



HAL
open science

L'intelligence artificielle serait-elle “ woke ” ?

Adrien Tallent

► **To cite this version:**

| Adrien Tallent. L'intelligence artificielle serait-elle “ woke ” ?. Revue Esprit, 2024. hal-04727724

HAL Id: hal-04727724

<https://hal.science/hal-04727724v1>

Submitted on 9 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

L'intelligence artificielle serait-elle « woke » ?

Paru dans la *Revue Esprit*, le 06 septembre 2024, en ligne < <https://esprit.presse.fr/actualites/adrien-tallent/l-intelligence-artificielle-serait-elle-woke-45455> >

*Adrien Tallent*¹

¹ Sciences, Normes, Démocratie (SND), Sorbonne Université, CNRS, Paris, France

En 2023, Google annonce triomphalement la sortie de son nouveau produit fondé sur de l'intelligence artificielle. Nommé Gemini, il doit offrir une solution à un problème récurrent des systèmes d'intelligence artificielle (SIA) : les biais. À peine Gemini sorti auprès du grand public, des premiers *tweets* et articles sont publiés montrant des exemples d'images proposées par Gemini lorsqu'on lui formule des requêtes (ou *prompt*) : des soldats de l'armée allemande de 1943 ayant la peau noire ou des traits asiatiques. Les commentaires de l'aile conservatrice de la politique américaine fustigent alors immédiatement le « wokisme » de l'IA et plus généralement de la Silicon Valley. Le débat sur le côté *woke* de l'IA s'inscrit alors dans un contexte plus large de polarisation politique aux États-Unis, particulièrement visible dans la campagne électorale de 2024. Ce débat touche à des questions de biais, de contrôle de l'information et de manipulation politique.

La question des biais

Alors que les algorithmes deviennent omniprésents dans nos vies quotidiennes, les individus se trouvent confrontés à des systèmes automatisés perçus comme essentiellement infaillibles, rationnels et objectifs, mais dont la conception même est le reflet des choix et des biais des concepteurs et des sociétés humaines. La question des biais discriminatoires est au cœur de nombreux travaux de recherche sur l'IA. Les SIA sont fondés sur des données. Celles-ci étant issues du monde réel, elles représentent alors les biais discriminatoires présents dans nos sociétés. ChatGPT n'a pas de conscience, et demeure programmé par des humains ; l'IA est certes biaisée, mais c'est parce que nous la nourrissons de nos propres préjugés ; à nous d'orchestrer notre utilisation de ces technologies, au lieu de fantasmer leur toute-puissance. Mais les SIA ne font pas que reproduire ces biais, ils les amplifient en les rendant systématiques et automatiques. Ils ne sont que « *des choix éminemment subjectifs, des opinions, voire des préjugés insérés dans des équations mathématiques* »¹.

Les SIA sont confrontés à deux types de problèmes. Le premier est historique : leurs données d'entraînement reflètent la domination historique des hommes blancs âgés. Le second est structurel : les SIA sont le miroir de nos inégalités actuelles. De nombreux exemples ont démontré et démontrent encore les risques de biais présents dans les systèmes automatisés au sens large. Ainsi, en 2016, de nombreux médias avaient couvert le cas d'un algorithme d'aide à la décision utilisés dans une cour de justice américaine. Destiné à prédire le risque de récidive des prévenus, le programme s'est révélé plus sévère avec les personnes de couleur noire. De la même manière en 2018, un logiciel de recrutement utilisé par Amazon discriminait les CV de femmes. Plus récemment, une campagne de sensibilisation par l'association #JamaisSansElles mettait en avant l'aspect discriminatoire des modèles d'IA génératives qui représentent un

¹ Cathy O'Neil, *Algorithmes : la bombe à retardement*, Les Arènes, 2018.

« CEO » (PDG en anglais) sous les traits d'hommes blancs de plus de cinquante ans, et « a secretary » (l'anglais a l'avantage du neutre) sous ceux de jeunes femmes souvent sexualisées. Ce faisant, les modèles d'IA perpétuent, renforcent et automatisent les discriminations et inégalités. C'est dans ce contexte que Google a lancé Gemini. L'objectif était alors de débiaiser les modèles d'IA en favorisant la représentation des groupes souvent sous-représentés et victimes de discriminations. Il semble qu'en cherchant à corriger les biais, les équipes de Google les aient remplacés par l'absurdité historique. Est-ce néanmoins plus grave que le fait de ne représenter les femmes et les minorités que dans des rôles subalternes, genrés et biaisés ? Les biais des SIA ainsi que l'expérience Gemini sont deux faces du même problème.

Les modèles d'IA doivent-ils représenter notre monde actuel et son histoire, ou bien refléter nos aspirations pour la société ? Il faudrait de la nuance. Néanmoins, il est compliqué pour un modèle d'IA de savoir quand les données historiques sont importantes et quand la représentation d'une vision idéale de la société est préférable.

L'IA dans la guerre culturelle

Dans un article publié sur le blog de l'entreprise, Google a reconnu que Gemini était « à côté de la plaque »² tout en précisant qu'il s'agissait d'une solution de création qui n'a pas de valeur de vérité.

Malgré le *mea culpa*, il n'en fallait pas davantage pour que les conservateurs américains se saisissent du sujet et y voit un excès de « wokisme ». Les systèmes d'IA seraient trop influencés par des idéologies progressistes, au détriment de la supposé neutralité. Il faut rappeler que cette neutralité n'existe pas. En France, certains éditorialistes, philosophes, et responsables politiques de droite et d'extrême droite se sont aussi saisis du sujet pour dénoncer le « wokisme » venu des États-Unis, important par la même occasion la guerre culturelle de la droite conservatrice américaine contre le progressisme de la Silicon Valley.

Si la Silicon Valley est historiquement proche du parti démocrate et l'est toujours en grande majorité, de plus en plus de figures font désormais exception. A tel point que certains médias évoquent un tournant politique dans la Valley³. Elon Musk – chantre de la non-intervention sur les réseaux sociaux et soutien de Trump – a ainsi construit son propre modèle d'IA, nommé Gork, pour qu'il soit « antiwoke ». Il a également particulièrement critiqué le « wokisme » de l'IA de Google en affirmant sur X (dont il est le propriétaire) que des IA comme Gemini pourraient « éventuellement tuer des gens ». Autre figure de proue de la croisade « antiwoke » d'une partie du monde la Tech, Peter Thiel – milliardaire admirateur de René Girard et un des premiers investisseurs de Facebook – a notamment financé la campagne victorieuse de J.D Vance dans l'Ohio en 2022⁴. Ce dernier est depuis devenu le candidat de Donald Trump pour la vice-présidence.

Comme pour les réseaux sociaux, le difficile contrôle des contenus de l'IA s'intègre dans une polarisation politique. On retrouve alors les oppositions entre la modération et les promoteurs du laissez-faire et de la liberté d'expression à tout prix – dont on sait déjà qu'elle est fondamentalement inéquitable. Donald Trump a annoncé qu'il abrogera l'*Executive Order* de

² Voir l'article publié par Google : <https://blog.google/products/gemini/gemini-image-generation-issue/>

³ Nicole Narea, "Why tech titans are turning toward Trump", *Vox*, 17 juillet 2024.
<https://www.vox.com/politics/361087/trump-silicon-valley-fundraising-musk-andreessen-horowitz>

⁴ Voir sur le sujet : Adrien Tallent, « Facebook, antichambre du trumpisme », *Esprit*, Mai 2022
<https://esprit.presse.fr/article/adrien-tallent/facebook-antichambre-du-trumpisme-43994>

Joe Biden sur l'IA en l'accusant d'aller à l'encontre de l'esprit d'innovation et « *censurer la liberté d'expression des américains* »⁵. Dans la perspective de l'élection présidentielle, si les conservateurs américains entendent attaquer le progressisme du monde de la Tech, Trump tente de séduire des barons de la Tech de plus en plus sensibles à ce discours anti-régulation.

Quelle valeur de vérité pour l'IA ?

Pour autant, les critiques interprètent en partie faussement les résultats donnés par une intelligence artificielle générative. Celle-ci a pour but de générer du contenu créatif. Le problème vient donc en partie du fait que les utilisateurs leur confèrent une valeur de vérité et projettent leurs propres systèmes idéologiques sur des systèmes sociotechniques déjà le fruit de volontés et d'idéologies humaines. Cette illusion rend le risque de manipulation politique d'autant plus important – l'explosion du nombre de *deepfake* politiques en témoigne.

L'aspiration de Google avec Gemini était d'aller à l'encontre des stéréotypes et d'essayer de représenter une certaine idée de la diversité, partant du principe que tout le monde voudrait cela. Il s'est révélé que ce n'était pas le cas. L'IA est un sujet politique. Face à ces défis, des experts et des militants appellent à une plus grande diversité parmi les ingénieurs et à une transparence accrue. Il est crucial de comprendre et de contrôler les biais inhérents aux systèmes d'IA pour éviter des manipulations à grande échelle. Les croisades « anti-woke » d'une partie désormais significative de la Silicon Valley pose alors en retour la question de la direction prise par les innovations technologiques et des conséquences d'une victoire de Trump sur les enjeux éthiques.

⁵ [Washington Examiner](#).