

# Le Mémo

## Les intelligences artificielles sont-elles sexistes ?

**Germain** : Au quotidien, l'outil de traduction de Google est souvent très utile. Mais il semble avoir des troubles de genre. Prenons un exemple : essayez de traduire la phrase allemande « Der Krankenpfleger ist abwesend », c'est-à-dire, « l'infirmier est absent ». Au lieu de l'infirmier, Google va vous suggérer « une infirmière ».

Autre exemple : si vous traduisez « la cheffe » en polonais, vous verrez votre patronne se transformer en patron. Et « quatre historiennes » se transforment, en allemand, en « Vier historiker », des historiens.

Mais pourquoi obtient-on ces résultats ? Est-ce que ce sont juste des erreurs de traduction ? Ou est-ce que les algorithmes sont programmés avec des biais sexistes ?

[JINGLE]

**Germain** : Bonjour Marine

**Marine** : Bonjour Germain

**Germain**: Bienvenue à toutes et tous dans le Mémo, le podcast qui décrypte la société numérique. Les algorithmes nous assistent de plus en plus au quotidien. Mais en même temps, on leur reproche de plus en plus de reproduire les discriminations qui existent déjà dans la société... et notamment celles envers les femmes. En septembre dernier, AlgorithmWatch, une organisation berlinoise, pointait d'ailleurs du doigt cette fâcheuse tendance des outils de traduction à ne pas prendre en compte la féminisation de nombreux métiers.

**Marine** : Et pourtant, Google tente depuis des années de corriger ces erreurs. Je lis dans 20 minutes que Google a promis de nouvelles améliorations en 2020. Ils annoncent « un modèle totalement différent pour éliminer les a priori liés au genre. »

**Germain** : Vaste programme... car ces erreurs ne concernent pas que la traduction. On les retrouve aussi dans la reconnaissance d'images.

**Marine** : Oui, le magazine WIRED raconte les résultats d'une étude menée par des chercheurs européens et américains. Ils ont utilisé des photos d'élus américains pour tester l'outil de reconnaissance d'images de Google. Eh bien, les législateurs hommes sont souvent identifiés avec des mots comme « officiel » ou « leader ». Les photos des femmes, elles, sont qualifiées avec des mots liés à leur apparence physique, tel que : « sourire » ou « beauté ». Ce qui signifie que quand on recherche « leader » sur le moteur de recherche de Google, on trouvera beaucoup plus de photos d'hommes que de femmes.

**Germain** : Peut-être qu'elles souriaient plus sur leur photo que leurs collègues masculins ?

**Marine** : En fait, non. C'est bel et bien un préjugé, qui surgit car l'intelligence artificielle derrière cet outil est elle-même alimentée par des bases de données biaisées.

**Germain** : C'est-à-dire ?

**Marine** : Les ingénieurs ont entraîné cet algorithme en utilisant une quantité massive d'images existant sur le net. Et de fait, sur internet, on trouve beaucoup de photos d'hommes d'affaires, mais peu d'images de femmes en réunion de travail.

**Germain** : Et donc les algorithmes reproduiraient les préjugés de notre société ?

**Marine** : En fait, ils font plus que simplement refléter les préjugés existants. Ils ont tendance à les normaliser et à les amplifier. Pour mieux comprendre, prenons l'exemple

d'Amazon. En 2014, l'entreprise met au point un programme informatique pour filtrer les CVs des candidats. L'agence de presse Reuters raconte à l'époque que pour ce géant aux plus d'un million d'employés, c'est le « graal » du recrutement. Et pourtant, dès les premiers résultats, l'algorithme semble avoir quelques soucis. « Les recruteurs attendaient une sélection froidement objective, mais se sont retrouvés avec un résultat nourri par une culture inégalitaire ». Les CVs contenant le mot « femme » (présidente du club d'échecs féminin par exemple), avaient tout simplement été éliminés par le programme !

**Germain** : Incroyable. Là, c'est de la discrimination pure et dure.

**Marine** : Oui, mais pas délibérée. Le programme se fondait simplement sur les candidatures reçues par Amazon ces dix dernières années.

**Germain** : Et laisse-moi deviner : les candidats étaient majoritairement des hommes.

**Marine** : Exactement ! Et c'est parce qu'on utilise du machine learning, cette capacité de l'intelligence artificielle à apprendre à partir de données. Or, si les données sont elles-même inégalement réparties, on n'est pas prêts de s'en sortir...

**Germain** : Et j'imagine que c'est un phénomène qui ne concerne pas uniquement les femmes ?

**Marine** : Tu as totalement raison! Je lis par exemple dans Quartz que les programmes de reconnaissance faciale ont été principalement alimentés par des photos d'hommes blancs. Donc ils ont du mal, soyons lucides, à reconnaître le reste de la population mondiale qui n'est pas blanche : c'est à dire 79% des habitants de notre planète.

**Germain** : Ce n'est donc pas juste un problème de discrimination, mais aussi une question d'efficacité !

**Marine** : Oui, et la solution se trouverait dans les bases de données. C'est là qu'il faut intervenir, d'après Flora Vincent, co-auteurice du livre *L'Intelligence artificielle, pas sans elles !* Pour elle, l'un des principaux soucis des outils de reconnaissance d'images est que les annotations (les étiquetages) sont souvent crowdsourcées. C'est-à-dire qu'elles sont réalisées par des internautes. Et certaines de ces annotations peuvent revêtir des stéréotypes... Comme cette image d'une femme blonde à côté d'un homme chauve. L'annotation associée disait "une secrétaire se fait sermonner par son patron." Flora Vincent explique au magazine *Usbek & Rica* : « qu'il y a des gens qui travaillent aujourd'hui sur des bases de données standardisées représentatives à la fois de différents sexes, nationalités et morphologies, afin de s'assurer qu'on ne laisse personne de côté ».

**Germain** : Comme beaucoup de développeurs utilisent les mêmes bases de données, ces changements peuvent avoir un impact exponentiel.

**Marine** : Oui, et c'est pour ça que la chercheuse américaine Joy Buolowini, citée dans l'entretien, appelle à faire de l'in-coding. « C'est la version informatique de l'écriture inclusive » explique Flora Vincent.

**Germain** : Mais ces bases de données standardisées et représentatives, d'où viendraient-elles ? Existe-t-il une culture sans aucune inégalité ?

**Marine** : Pas vraiment. Ces données sont fabriquées pour être inclusives. À ce propos, on peut lire dans le magazine d'affaire *Forbes* que « La recette pour des algorithmes éthiques est simple : il suffit de les alimenter avec des données équitables ». C'est ce qu'on appelle des données de synthèse, ou « synthetic data » en anglais.

**Germain** : Donc on utiliserait des données fictives pour se construire un avenir bien réel mais moins biaisé... On plante des graines d'égalité, en espérant qu'elles germent. Et où en est-on dans la mise en place de ces mesures ?

**Marine** : D'après le site d'actualité web Les Numériques, changer les algorithmes est une chose, mais « la mise à mal des préjugés requiert un travail de fond bien plus colossal. Plusieurs collectifs tirent d'ailleurs la sonnette d'alarme sur le manque de femmes dans l'informatique. Je lis qu'aujourd'hui, [en France], il y a 33% de femmes dans les métiers du numérique... mais si l'on resserre autour des métiers purement techniques, les développeurs, les ingénieurs, les data scientists... ce pourcentage tombe à 12% ».

**Germain** : D'ailleurs les collectifs ne sont pas les seuls à se saisir du sujet en France !

**Marine** : En effet, suite à un séminaire sur les algorithmes l'été dernier, la CNIL et le Défenseur des droits ont appelé à une prise de conscience collective, et engagé « les pouvoirs publics... à prendre des mesures pour éviter que les discriminations soient reproduites et amplifiées par ces technologies ».

**Germain** : Le mieux serait que des femmes soient aux commandes de ce genre d'initiatives et impliquées dans la tech. Mais j'imagine que le chemin à parcourir reste long...

**Marine** : Oui, comme en témoigne le licenciement de cette chargée de l'éthique chez Google, dont l'histoire est relayée par le Financial Times. Margaret Mitchell a été écartée en février 2021 lorsqu'elle a commencé à s'intéresser de trop près au sort d'une collègue. Elle avait été licenciée pour avoir publié un rapport sur les limites de l'IA et notamment sur les préjugés de genre. « La pression pour réformer ces géants de la tech ne vient pas des actionnaires ou des autorités extérieures, mais plutôt des employés eux-mêmes ».

**Germain** : Mesures politiques externes, pressions internes – quels sont les meilleurs outils pour faire évoluer les choses ? On se retrouve justement la semaine prochaine pour se pencher plus en détail sur les défis de la mixité dans le numérique. Merci Marine, et merci à vous de nous avoir écoutés. J'espère que cet épisode vous a plu.

Vous pouvez retrouver les articles ayant servi à l'écriture dans la description. À la semaine prochaine !

## SOURCES

- [Pour les algorithmes de traduction, les infirmiers sont des infirmières](#) (ICT Journal)
- [Google Traduction : Une nouvelle approche pour limiter les a priori liés au genre](#) (20 minutes)
- [When AI Sees a Man, It Thinks 'Official.' A Woman? 'Smile'](#) (Wired)
- [Amazon scraps secret AI recruiting tool that showed bias against women](#) (Reuters)
- [AI will reproduce and enshrine age-old biases—if we let it](#) (Quartz)
- [« On peut utiliser l'IA pour rendre la société plus égalitaire »](#) (Usbek & Rica)
- [Synthetic Data Promises Fair AI And Privacy Compliance, But How Exactly Does It Work?](#) (Forbes)
- [L'intelligence artificielle est sexiste, mais elle se soigne](#) (Les Numériques)
- [Algorithmes et discriminations : le Défenseur des droits, avec la CNIL, appelle à une mobilisation collective](#) (CNIL)
- [After Google drama, Big Tech must fight against AI bias](#) (Financial Times)