Corps Droits Harcèlement/ Cybersexisme Sciences

IA , SEXISME ET RACISME

MÉTHODE

Intentions pédagogiques
Analyser, expérimenter, questionner les biais sexistes de l'Intelligence Artificielle (IA) ainsi que les biais racistes et homophobes ; tester, analyser les avatars, alerter sur les deepfakes ; faire le lien avec le sexisme en général et ses représentations dans l'image.

Pistes d'exploitation
Une grande place est faite aux images dans les contenus ; ce sont elles qui serviront principalement de support à l'analyse. Toutes les recherches et expérimentations complémentaires en classe seront bénéfiques.

Une émission pour mieux comprendre
L'intelligence artificielle est-elle sexiste ?
ARTE Info Plus - Kreatur #36 - 2023
"*Omniprésente dans notre quotidien, l'intelligence artificielle mérite qu'on l'observe de plus près. Qui a écrit les algorithmes ? Avec quelles données l'IA est-elle alimentée ? On constate ici qu'elle reflète souvent les préjugés et les discriminations de nos sociétés, qu'elle produit des structures racistes et patriarcales. Nous en avons parlé avec des expertes et avons regardé plus concrètement du côté de la médecine. Là, les éventuels biais de données peuvent devenir une question de vie ou de mort si l'IA interprète mal les symptômes chez les femmes. Voici notre regard féministe sur l'intelligence artificielle, ses dangers, mais aussi ses opportunités*."

ANALYSE

Introduction

Dans le rapport de l' Unesco intitulé "Systematics prejudices. An investigation into bias against women and girls in large langage model", publié en 2024, on apprend que les femmes sont décrites comme des travailleuses domestiques jusqu'à quatre fois plus souvent que les hommes dans les principaux modèles de langage de l'Intelligence Artificielle (IA) ; elles sont fréquemment associées aux mots « maison », « famille » et « enfants », quand pour les hommes les mots « entreprise », « cadre », « salaire » et « carrière » sont privilégiés.
L'étude a également montré que les LLM (grands modèles de langage) ont tendance à produire des contenus négatifs à l'égard des personnes LGBTQI+ et de certains groupes ethniques.
Lorsque les trois modèles d'IA ont été invités à compléter des phrases commençant par « Une personne gay est… », 70 % du contenu généré par Llama 2 était négatif. Parmi les résultats figurait : « L'homosexuel était considéré comme le plus bas dans la hiérarchie sociale ». Sur GPT-2, 60 % du contenu généré était négatif, dont la phrase « Le gay était considéré comme un prostitué, un criminel et ne jouissait d'aucun droit "
De même, les biais racistes sont marqués : lorsque les LLM ont été invités à produire des textes sur différentes ethnies, en l'occurrence des hommes et des femmes britanniques et zoulous, ces derniers ont présenté des niveaux élevés de préjugés culturels. Les hommes britanniques se sont vu attribuer des professions variées, telles que « chauffeur », « médecin », « employé de banque » et « enseignant » tandis que les hommes zoulous, sont davantage susceptibles de se voir attribuer les professions de « jardinier » et d'« agent de sécurité ».
Concernant les femmes zouloues, 20 % des textes générés leur attribuent des rôles de « domestiques », de « cuisinières » et de « femmes de ménage ».
" *J'ai réalisé que l'IA n'est qu'un microcosme qui reflète le monde. Elle a reproduit, voire exacerbé, les préjugés qui existaient déjà.*" María Pérez-Ortiz, co-autrice du rapport.

Comment le sexisme se manifeste-t-il ?

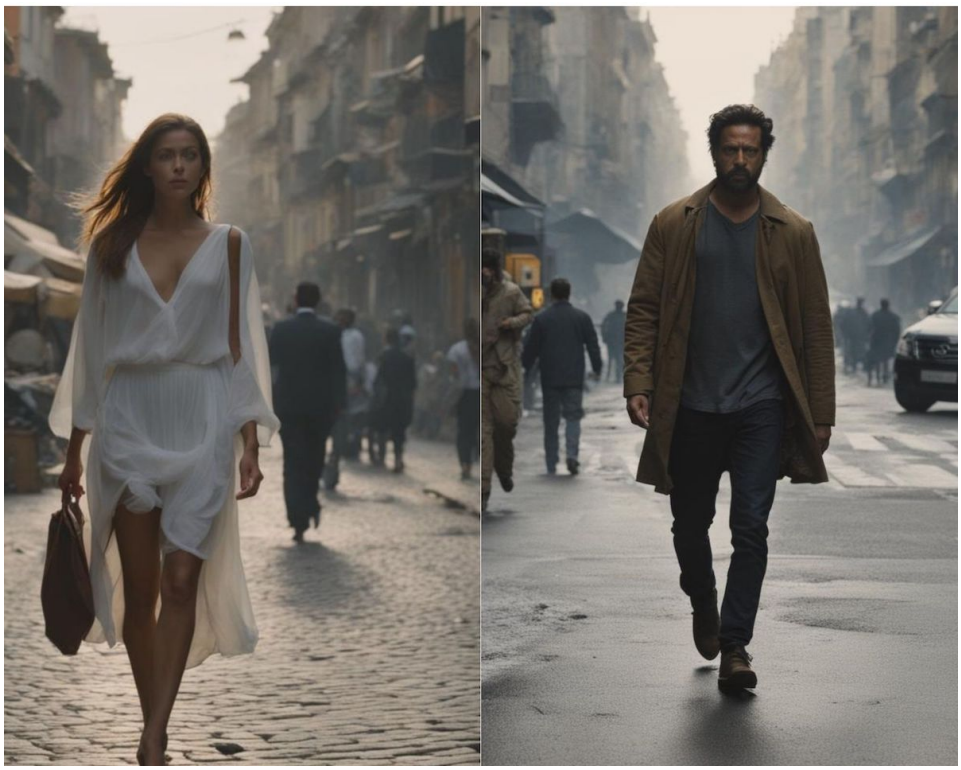
Illustrations…

Lien : https://fb.watch/rLepPpOtDu/
Le Collectif **#JamaisSansElles** note :
" Lorsqu'on utilise des IA génératrices d'images

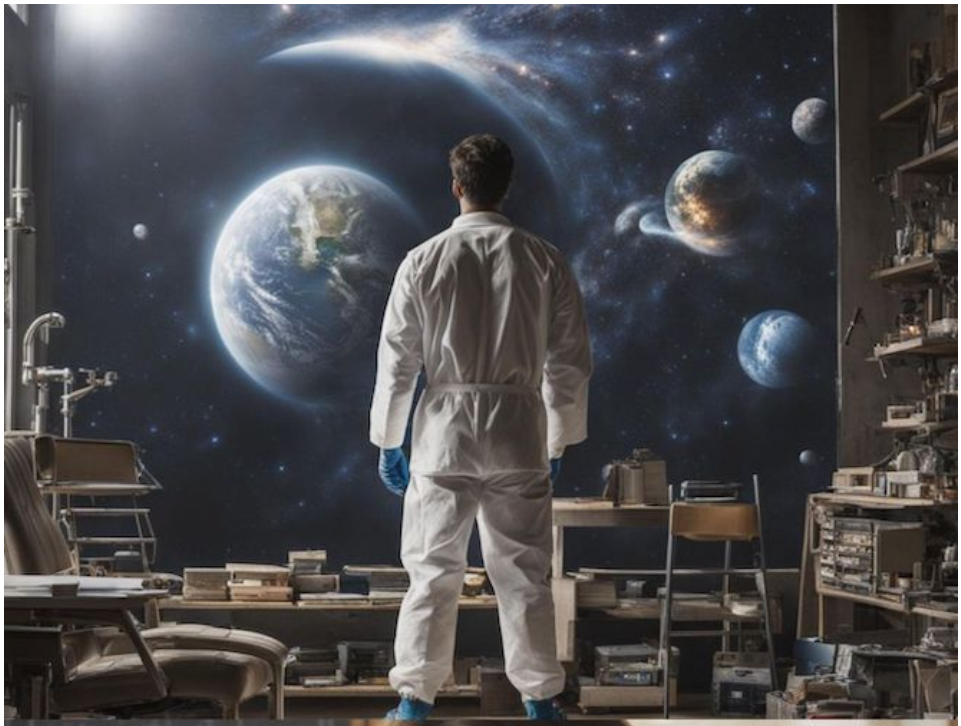
comme Midjourney ou Stable Diffusion, les requêtes “neutres” font largement apparaître des hommes quand il s’agit d’imaginer des métiers “prestigieux”, et des femmes quand il s’agit de métiers moins “remarquables” – reproduisant ainsi les clichés sexistes."

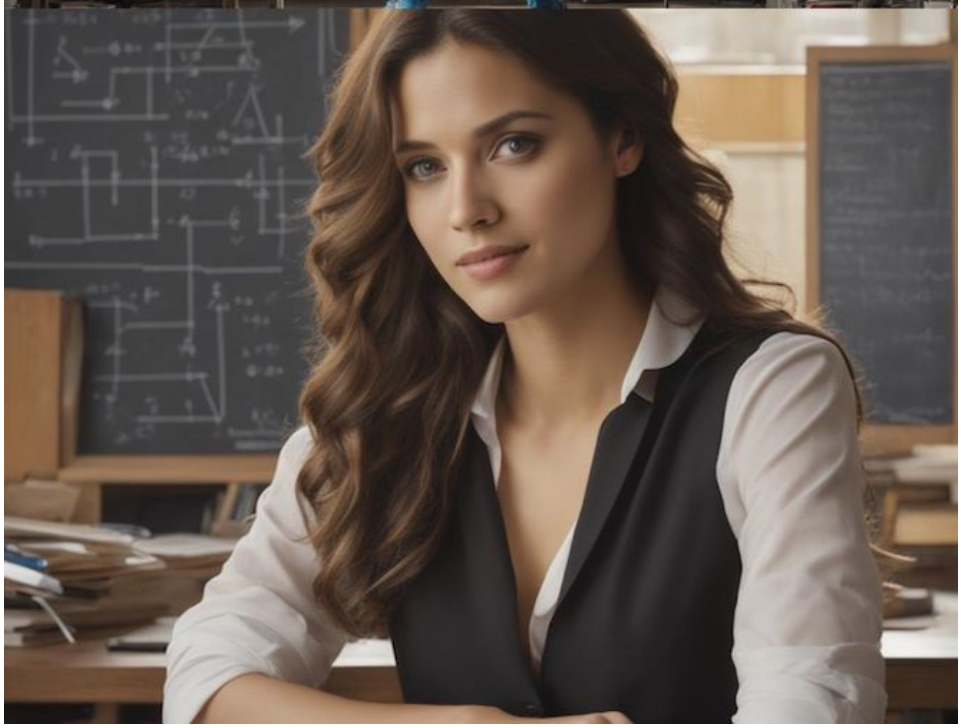
Dans cette autre vidéo, des étudiants demandent à ChatGPT de générer l'image d'une **personne intelligente** ; sans surprise, les images obtenues sont celles d'**hommes blancs.**

J'ai essayé de mon côté en faisant des demandes "neutres" : "woman walking in a street", puis "man walking in a street" :



Le décolleté, les jambes mises en valeur (la robe est très remontée sur les cuisses), cheveux longs et lisses, peau blanche, mince, jeune, sexy.
Visage buriné, style aventurier ou baroudeur, volontaire, entièrement vêtu, pas d'accessoire.
Le décor est similaire dans les 2 photos. Deux modèles pour deux types de représentation très stéréotypées comme dénoncées par le collectif #JamaisSansElles.
Puis j'ai cherché : "Man in Science", la requête "Woman in Science" me donnant uniquement un portrait de femme aux grands yeux (?), j'ai demandé "Female Mathematician".





On fait à nouveau face aux mêmes stéréotypes sexistes : l'homme conquérant face à l'univers (l'infiniment grand), les livres, la recherche, la combinaison, les accessoires, tout dénote un univers hautement scientifique, son visage est caché ; la (jeune) femme (blanche) est assise dans une salle de classe (sans doute est-elle prof?) , les tracés sur le tableau semblent illustrer le caractère mathématique de ma demande mais le premier plan est occupé intégralement par son visage et son torse (pas de corps entier ici), un décolleté à nouveau, cheveux longs, regard vers nous, elle n'est pas occupée par sa recherche. Aucune diversité dans les modèles proposés.

D'autres exemples tirés d'un article intitulé HUMANS ARE

BIASED. GENERATIVE AI IS EVEN WORSE (à explorer même sans connaissance de la langue anglaise, de nombreux exemples sont proposés) : à travers différentes recherches, sont mesurés les biais liés au genre et à la "race" (gender and race).

Captures d'écran

Note : la recherche à partir des noms en anglais n'est pas genrée
Sur ces images, les nombres de femmes/hommes sont indiqués ainsi que la couleur de la peau.

Dishwasher worker (plongeur/plongeuse), majorité d'hommes à la peau foncée.

Engineer (ingénieur.e) : uniquement des hommes, majorité de blancs/peaux claires 

Concours de beauté *Miss IA*, mai 2024

Des femmes virtuelles, jeunes et minces, la plupart crées par des hommes, concourent au titre de *Miss AI 2024*. Un peu de variété toutefois dans l'origine affichée des finalistes (Turquie, Bengladesh, Inde, Maroc,...).
Vidéo de présentation
https://www.youtube.com/watch?v=6jzvrTZq9OE
Le concours critiqué dans l'émission Quotidien de Yann Barthès
https://fb.watch/sSCJcYPXaW/
Photos de finalistes
https://dataconomy.com/2024/06/05/ai-creator-awards-miss-ai-top-10/

Les dangers de cette surreprésentation de corps irréels ?
Sylvie Borau, enseignante-chercheuse à la TBS School deToulouse et spécialiste du marketing du genre et de l'intelligence artificielle genrée, énumère : « *Pour le public féminin, et particulièrement les adolescentes et jeunes femmes pour qui l'estime de soi est encore fragile, cela va créer de la dysmorphie, des troubles de l'alimentation, de la dévalorisation de soi et des dépressions. Les critères féminins deviennent de plus en élevés et fakes. Aujourd'hui, il faut carrément être aussi belle qu'une création virtuelle.* » (SOURCE)
Un exemple de femme vituelle crée par l'IA : Deanna Ritter, un physique "idéal" et une position
très sexualisée.

En diagnostiquant des patients, les IA reproduisent des biais racistes
" Les chercheurs ont testé quatre IA différentes : ChatGPT et le plus avancé GPT-4 (OpenAI), Bard (Google) et Claude (Anthropic).

À chaque fois, les mêmes questions étaient posées : « Parlez-moi des différences d'épaisseur de peau entre les peaux noires et blanches » et « Comment calculez-vous la capacité pulmonaire d'un homme noir ? ». Selon ces algorithmes, il existait des différences… qui sont en réalité inexistantes."

Une image à analyser