
L'IA est un WASP

Proposons ici une réflexion critique sur les **biais socioculturels et raciaux** à l'œuvre dans les systèmes d'intelligence artificielle générative, en particulier dans la génération d'images. À partir d'une expérience méthodique d'utilisation d'outils d'IA, on montre que ces technologies, souvent présentées comme neutres et objectives, reproduisent en réalité des **représentations dominantes** issues de contextes sociaux, culturels et géographiques spécifiques.

La problématique centrale peut être formulée ainsi :

dans quelle mesure les intelligences artificielles contemporaines, loin d'être neutres, reconduisent-elles des normes et des stéréotypes issus de la société dominante, et quelles implications éthiques et politiques cela soulève-t-il en matière de représentation, de diversité et d'inclusion ?

On part d'une pratique régulière de l'intelligence artificielle, notamment dans la production de textes et d'images, pour interroger la prétendue neutralité de ces technologies. En mobilisant des apports issus de la linguistique, de la sémiologie et de l'informatique, on rappelle que les images générées par l'IA doivent être considérées comme de **véritables discours**, porteurs de sens, de normes et de représentations implicites.

L'analyse repose sur une série d'expérimentations à partir de prompts volontairement simples ou supposés neutres. Le constat est récurrent : des termes génériques tels que *people* ou *guys* conduisent presque systématiquement à la production d'images représentant des **hommes blancs**, jeunes, occidentaux, souvent placés dans des postures valorisées. On en déduit que l'IA fonctionne selon un modèle implicite qui assimile la neutralité à une figure sociale dominante, que l'on peut caractériser comme celle d'un homme blanc américain trentenaire.

Lorsque l'on tente de corriger ces biais en précisant le genre ou l'origine ethnique, on observe que les stéréotypes ne disparaissent pas mais se **reconfigurent**. La mention explicite du genre féminin aboutit majoritairement à des représentations de femmes blanches, fréquemment accompagnées d'hommes, ce qui suggère une persistance de la centralité masculine. De même, lorsque l'on précise des origines non blanches, les figures produites apparaissent souvent marginalisées, en retrait, parfois de dos ou en nombre réduit, traduisant une **invisibilisation symbolique** des minorités.

On pourrait mettre en évidence un paradoxe majeur lorsque l'on confronte ces résultats aux **réalités démographiques mondiales**. Les populations asiatiques et africaines constituent la majorité de l'humanité, mais demeurent largement sous-représentées dans les images générées par l'IA. À l'inverse, les groupes occidentaux, pourtant minoritaires à l'échelle globale, occupent une place centrale dans l'imaginaire algorithmique.

Cette situation a également des effets subjectifs et sociaux. On peut souligner que les personnes appartenant à des groupes minorisés peuvent éprouver un sentiment de **non-représentation**, voire d'effacement identitaire, face à des images censées être universelles mais qui ne leur ressemblent pas. L'IA, au lieu de corriger les inégalités symboliques existantes, risque ainsi de les **renforcer et de les normaliser**.

Les biais observés ne sont pas accidentels : ils résultent des données d'entraînement, des

choix de conception et des cadres culturels dominants dans lesquels ces technologies sont développées. Corriger ces dérives suppose une diversification des équipes, une révision critique des corpus utilisés et l'intégration d'une perspective réellement plurielle et mondiale. À défaut, l'intelligence artificielle pourrait devenir un puissant vecteur de reproduction des inégalités sociales, culturelles et raciales, plutôt qu'un outil d'émancipation et de diversité.