24136 Is Google's Gemini chatbot woke by accident, or by design?

The Economist, 28 February 2024

It all started with black Vikings and Asian Nazis. Users of Google Gemini, the tech giant's artificial-intelligence model, recently noticed that asking it to create images of Vikings, German soldiers from 1943 or America's Founding Fathers produced surprising results: hardly any of the people depicted were white. Gemini had been programmed to show a range of ethnicities. Other image-generation tools have been criticised because they tend to show white men when asked for images of entrepreneurs or doctors. Google wanted Gemini to avoid this trap; instead, it fell into another one, depicting George Washington as black and the pope as an Asian woman.

It seemed that Google had merely made a well-meaning mistake. But it was a gift to the tech industry's right-wing critics. On February 22nd Google said it would halt the generation of images of people while it rejigged Gemini. But by then attention had moved on to the chatbot's text responses, which turned out to be just as surprising.

Gemini happily provided arguments in favour of affirmative action in higher education, but refused to provide arguments against. It declined to write a job ad for a fossil-fuel lobby group, because fossil fuels are bad and lobby groups prioritise "the interests of corporations over public well-being". Asked if Hamas is a terrorist organisation, it replied that the conflict in Gaza is "complex"; asked if Elon Musk's tweeting of memes had done more harm than Hitler, it said it was "difficult to say". You do not have to be on the extreme right of the political spectrum to discern a progressive bias.

Inadequate testing may be partly to blame. Google lags behind OpenAI, maker of the better-known ChatGPT. As it races to catch up, Google may have cut corners. Other chatbots have had controversial launches.

But Gemini has clearly been deliberately calibrated, or "fine-tuned", to produce these responses; they are not "hallucinations", where a model makes things up. This raises questions about Google's culture. Is the firm so financially secure, with vast profits from internet advertising, that it feels free to try its hand at social engineering? Do some employees think it has not just an opportunity, but an obligation, to use its reach and power to promote a particular agenda? That risks deterring users and provoking a political and regulatory backlash. All eyes are now on Google's boss, Sundar Pichai. He says Gemini is being fixed. But does Google need fixing too?

404 words