chapitre II

Pourquoi parle-t-on de discrimination en intelligence artificielle?

Résumé

1. Comment la discrimination peut-elle s'immiscer dans les systèmes d'intelligence artificielle?

Rappelons que c'est en analysant des exemples qu'un ordinateur peut dégager différentes tendances et structures pour faire des prédictions dans des contextes qu'il n'a jamais vus. Comme un enfant, plus il voit des données (ou des exemples), meilleur est son apprentissage. Au premier regard, le fonctionnement des systèmes d'intelligence artificielle peut sembler neutre et objectif. En effet, ceux-ci ne se basent que sur des données réelles et tous les processus d'apprentissage et de prédiction sont exécutés par un ordinateur. Cependant, il faut garder en mémoire que des êtres humains sont impliqués à toutes les étapes de la conception de ces systèmes. Étant donné qu'il est peu probable que ces interventions humaines soient complètement objectives et dépourvues de biais, les systèmes d'intelligence artificielle risquent donc d'incorporer ces préjugés à leur propre fonctionnement.

Il existe plusieurs façons dont les biais peuvent s'immiscer dans les systèmes d'IA. D'abord, si les données utilisées pour entraîner un système d'IA reflètent un contexte préexistant de discrimination, le système d'IA apprendra nécessairement à répliquer ces diverses tendances discriminatoires, même si celles-ci ne sont pas le résultat d'une intention délibérée. Ensuite, même lorsque certains types de données sont plus neutres à la base, c'est-à-dire que leur accessibilité n'est pas nécessairement influencée par des biais humains, le choix des données à utiliser peut aussi avoir un impact sur la capacité du système d'IA à représenter de manière équitable la réalité. Au-delà du choix des données, les personnes responsables peuvent aussi inclure leurs préjugés personnels de manière inconsciente dans différentes décisions relatives au design d'un système d'IA, comme le choix des variables dont on peut tenir compte pour faire certains types de prédictions ou le concept même du système qui contribue à perpétuer une certaine forme de discrimination.

2. Quelles sont les conséquences de la discrimination causée par l'intelligence artificielle?

Bien évidemment, les individus comme la société de manière plus générale peuvent être grandement touchés par les différents biais qu'ont appris les systèmes d'IA. En effet, les résultats discriminatoires des systèmes d'IA peuvent mener à des décisions injustes et non équitables, qui peuvent bouleverser la vie des individus affectés. De plus, lorsqu'elles sont nombreuses, les erreurs qui sont causées par les biais des systèmes d'IA contribuent à renforcer des structures discriminatoires déjà présentes dans notre société. D'une part, ces erreurs favorisent l'exclusion de groupes de personnes déjà défavorisées. D'autre part, s'il se nourrit de ses propres prédictions erronées, le système d'IA peut ainsi accentuer les biais qu'il a déjà intégrés à partir des données humaines. Ces erreurs sont d'autant plus graves lorsqu'elles touchent à des enjeux aussi sensibles que la santé, la justice et les finances personnelles.

3. Comment peut-on lutter contre la discrimination en intelligence artificielle?

D'abord, il est possible de poser certaines actions d'un point de vue technique pour rendre la performance des systèmes d'IA moins sujette aux biais. On peut aussi évaluer de manière plus robuste la façon dont ces systèmes tiennent compte des problèmes de discrimination. En outre, il est important de communiquer plus clairement les failles que ces systèmes présentent puisqu'on est encore loin d'avoir créé un système d'IA qui fonctionne de manière parfaitement juste et équitable. Au-delà des solutions techniques, l'industrie et le milieu de la recherche en IA peuvent explorer d'autres avenues, plus sociales ou conceptuelles, pour tenter de contrer l'influence des biais en IA. On peut, entre autres, penser à l'inclusion et la collaboration d'une plus grande diversité de personnes dans la conception des systèmes d'IA. La réglementation a aussi son rôle à jouer pour, notamment, garantir certains standards qui assureraient des processus de développement et de performance éthiques, même s'il demeure assez compliqué de mettre en place ces mesures et de les faire respecter.

Finalement, on a également un rôle à jouer dans la lutte contre la discrimination, même si on se limite à utiliser l'IA et non à la développer. Ces actions n'ont peut-être pas le même genre de visibilité ou de portée que celles posées par les compagnies ou les gouvernements. Elles demeurent tout de même très importantes. On peut, par exemple, chercher à se renseigner davantage, faire un signalement à une autorité compétente lorsqu'on remarque un biais dans un système d'IA, faire preuve de transparence quand on a soi-même recours à un système d'IA et rester critique quand on utilise une technologie pour s'assurer qu'on fait usage de l'IA de manière à rester en accord ses valeurs.

Ressources

Coded Bias : Algorithmes et discrimination [documentaire]. Netflix. 2020. 1h25. https://www.netflix.com/ca-fr/title/81328723

Dioré de Périgny, C. *Vaincre les biais de l'intelligence artificielle* [article en ligne]. Québec Science. 28 mars 2019. https://www.quebecscience.qc.ca/societe/vaincre-biais-intelligence-artificielle/

Équité : types de biais [cours en ligne]. Google. https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias?hl=fr

Gender Shades [documentaire en ligne]. MIT Media Lab. 2018. 5 min. https://www.youtube.com/watch?v=TWWsW1w-BVo (en anglais, sous-titres en français disponibles).

Hsu, J. *How using a medical startup's AI "symptom checker" could go very wrong* [article en ligne]. Quartz. 13 décembre 2019. https://qz.com/1766418/how-using-babylon-healths-ai-symptom-checker-could-go-wrong (en anglais seulement).

L'efficacité d'un logiciel censé prédire la récidive à nouveau critiquée [article en ligne]. Le Monde. 17 janvier 2018. https://www.lemonde.fr/pixels/article/2018/01/17/l-efficacite-d-un-logiciel-cense-predire-la-recidive-a-nouveau-critiquee_5243218_4408996.html

Les outils d'embauche intelligents peuvent développer des biais racistes ou sexistes [article en ligne]. Radio-Canada. 19 janvier 2023. https://ici.radio-canada.ca/nouvelle/1949452/outil-recrutement-intelligence-artificielle-prejuges-embauche

Mao, F. *Robodebt: Illegal Australian welfare hunt drove people to despair* [article en ligne]. BBC News. 7 juillet 2023. https://www.bbc.com/news/world-australia-66130105. (en anglais seulement)

Viniacourt, E. *Aux Etats-Unis, une femme noire enceinte de huit mois victime des biais racistes de la reconnaissance faciale* [article en ligne]. Libération. 7 août 2023. https://www.liberation.fr/international/amerique/aux-etats-unis-une-femme-noire-enceinte-de-huit-mois-victime-des-biais-racistes-de-la-reconnaissance-faciale-20230807_IEHH3U2FFVFMFFVTUSOV53UPJU/

Sabourin, M.-A. *Quand l'intelligence artificielle dérape* [article en ligne]. L'actualité. 5 décembre 2018.
https://lactualite.com/sante-et-science/quand-lintelligence-artificielle-derape/

West, M., Kraut, R. et Chew, H. E. *Je rougirais si je pouvais : réduire la fracture numérique entre les genres par l'éducation* [rapport en ligne]. UNESCO. 2020.
https://unesdoc.unesco.org/ark:/48223/pf0000367416_fre