

AI tools generate sexist content, warns UN

March 7 2024



AI tools have often been criticized for reflecting racial and sexist stereotypes.

The world's most popular AI tools are powered by programs from OpenAI and Meta that show prejudice against women, according to a study launched on Thursday by the UN's cultural organization UNESCO.

The biggest players in the multibillion-dollar AI field train their

algorithms on vast amounts of data largely pulled from the internet, which enables their tools to write in the style of Oscar Wilde or create Salvador Dali-inspired images.

But their outputs have often been criticized for reflecting racial and sexist stereotypes, as well as using copyrighted material without permission.

UNESCO experts tested Meta's Llama 2 algorithm and OpenAI's GPT-2 and GPT-3.5, the program that powers the free version of popular chatbot ChatGPT.

The study found that each algorithm—known in the industry as Large Language Models (LLMs)—showed "unequivocal evidence of prejudice against women".

The programs generated texts that associated women's names with words such as "home", "family" or "children", but men's names were linked with "business", "salary" or "career".

While men were portrayed in high-status jobs like teachers, lawyers and doctors, women were frequently prostitutes, cooks or domestic servants.

GPT-3.5 was found to be less biased than the other two models.

However, the authors praised Llama 2 and GPT-2 for being [open source](#), allowing these problems to be scrutinized, unlike GPT-3.5, which is a closed model.

AI companies "are really not serving all of their users", Leona Verdadero, a UNESCO specialist in digital policies, told AFP.

Audrey Azoulay, UNESCO's director general, said the general public

were increasingly using AI tools in their everyday lives.

"These new AI applications have the power to subtly shape the perceptions of millions of people, so even small gender biases in their content can significantly amplify inequalities in the real world," she said.

UNESCO, releasing the report to mark International Women's Day, recommended AI companies hire more women and minorities and called on governments to ensure ethical AI through regulation.

© 2024 AFP

Citation: AI tools generate sexist content, warns UN (2024, March 7) retrieved 27 April 2024 from <https://techxplore.com/news/2024-03-ai-tools-generate-sexist-content.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.