



# IA ET ÉTHIQUE

## ENJEUX ET LIMITES

**Est-ce que l'on peut "craquer" des modèles d'intelligence artificielle pour récupérer des données sensibles, par exemple des données personnelles utilisées pour les entraîner ?**

C'est parfois possible. Les personnes qui créent ces systèmes ont donc le souci réglementaire ou éthique d'éviter, de vérifier et d'homologuer le système pour s'assurer que ça n'est pas possible. On peut, par exemple, s'assurer que les données ont été anonymisées. Les concepteurs, leurs superviseurs, éventuellement le comité d'éthique, doit vérifier qu'une procédure d'anonymisation, ou de pseudonymisation a été correctement suivie. Si les données ont été pseudonymisées, il y a quelque part une table qui met en correspondance l'identifiant d'une personne dans le jeu d'entraînement et son identité réelle. Il faut vérifier qu'il n'y aura aucune correspondance entre cette table et le système IA. Cependant certaines données sensibles peuvent permettre par elles-mêmes de reconstituer une identité. Par exemple si le jeu de données comprend une image ou un film de moi, ou simplement un enregistrement de ma voix, même si le système d'IA a appris sur un jeu de données ayant "oublié" qu'il s'agit de moi, un utilisateur pourrait me reconnaître. Il ne faut pas que la donnée puisse permettre en tant que telle de retrouver la personne. Il faut donc être prudent, ne pas utiliser des images trop identifiantes, qui comprennent un tatouage par exemple.

**La montée en puissance de l'IA et des moyens de surveillance est-elle une menace pour la vie privée, la démocratie ?**

La réponse à cette question est oui, nous le savons tous. Je pense en particulier à un livre que j'ai lu récemment de David Colon sur la guerre de l'information<sup>1</sup>. À partir par exemple de nos profils Facebook et de dix questions ou post que nous avons affichés, l'IA peut reconstituer assez précisément notre profil, nos idées ou nos habitudes, mais également, ce qui est plus dangereux, notre éventuelle fragilité psychologique, notre fragilité de personnalité. Et il y a, en ce bas monde, des organismes, je dirais même des États, qui se sont spécialisés dans le repérage des personnages fragiles, et la création de communautés artificielles sur les réseaux sociaux, où les personnes discutent entre elles, et reçoivent des incitations à des manifestations ou des actions violentes. Cette manipulation repose sur des techniques psychologiques d'influence, mais avant cela sur l'IA qui permet de repérer les personnes influençables. Je pense que c'est une vraie menace pour la démocratie.

**En quoi les caractéristiques des jeux de données utilisés pour entraîner les modèles peuvent-elles affecter leurs résultats en vie réelle ?**

Les modèles peuvent en effet être biaisés par les données utilisées et entraîner des problèmes éthiques. Par exemple, s'il est observé qu'il y a plus d'accidents

chez les jeunes hommes de 25 à 30 ans que chez les femmes de 50 à 65 ans, une assurance peut en tirer la conclusion qu'elle doit appliquer une surprime aux jeunes hommes de 25 à 30 ans. Cela peut se concevoir d'un point de vue optimisation des coûts et des "rewards". Cependant cela pose un problème de justice, et donc d'éthique, vis-à-vis du très grand nombre de jeunes hommes de 25 à 30 ans qui conduisent parfaitement bien et dont la surprime ne se justifie pas du tout par rapport aux accidents qui pourraient être causés par les distractions auxquelles moi, une femme plus âgée, vais peut-être être sujette.

Un autre exemple serait un système IA de prédiction des risques d'apparition d'une maladie entraîné sur des cohortes de sujets caucasiens mâles issus de milieux confortables. L'exactitude des prédictions de ces systèmes pourrait être dramatiquement différente pour des femmes, des gens en situations de précarité ou d'origine ethnique différente. Mais les choix et les biais des systèmes IA peuvent affecter l'équité de l'accès aux soins de manière plus subtile. Imaginons un système d'IA en médecine personnalisée qui analyse les symptômes des patients et en déduit des recommandations, par exemple, de ne pas prendre trop de beurre parce que le taux de cholestérol avait augmenté (Entre parenthèse, je pense que tout système à visée diagnostique devrait être supervisé par un médecin qui retourne



ITW DE MIREILLE RÉGNIER  
PROPOS RECUEILLIS PAR  
NICOLAS GAMBARDILLA  
8602C



vers le patient une fois le diagnostic posé). Mais on ne s'exprime pas de la même façon dans tous les milieux. Si un système d'IA générative s'exprime avec un certain registre de langage et des expressions langagières dus à des corpus de textes d'entraînement biaisés vers certaines populations, il se peut que les recommandations soient moins compréhensibles pour d'autres couches de la population.

**Les chercheurs en IA sont-ils conscients des problèmes éthiques ? Les prennent-ils en compte ?**

Les chercheurs en IA sont très divers et la réponse à cette question est nécessairement variable. Je pense que la plupart d'entre eux sont conscients de l'impact sur la société et de certains problèmes éthiques. Sont-ils conscients, lorsqu'ils font une recherche, du problème éthique, de l'impact de leur travail, ou même de la manière dont ils font leurs recherches ? C'est moins clair. D'abord, l'être humain étant ce qu'il est, quand on veut faire de la recherche, on va avoir tendance à minimiser les points contradictoires. Il y a donc une tendance évidente et naturelle à diminuer les problèmes éthiques, y compris par l'affirmation « je suis

conscient du problème et je le traite » ; affirmation selon moi contradictoire avec le fait que justement une question éthique repose sur la discussion et doit être résolue à plusieurs. Est-ce les chercheurs en IA prennent ces questions suffisamment en compte ? La réponse est non. Il y a un travail collectif à faire. En revanche, pour être plus positive, je pense qu'il y a une vraie conscience et un désir sincère de discuter, d'approfondir et d'avoir une bonne recherche dans son impact sur la société.

**Quelles sont les structures mises en place pour informer, conseiller ou guider en matière d'éthique de l'IA ?**

Actuellement, toutes les universités mettent en place des comités d'éthique de la recherche (CER) dont un des rôles est d'analyser les projets qui soulèvent des questions éthiques. Est-ce que les chercheurs vont les saisir ? Mon opinion est que si ces structures sont perçues comme une contrainte technique de plus, les chercheurs vont les éviter. Il est très important que ces structures se positionnent comme des structures d'appui aux chercheurs pour un dialogue réfléchi sur le sujet : quelles sont les problématiques susceptibles de

se poser ? Quelles sont les manières simples d'y répondre ? En reconnaissant l'importance de faire avancer la connaissance. Les organismes publics de recherche, comme l'Inria, mettent en place des structures similaires comme le Coerle que je préside. Un certain nombre d'entreprises mettent aussi en place des comités d'éthique ou de réflexions éthiques. C'est une tendance positive, mais il faut être attentif à ne pas faire de l'"ethic washing" ou de la norme déguisée.

Ces comités d'éthique ne sont pas spécifiques de l'IA, mais il y a une augmentation des questions et des soumissions liées à l'éthique de l'IA. À mon avis, ces requêtes ne sont pas à la mesure des problèmes qui sont effectivement soulevés. De plus, il n'y a pas assez de recommandations ou même d'articles qu'un chercheur en IA pourrait lire qui le ferait réfléchir et qu'il pourrait dès lors appliquer spontanément dans sa recherche. De ce point de vue-là, je salue la création du Comité Consultatif National d'Éthique du numérique dont la mission couvre la recherche comme les applications du numérique. Je pense qu'il jouera un vrai rôle. ■

Interview de Mireille RÉGNIER, présidente du Coerle (Comité opérationnel d'évaluation des risques légaux et éthiques de l'Inria (Institut national de recherche en science et technologies du numérique).

1. David COLON. *La Guerre de l'information : les États à la conquête de nos esprits*. Éditions Tallandier 2023. Livre récompensé par le prix de la Revue des Deux Mondes, le prix La Plume et l'Épée, et le prix Corbay