, Earl Isaac et Bill Fair inventèrent le modèle FICO (qui ne se basait pas sur les origines ethniques mais se concentrait uniquement sur le risque de défaut de paiement de l'emprunteur), qui persiste encore aujourd'hui, même si un modèle plus cruel d'ADM est plus répandu : les e-scores (ou scores de crédit). Ce deuxième modèle est « arbitraire, obscur, non réglementé et souvent injuste » selon l'auteure, qui le caractérise d'ADM. Dans ce modèle, différentes données rentrent en compte dans l'obtention d'un prêt, qui serviraient à prédire la solvabilité de l'emprunteur. Mais, lorsqu'on se penche réellement sur ces données utilisées, on se rend compte que certaines n'ont rien à voir avec le fait d'être solvable ou non. Par exemple, les e-scores prennent en compte le code postal, l'historique internet, ou encore les achats antérieurs d'un individu. On se rend alors compte que cet ADM attribue les prêts selon certains préjugés (ici, le code postal pourrait révéler que vous habitez dans un quartier mal famé et donc que vous faites partie d'une minorité ethnique et que vous ne pourrez donc pas rembourser votre prêt). En se basant

sur des préjugés, l'ADM accordera alors à l'individu un crédit à taux d'intérêts très élevés par exemple, alors même que ce dernier a déjà du mal financièrement parlant. Le citoyen n'est alors pas analysé comme individu mais plutôt comme un membre d'un groupe, à cause des préjugés qui découlent en généralités.

Dans son livre, l'auteure nous apprend qu'en plus des préjugés des e-scores concernant l'obtention d'un prêt immobilier, ils jouent également un rôle sur le recrutement. En effet, les employeurs viennent se renseigner sur les scores de crédit des individus (avant même de les rappeler pour un entretien par exemple), pour savoir s'ils vont les recruter ou non. L'auteure nous apprend qu'une loi visant à interdire l'utilisation des scores de crédit dans le recrutement est en vigueur dans 10 états, ce qui montre que même s'il y a eu du progrès, il reste 40 états à améliorer.

L'utilisation des scores de crédit dans le monde immobilier ou le monde de l'emploi ne fait qu'aggraver la pauvreté et l'endettement, en effet, c'est un réel cercle vicieux pour les plus pauvres et les minorités (surtout les personnes de couleur et les femmes). On retrouve bien ici une des caractéristiques de l'ADM qui est le manque d'équité.

On voit que les instances de régulation veulent de nouvelles lois pour encadrer le marché des données personnelles aux Etats-Unis, pour venir compléter celles déjà en place mais qu'il faudrait moderniser. Par exemple, il existe le Fair Credit Reporting Act de 1970, qui vient limiter la collecte de données concernant les informations de crédit des individus, pour plus d'équité et de confidentialité, mais aussi l'Equal Credit Opportunity Act, datant de 1974 donnant l'opportunité à tous les citoyens d'obtenir un prêt tout en ne prenant pas en compte le sexe ou les origines ethniques lors de l'établissement du score de crédit. On voit que ces lois existent depuis longtemps, cependant elles ne semblent pas être respectées par les ADM aujourd'hui qui, comme le dit l'auteure dans son livre, se basent sur des données qui sont en fait le moyen caché d'obtenir l'origine ethnique ou le sexe d'une personne pour établir son score de crédit (par exemple, le code postal).

Ces ADM existent depuis longtemps dans le secteur bancaire : à travers les banquiers en 1950, la crise des subprimes en 2008, ou encore de nos jours.

Afin d'illustrer les dégâts des ADM, Cathy O'Neil nous présente le cas de la réforme scolaire de 2007 à Washington. Le nouveau maire, dans un souci d'améliorer les écoles les moins performantes de sa ville, fait appel à Michelle Rhee, experte des réformes éducatives. Le problème, selon l'avis général, était que les élèves n'apprennent pas correctement parce que les professeurs sont incompetents. En 2009, Michelle Rhee va donc lancer un plan de réforme visant à éliminer les enseignants les moins performants.

Pour cela, un outil d'évaluation appelé "IMPACT" est mis en place. En deux ans, 206 enseignants furent renvoyés. Parmi eux, une institutrice du nom de Sarah Wysocki, qui

enseignait à la MacFarland Middle School depuis deux ans et qui apparemment n'avait aucune raison de s'inquiéter d'un éventuel licenciement, puisqu'elle recevait déjà d'excellents avis de la part de son directeur et des parents d'élèves. Elle obtint cependant une note déplorable lors de son évaluation "IMPACT". Le problème venait du nouveau système de notation connu sous le nom de "modèle de la valeur ajoutée", qui réduit au minimum les critères d'appréciations humains (subjectif) et met l'accent sur les résultats concrets des élèves (leur réussite en mathématique et en lecture, par exemple). Sarah Wysocki trouva ces résultats injustes et décida de s'intéresser de plus près à cette notation. Au cours de ses recherches, elle découvrit que le district de Washington avait engagé le cabinet de conseil Mathematica Policy Research de Princeton pour concevoir le système d'évaluation. Les chercheurs ont conscience que mesurer le progrès d'un élève prend en compte de nombreuses variables, comme les troubles de l'apprentissage, et prendre tous ces facteurs en compte n'est pas une mince affaire. Les modèles utilisés par Mathematica laissent de côté beaucoup de facteurs et se basent sur un échantillon trop restreint pour que les résultats soient fiables. Un autre facteur important relevé par l'auteur est que, pour la réussite de tels systèmes statistiques, il faut un feed-back afin d'actualiser le modèle en fonction de la réalité. Ce qui n'est pas le cas du modèle de la valeur ajoutée, qui, comme de nombreuses ADM, définit sa propre réalité et l'utilise ensuite pour justifier ses résultats. Il n'y a donc aucun moyen de vérifier l'efficacité de cette méthode puisqu'il n'y a aucun retour. Les algorithmes d'évaluation entraînent une forte motivation des enseignants concernant la réussite de leurs élèves, quitte à truquer les résultats des examens, ce qui vient alors fausser les résultats des évaluations. Il est donc possible que ceux de Sarah Wysocki aient débuté l'année avec des scores artificiellement gonflés, ce qui expliquerait pourquoi elle a obtenu un score médiocre.

Un exemple similaire concernant les universités est également présenté dans l'ouvrage. En 1983, le magazine US News & World Report décide de faire un classement des meilleures universités sur la base de nouveaux critères d'évaluation, afin de permettre à des millions de jeunes de prendre la meilleure décision.

Il publia alors un premier classement, fondé sur des analyses de données, en 1988, et les résultats paraissaient sensés. Ce classement, alors devenu une référence nationale, commence à avoir des effets néfastes car il s'auto-renforçait. Par exemple, si une université était mal classée, sa situation se dégradait, les meilleurs élèves ainsi que les meilleurs professeurs faisaient en sorte de les éviter. Le problème se posa de maintenir ou améliorer leur place dans le classement en fonction de 15 critères définis par le magazine US News & World Report, les universités tentent donc désormais d'augmenter le nombre de diplômés et de réduire le nombre d'étudiants par classes. Les universités se sont mises à leurrer le système et à engendrer des faux positifs. Un second problème est que l'un des critères le plus pris en compte était la réputation, ce qui est complètement

absurde car elle est complètement subjective. Les manoeuvres pour tenter d'améliorer les scores ne font que se multiplier et les universités les moins côtés en pâtissent.

ADM ET TESTS DE PERSONNALITE : LES CAS DE LA PRISON ET DE L'EMPLOI

L'une des principales dérives des ADM est la détermination de types de personnalités. En fonction des données collectées sur une personne ou de sa réponse à un questionnaire donné, il serait possible de déterminer certains aspects de sa personnalité. Dès lors, on pourrait identifier certaines prédispositions qui le conduirait à mener une action plutôt qu'une autre. L'algorithme sert à identifier des caractéristiques chez les personnes étudiées et à déterminer, à partir de ces informations, comment elles agissent ou pourraient agir dans des situations données.

Les Big Data ont en premier lieu été utilisés dans une logique pénale. L'objectif était de prédire quels crimes auraient lieu, dans quel endroit et quand. Pour ce faire, la police a dû déterminer qui étaient les criminels en général. En collectant des données, la police a pu établir que les criminels avaient une allure spécifique (accoutrement, démarche, tatouages, etc...). La police, à partir de ces critères, a adopté le "stop and frisk" : elle interpellait les individus correspondants à ces caractéristiques pour les fouiller ensuite. Cependant, cette mesure d'origine citoyenne, posait problème en termes d'égalité : 85% des jeunes interpellés étaient en effet des jeunes d'origine hispanique ou afro-américaine, alors que, dans la majorité des cas, ces personnes ne s'apprêtaient aucunement à commettre un crime.

L'initiative est allée encore plus loin : les algorithmes n'ont pas seulement cherché à savoir qui étaient les criminels, mais également la durée de la future peine de prison ou encore lesquels d'entre eux récidiveront. A partir de certains critères, tel que le quartier d'origine du criminel, les algorithmes ont tenté d'établir la liste des criminels récidivants. Ces critères, choisis subjectivement, remettent en question la pertinence des conclusions que formulent les algorithmes.

Cependant, ces algorithmes conçus dans une logique d'égalité, perpétuent pourtant des inégalités. Selon leur éducation par exemple, les candidats auront tendance à adopter une réponse plutôt qu'une autre. Le questionnaire est révélateur des préjugés, et a notamment tendance à pénaliser ceux qui sont issus de quartier défavorisés, sous prétexte qu'ils répondront avoir été confrontés à la police à plusieurs reprises par exemple. Ces tests échouent à proposer une politique pénale égalitaire : certains profils sont plus entravés par ces tests et jugés dangereux, et ont par conséquent des peines plus lourdes, en fonction de critères arbitraires.

La discrimination par les tests de personnalité a également lieu dans le cadre de l'emploi. Pour l'illustrer, l'auteur prend le cas d'un brillant étudiant nommé Kyle. Kyle souhaitait avoir, en parallèle de ses études, un emploi à mi-temps en grande surface. Chaque candidat devait passer un test de personnalité établi par l'entreprise Kronos. Kyle, rétabli d'anciens troubles psychiatriques, reconnaissait pourtant ce test comme étant similaire à celui des "5 facteurs" qu'on lui avait fait passer à l'hôpital. L'objectif de ce test est d'évaluer la personnalité de la personne interrogée à partir de 5 caractéristiques : "la propension à être extraverti, agréable, consciencieux, névrotique, ou ouvert d'esprit" (p.163). A ce poste, Kyle fut refusé. Il postula à d'autres postes similaires et la réponse était systématiquement négative. Kyle décida de porter plainte : il considère que ce test, d'origine médicale, en étant utilisé dans les recrutements, s'opposait à la loi contre les discriminations relatives au handicap.

Pourtant, initialement, ce test avait pour objectif de rendre le processus de recrutement plus égalitaire, en empêchant que le recruteur jauge le candidat selon des critères subjectifs lors d'un entretien. La pratique des tests de personnalité se banalise ; elle affiche notamment une croissance de 10 à 15% par an. De plus, l'auteure questionne l'efficacité de ces tests de personnalité : ils ont, en effet, une très faible valeur prédictive. Une étude de Franck Schmidt, un professeur de l'université de l'Iowa, montre que ces tests sont trois fois moins efficaces qu'un examen cognitif par exemple. Dans le cas de Kyle, sans ces tests de personnalité il n'aurait eu aucune difficulté à l'emploi. Cependant, en raison de son ancienne maladie, sa réponse au questionnaire ne lui permettait pas d'être recruté. Là encore, plutôt que de permettre l'égalité, ces tests de personnalité enracinent des discriminations.

S'il est admis que les tests de personnalité ne résolvent pas les problèmes qu'ils sont censés régler, leur démocratisation progresse en raison du fait qu'ils soient bien plus économiques que toute autre forme de solution. Ces tests ont tendance à se généraliser, en dépit de leur nature inégalitaire et de leur efficacité questionnable. Plus grave encore pour l'auteur, ces tests sont jugés objectifs, car ils sont issus des mathématiques : ainsi, leurs conclusions font office de vérité. En conséquence, la pratique de ces tests ne permet pas de restructurer des institutions afin de les perfectionner, mais permet surtout de cristalliser des inégalités, lesquelles n'ont pas de fondement logique.

CONCLUSION

Pour les raisons telles que la discrimination ou les préjugés, l'auteure propose que les Etats-Unis imitent le modèle européen et le RGPD, afin de garantir plus d'équité et de transparence aux citoyens, qui sont évalués au quotidien. De plus, il faudrait mettre en place des normes plus strictes pour les data scientists qui viennent analyser ces données, sans même chercher à comprendre si elles sont injustes ou non. Néanmoins, il existe des

lois, notamment pour les algorithmes liés au secteur de la santé comme le Americans with Disabilities Act ou le Health Insurance Portability and Accountability Act qui interdisent la discrimination, surtout pour le recrutement, basée sur les tests de personnalité, qui permettent de savoir, par les questions posées de manière implicite, si une personne est handicapée ou non. Ces algorithmes peuvent permettre aux recruteurs de refuser un candidat, en se basant sur des critères discriminatoires.

Ce livre est intéressant car on découvre comment fonctionnent les algorithmes qui rythment notre quotidien et les dégâts qu'ils peuvent causer. Les exemples cités dans le livre nous permettent de mieux comprendre les failles et les injustices qu'ils perpétuent. Cependant, l'auteur se répète beaucoup et utilise plusieurs fois les mêmes exemples dans différentes parties, ce qui devient lassant. L'aspect redondant des conséquences des algorithmes dans chaque exemple se retrouve à de nombreuses reprises dans le livre : ce qui allonge les chapitres et essouffle la lecture. En dépit de ce côté répétitif, dans l'ensemble, nous avons toutes bien aimé ce livre !

#CulturesNum est un programme réalisé par les étudiants du **Master Communication des Organisations de l'Université Bordeaux Montaigne** sous la direction d'**Aurélie Laborde**, en collaboration avec **UNITEC**. Depuis 2016, des ouvrages récents sur la société numérique sont synthétisés pour mieux appréhender les questions de fond pour notre société : big data, smart-cities, post humanisme, avenir du travail et de la consommation, etc...