

Gemini's flawed AI racial images seen as warning of tech titans' power

March 17 2024, by Julie JAMMOT



A Google Gemini AI gaffe when it came to creating images on command spotlighted the challenge of eliminating cultural bias in such tech tools without ridiculous results.

For people at the trend-setting tech festival here, the scandal that erupted after Google's Gemini chatbot cranked out images of Black and Asian

Nazi soldiers was seen as a warning about the power artificial intelligence can give tech titans.

Google CEO Sundar Pichai last month slammed as "completely unacceptable" errors by his company's Gemini AI app, after gaffes such as the images of ethnically diverse Nazi troops forced it to temporarily stop users from creating pictures of people.

Social media users mocked and criticized Google for the historically inaccurate images, like those showing a female black US senator from the 1800s—when the first such senator was not elected until 1992.

"We definitely messed up on the image generation," Google co-founder Sergey Brin said at a recent AI "hackathon," adding that the company should have tested Gemini more thoroughly.

Folks interviewed at the popular South by Southwest arts and tech festival in Austin said the Gemini stumble highlights the inordinate power a handful of companies have over the [artificial intelligence](#) platforms that are poised to change the way people live and work.

"Essentially, it was too 'woke,'" said Joshua Weaver, a lawyer and tech entrepreneur, meaning Google had gone overboard in its effort to project inclusion and diversity.

Google quickly corrected its errors, but the underlying problem remains, said Charlie Burgoyne, chief executive of the Valkyrie applied science lab in Texas.

He equated Google's fix of Gemini to putting a Band-Aid on a bullet wound.

While Google long had the luxury of having time to refine its products, it

is now scrambling in an AI race with Microsoft, OpenAI, Anthropic and others, Weaver noted, adding, "They are moving faster than they know how to move."

Mistakes made in an effort at cultural sensitivity are flashpoints, particularly given the tense political divisions in the United States, a situation exacerbated by Elon Musk's X platform, the former Twitter.

"People on Twitter are very gleeful to celebrate any embarrassing thing that happens in tech," Weaver said, adding that reaction to the Nazi gaffe was "overblown."

The mishap did, however, call into question the degree of control those using AI tools have over information, he maintained.

In the coming decade, the amount of information—or misinformation—created by AI could dwarf that generated by people, meaning those controlling AI safeguards will have huge influence on the world, Weaver said.

Bias-in, Bias-out

Karen Palmer, an award-winning mixed-reality creator with Interactive Films Ltd., said she could imagine a future in which someone gets into a robo-taxi and, "if the AI scans you and thinks that there are any outstanding violations against you... you'll be taken into the local police station," not your intended destination.

AI is trained on mountains of data and can be put to work on a growing range of tasks, from image or audio generation to determining who gets a loan or whether a medical scan detects cancer.

But that data comes from a world rife with cultural bias, disinformation

and social inequity—not to mention online content that can include casual chats between friends or intentionally exaggerated and provocative posts—and AI models can echo those flaws.

With Gemini, Google engineers tried to rebalance the algorithms to provide results better reflecting human diversity.

The effort backfired.

"It can really be tricky, nuanced and subtle to figure out where bias is and how it's included," said technology lawyer Alex Shahrestani, a managing partner at Promise Legal law firm for tech companies.

Even well-intentioned engineers involved with training AI can't help but bring their own life experience and subconscious bias to the process, he and others believe.

Valkyrie's Burgoyne also castigated big tech for keeping the inner workings of generative AI hidden in "black boxes," so users are unable to detect any hidden biases.

"The capabilities of the outputs have far exceeded our understanding of the methodology," he said.

Experts and activists are calling for more diversity in teams creating AI and related tools, and [greater transparency](#) as to how they work—particularly when algorithms rewrite users' requests to "improve" results.

A challenge is how to appropriately build in perspectives of the world's many and [diverse communities](#), Jason Lewis of the Indigenous Futures Resource Center and related groups said here.

At Indigenous AI, Jason works with farflung [indigenous communities](#) to design algorithms that use their data ethically while reflecting their perspectives on the world, something he does not always see in the "arrogance" of big tech leaders.

His own work, he told a group, stands in "such a contrast from Silicon Valley rhetoric, where there's a top-down 'Oh, we're doing this because we're going to benefit all humanity' bullshit, right?"

His audience laughed.

© 2024 AFP

Citation: Gemini's flawed AI racial images seen as warning of tech titans' power (2024, March 17) retrieved 28 April 2024 from <https://techxplore.com/news/2024-03-gemini-flawed-ai-racial-images.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.