

Mixité et intelligence artificielle

Peu de sociologues du genre se penchent sur la question du développement technique de secteurs innovants, comme la recherche. L'intelligence artificielle ne fait pas exception, et se doit d'être aussi questionnée.

Aude BERNHEIM et Flora VINCENT, membres du comité d'orientation du Laboratoire de l'égalité, cofondatrices de Wax Science*

Bien que le fameux « geek » s'apparente aujourd'hui plus à un stéréotype qu'à une réalité, il reflète une évidence concrète : l'absence de femmes dans le secteur de l'intelligence artificielle. Elles ne représentent que 10% des étudiantes et étudiants dans les cursus d'informatique à l'université, et ne sont que 11% dans les secteurs de l'informatique et de la communication en écoles d'ingénieurs. Les conséquences sont nombreuses, et dépassent largement la question des (in)égalités salariales ou de poste. Le domaine tout entier souffre d'un mal assez commun : le manque de pensée inclusive. Cette pensée s'assure que ce que l'on invente, développe, ou cherche à comprendre en vue de toucher le spectre le plus large possible d'individus ne laisse pas, en réalité, une partie de la population sur le bas-côté. Derrière cela se cache l'objectif de garantir que les conséquences de nos actions ou de nos produits vont inclure le plus grand nombre. Ce concept nous rappelle que l'autre est là, différent, et qu'il doit être pris en compte.

Lorsqu'un ingénieur - la trentaine, blanc, évoluant dans le milieu de l'intelligence artificielle - développe un algorithme de reconnaissance faciale, un de ses premiers réflexes sera de le tester sur lui, sur ses collègues, son entourage, et de l'ajuster en conséquence avant de décréter qu'il « fonctionne bien » ; il reconnaît efficacement les contours du visage, des lèvres, des yeux, la couleur des cheveux. En 2016, Joy Buolamwini, alors étudiante en thèse au MIT, utilise un de ces algorithmes et réalise qu'il bute à identifier son visage. Elle creuse la question, et réalise que l'algorithme est plus performant pour reconnaître des hommes blancs que des femmes à la peau sombre. La même année, Richard Lee, ingénieur né à Taiwan, se voit refuser ses photos d'identité sur un site administratif ; l'algorithme réclame qu'il a « les yeux fermés », alors que la photo

* Auteures du livre *L'intelligence artificielle, pas sans elles !*, Belin, mars 2019.

« **Le courant des « innovations genrées » prend racine dans la convergence entre une pensée féministe et des sciences dites « dures » comme l'ingénierie, les mathématiques, la technologie, la biologie.** »

montre le contraire. Il aurait peut-être suffi d'une Joy Buolamwini et d'un Richard Lee dans l'équipe des développeurs pour dire : « Nous devrions peut-être tester l'algorithme sur un visage de femme à la peau sombre et/ou un profil asiatique ? » Et ainsi éviter des discriminations algorithmiques. L'intelligence artificielle est, elle aussi, contaminée par les stéréotypes de nos sociétés et propage des inégalités. Reconnaissance faciale moins bonne pour les femmes, traduction automatique à contenu stéréotypé, propositions de salaires plus basses⁽¹⁾ et CV pénalisés automatiquement⁽²⁾... Les exemples dénonçant la contagion sexiste des algorithmes sont multiples.

A la source des discriminations algorithmiques

Les causes techniques de ces biais sont comprises et étudiées. Les bases de données utilisées pour calibrer et tester les algorithmes ne sont pas forcément représentatives du public ciblé par le produit en question : par exemple, utiliser une base de données avec 90% de visages d'hommes pour entraîner un algorithme de reconnaissance faciale. La subjectivité de la traduction de l'objectif d'un algorithme en formule mathématique est une source plus complexe mais non moins importante de biais. Si une personne développeuse doit créer un algorithme pour « sélectionner le meilleur ou la meilleure élève d'une classe », les traductions algorithmiques sont nombreuses. Est-ce l'élève qui a la meilleure note à l'examen final ? Ou la meilleure moyenne sur l'année ? Ou la meilleure progression sur une période donnée ? Si c'est un mélange entre les trois, en quelles proportions ? Ces choix sont politiques, sociaux, ils impliquent une part subjective et sont, dès lors, potentiellement biaisés, surtout si les personnes les effectuant ne sont

(1) Les offres d'emplois ciblées de Google Ad proposent des postes moins rémunérés aux femmes.

(2) On a découvert en 2015 qu'un programme pilote d'Amazon pénalisait des CV contenant le mot « femme », pour des postes techniques.



pas formées à débusquer les stéréotypes. Ceux des autres, mais aussi les leurs.

Une fois l'algorithme écrit, il s'agit de fixer les critères d'évaluation pour déterminer s'il est performant et équitable. Malheureusement, l'algorithme parfaitement juste et équitable n'existe pas, il faut donc fixer des seuils, trouver des compromis, entre taux d'erreurs acceptables, possibilité de généraliser à d'autres populations ou sujets, et rapidité d'exécution. Le milieu de l'informatique est de nature profondément collaboratif : cela garantit une homogénéisation et une reproductibilité des pratiques, notamment à travers la réutilisation des codes informatiques. La contrepartie, c'est la non-remise en question de certaines habitudes, qui induit une propagation d'erreurs : utiliser les mêmes bases de données biaisées, les mêmes compromis sans vérifier leurs conséquences, etc.

Pensée féministe et « sciences dures »

Ces dysfonctionnements sont longtemps restés sous silence, pour une raison simple : l'absence de questionnement féministe dans le développement même des sciences dites « dures ». Pas ou peu de réflexion sur l'innovation technologique ou scientifique qui intègre la variable du sexe et du genre. Qui se pose la question suivante : mon produit final touche-t-il des hommes et des femmes ? Et si oui, les hommes et les femmes sont-ils impactés différemment par ce produit ? Ce courant de pensée, né aux États-Unis et notamment porté par des scientifiques, comme Londa Schiebinger (Stanford), porte le nom d'« innovations genrées » (Gendered Innovations). Il prend racine dans la convergence entre une pensée féministe et des sciences dites « dures » comme l'ingénierie, les mathématiques, la technologie, la biologie. Ce courant montre l'importance d'étudier

L'intelligence artificielle est, elle aussi, contaminée par des stéréotypes de genre.

Reconnaissance faciale moins bonne pour les femmes, propositions de salaires plus basses, CV pénalisés automatiquement...

Les exemples dénonçant la contagion sexiste des algorithmes sont multiples.

les effets secondaires de médicaments sur des souris mâles et femelles, de réaliser que ces mêmes souris ne réagissent pas de la même manière quand elles sont manipulées par des scientifiques hommes ou femmes, que la façon d'optimiser des trajectoires de bus dans une ville change lorsqu'on sépare les données de déplacement par sexe, ou encore qu'il est nécessaire de créer des mannequins de « crash test » qui n'ont pas tous la morphologie d'un homme standard de soixante-dix kilos... Bien que fondamentale, il faut donc dépasser la problématique de l'absence de femmes en tant qu'actrices dans les secteurs scientifiques et intégrer cette variable du sexe et du genre dans l'implémentation même de ces technologies. Peu de scientifiques sont réellement formés sur l'impact de la variable du sexe et du genre, dans leurs recherches.

Sur ce sujet les lanceurs d'alerte commencent à se mobiliser. D'un œil critique, ils interrogent le contenu et l'éthique de ces outils surpuissants : comment l'algorithme a-t-il été pensé ? Comment est-il évalué ? Comment a-t-il été codé ? Que contient le jeu d'apprentissage ? Certains groupes sont-ils discriminés dans les prédictions effectuées ? Cet algorithme est-il équitable ? Qu'entendons-nous par équitable ? Les réponses sont à inventer dans les sciences sociales mais aussi dans les mathématiques fondamentales ou la physique. Arvind Narayanan, un informaticien de Princeton, explore par exemple vingt-et-une définitions de ce qu'est un algorithme équitable, et démontre comment chaque définition se place à l'interface entre une notion mathématique et une volonté politique. Ces questions sont le résultat d'une pensée inclusive, qui ouvre la voie à de nouvelles directions de recherche, en pas vers une société prête à utiliser sa diversité comme un moteur d'innovation et de créativité plutôt que comme un quota à remplir. ●