« LES ALGORITHMES REFLÈTENT LE SEXISME DE LA SOCIÉTÉ »

Entretien avec **AUDE BERNHEIM ET FLORA VINCENT /**

CO-FONDATRICES DE WAX SIENCE,
AUTEURES DE *L'INTELLIGENCE ARTIFICIELLE, PAS SANS ELLES !*
(BELIN/LABORATOIRE DE L'ÉGALITÉ, 2019)

Propos recueillis par Marie-Cécile NAVES /
DIRECTRICE DE RECHERCHE À L'IRIS

SEPTEMBRE 2020

OBSERVATOIRE GENRE ET GÉOPOLITIQUE

IRIS
INSTITUT DE RELATIONS INTERNATIONALES ET STRATÉGIQUES

Aude Bernheim et Flora Vincent sont toutes deux docteures en sciences, formées respectivement à l'Institut Pasteur et à l'École Normale Supérieure. Pendant leurs études au Centre de Recherche Interdisciplinaire, elles ont cofondé WAX Science, une association de promotion des sciences et des femmes dans la science, à travers le développement et la diffusion d'outils innovants, à l'interface du numérique, du design, des sciences et de l'égalité femmes-hommes.

Elles ont publié en 2019 l'ouvrage L'intelligence artificielle, pas sans elles !, aux éditions Belin/Laboratoire de l'égalité (préface de Cédric Villani). Elles répondent aux questions de Marie-Cécile Naves, directrice de l'Observatoire genre et géopolitique de l'IRIS.

MARIE-CÉCILE NAVES : **En quoi, et comment, l'intelligence artificielle est-elle « aveugle » aux questions de genre ? Quelles en sont les conséquences ?**

AUDE BERNHEIM ET FLORA VINCENT : L'intelligence artificielle (IA) correspond à un ensemble de techniques mathématiques et informatiques permettant de résoudre des problèmes de façon automatisée. Ces dernières années, l'émergence de nouvelles technologies comme la *machine learning* et la disponibilité de données en grand nombre ont permis des progrès fulgurants trouvant leurs applications dans de nombreux secteurs : banque, justice, hôpitaux, etc.

Les algorithmes n'ont pas d'intentions en soi, ils reflètent la société et les points de vue de ses créateurs. Or, notre société est sexiste. Des algorithmes de traduction, lorsqu'ils passent d'une langue non genrée (comme le turc) à une langue genrée, proposent des

associations : par exemple, une personne célibataire devient un homme célibataire alors qu'une personne mariée, une femme mariée. Ces stéréotypes peuvent devenir particulièrement injustes lorsqu'il s'agit de trier des CV (jetant systématiquement ceux de femmes pour des postes techniques) ou de proposer des salaires (automatiquement moins élevés pour les femmes). Les IA reproduisent les préjugés genrés de notre société, les propagent et les amplifient.

En étudiant quelques exemples, on peut mettre au jour et comprendre les mécanismes de contagion sexiste et/ou raciste des algorithmes. Un premier point consiste à s'intéresser au jeu d'apprentissage de ces algorithmes. Typiquement, lors de la création d'un algorithme, celui-ci est entraîné sur une base de données : photos ou CV, par exemple. Si, pour entraîner un algorithme à reconnaître des visages, la base de données contient plus de visages d'hommes blancs que de femmes noires, l'algorithme saura à la fin mieux reconnaître les visages d'hommes blancs que de femmes noires. Idem si, lorsqu'on cherche à recruter quelqu'un, on nourrit l'algorithme de CV plus masculins que féminins (typiquement en donnant les CV de personnes ayant occupé le poste précédemment ; si ces postes étaient majoritairement occupés par des hommes, l'algorithme apprendra à recruter des hommes). Il est donc nécessaire de faire un effort de diversification des contenus de bases de données, pour renverser les biais par ailleurs existants dans notre société.

Mais concevoir un algorithme ne se limite pas à « corriger » les bases de données. Les biais se nichent à d'autres étapes de leur création. Il faut par exemple questionner à travers le prisme du genre comment un objectif est transcrit en langage mathématique/informatique... Par exemple, comment trouver qui est le ou la meilleur.e étudiant.e pour un cursus ? Comment mathématiquement définir les critères : est-ce la moyenne des notes, la meilleure progression sur l'année, l'importance des matières ? Sans s'en rendre compte, on peut introduire des préjugés.

Ainsi, récemment, un article dans la revue *Science*[1] détaille l'existence de biais raciaux dans le domaine de la recommandation automatique, faite à certains patients, pour bénéficier d'un programme fédéral de soins aux États-Unis. Or, ce programme a été proposé plus facilement à des Blancs qu'à des Noirs. Pourquoi ? L'algorithme a cherché à reconnaître qui est le plus malade et, pour le définir, prend comme critère la hauteur des dépenses de santé de cette personne. Cependant, cette définition de l'état de santé entraîne un biais : il a été documenté que, pour un nombre de maladies similaires, les Noirs et les Blancs avaient des dépenses de santé différentes (inférieures pour les Noirs). C'est pourquoi, en choisissant ce critère des dépenses, l'algorithme a sélectionné des Blancs souffrant de moins de pathologies que certains Noirs. L'article de *Science* suggère de reformuler l'algorithme pour qu'il privilégie un autre critère. Une fois l'algorithme entraîné, entre en ligne de compte la performance de celui-ci, ses paramètres de précision, comme les faux négatifs ou faux positifs. On peut lui demander de garder un taux de faux négatifs faible, c'est un choix que l'on fait en amont qui lui aussi peut avoir des conséquences inattendues en termes de biais sexistes et racistes.

Dès lors, qu'est-ce qu'un bon algorithme, non biaisé ? Ce n'est pas qu'une question informatique et aujourd'hui la recherche est à l'intersection de plusieurs disciplines : philosophie, mathématique, informatique …

MARIE-CÉCILE NAVES : Quelle est la place des femmes dans l'IA ?

AUDE BERNHEIM ET FLORA VINCENT : Il y a encore très peu de femmes dans l'IA avec aujourd'hui seuls 12 % des employés du secteur qui sont des femmes. Cette rareté est plus criante que dans les sciences en général, alors que l'informatique était au départ un domaine très féminisé. Que s'est-il passé ? Plusieurs hypothèses peuvent expliquer cela. D'une part, la dimension culturelle d'un marketing très genré a découragé les femmes et

1 Ziad obermyer, brian powers, christine vogeli, sendhil mullainathan, « dissecting racial bias in an algorithm used to manage the health of populations », *science,* vol. 366, n° 6464, octobre 2019, p. 447-453.

encouragé les hommes : le message selon lequel « les ordinateurs sont pour les hommes », l'imaginaire très stéréotypé du geek, etc. D'autre part, lorsque l'informatique est devenue lucrative, les femmes en auraient été exclues petit à petit. Les progrès sont faibles malgré les nombreuses initiatives pour promouvoir les femmes, car les stéréotypes de genre sont difficiles à déconstruire, en particulier dans l'IA. Aujourd'hui encore, les dirigeants les plus connus de l'IA sont des hommes, comme Cédric Villani ou Yann Lecun en France ou Elon Musk ou Demis Hassabis à l'international.

Au-delà de la difficulté à pouvoir attirer de nouveaux talents, les femmes travaillant dans ce secteur sont découragées au fur et à mesure d'y participer, ce milieu souffrant de beaucoup de sexisme. Ainsi, une étude a montré que sur GitHub, une plateforme sur laquelle on échange du code, les contributions des femmes étaient mieux notées par la communauté que celles des hommes jusqu'à ce que les internautes apprennent qu'elles avaient été écrites par des femmes : ensuite elles étaient moins bien notées. Par sexiste et biais de genre, on se prive donc d'un code de meilleure qualité !

Mais le sexisme n'est pas une fatalité. Par exemple, l'« École 42 », quand elle a été fondée, s'appuyait sur une culture traditionnelle des geeks, en termes de modes de vie, d'interactions aux autres. La nouvelle directrice, Sophie Vigier, a transformé le fonctionnement et la culture de son école pour qu'il y ait plus de femmes, pour mettre fin à l'entre-soi masculin.

Féminisation du domaine et production de biais sexistes sont-elles liées ? En partie oui. De nombreux lanceurs d'alerte sur les biais des algorithmes sont des individus différents de la majorité dans le secteur. En testant sur eux-mêmes les algorithmes, ils observent que ceux-ci fonctionnent moins bien. En intégrant plus de diversité dans les équipes, les probabilités sont plus grandes de faire émerger en amont, pendant leur conception, des biais potentiels. Ainsi, plusieurs études ont montré que dans des publications

scientifiques, les équipes mixtes prenaient plus en compte les questions de genre et de « race ». Cependant, une mixité dans les équipes ne garantit pas du tout une absence de biais, notamment car les femmes et les hommes peuvent n'être pas conscients des préjugés qu'ils transmettent. Le processus de conscientisation, par la formation, est essentiel.

MARIE-CÉCILE NAVES : **Comment combattre les stéréotypes et remédier à ces inégalités ?**

AUDE BERNHEIM ET FLORA VINCENT : Il n'est jamais trop tôt ni trop tard. Cela commence par mettre des ordinateurs dans les mains des filles, dès le plus jeune âge. L'apprentissage du code doit démarrer le plus tôt possible. Il est aussi essentiel de former celles et ceux qui créent des algorithmes, sans oublier celles et ceux qui vont les utiliser (managers et responsables RH, par exemple), qui doivent en comprendre les biais : la technologie n'est pas neutre, la distance critique est nécessaire.

Il est aussi possible d'aller plus loin et d'imaginer ce que seraient des algorithmes égalitaires. Pourrat-on utiliser des algorithmes pour faire avancer l'égalité ? On peut imaginer des logiciels de traduction automatique qui produiraient un langage inclusif. Il est aussi possible d'utiliser les capacités d'analyse des algorithmes pour promouvoir la prise de conscience d'inégalités. Par exemple, en 2016, Google a fait regarder à un algorithme les plus gros succès du box-office américain de cette année-là. Résultat : dans ces films, les hommes parlent deux fois plus que les femmes. En France, ce type d'analyses a été répliqué sur des programmes radio et télévisuel et arrive aux mêmes conclusions. La preuve par les chiffres entraîne une prise de conscience des inégalités, d'autant qu'il est très facile de réévaluer les chiffres l'année suivante pour voir si ceux-ci évoluent après par exemple la mise en place de mesures dédiées. Enfin, en matière de recrutement, on peut forcer un algorithme à pousser à la parité, en demandant dès le départ, de sélectionner autant de CV de femmes que de CV d'hommes (et même d'aller chercher

d'autres profils sur LinkedIn, par exemple). La question de se priver de talents pourrait être un moyen de convaincre les recruteurs et les recruteuses.■

« LES ALGORITHMES REFLÈTENT LE SEXISME DE LA SOCIÉTÉ »

Entretien avec **AUDE BERNHEIM et FLORA VINCENT** / CO-FONDATRICES DE WAX SIENCE, AUTEURES DE *L'INTELLIGENCE ARTIFICIELLE, PAS SANS ELLES !* (BELIN/LABORATOIRE DE L'ÉGALITÉ, 2019)

Propos recueillis par **Marie-Cécile NAVES** / DIRECTRICE DE RECHERCHE À L'IRIS

OBSERVATOIRE GENRE ET GÉOPOLITIQUE / SEPTEMBRE 2020
Sous la direction de Marie-Cécile Naves, directrice de recherche à l'IRIS.
naves@iris-france.org

L'Observatoire 'Genre et géopolitique' de l'IRIS a pour ambition d'être un lieu de réflexion et de valorisation de la recherche inter et pluridisciplinaire sur la manière dont le genre, en tant que concept, champ de recherches et outil d'analyse du réel, peut être mobilisé pour comprendre la géopolitique et être un outil d'aide à la décision sur des questions internationales.

© IRIS
Tous droits réservés

INSTITUT DE RELATIONS INTERNATIONALES ET STRATÉGIQUES
2 bis rue Mercœur
75011 PARIS/France

T. + 33 (0) 1 53 27 60 60
contact@iris-france.org
@InstitutIRIS

www.iris-france.org