# Google chief admits ‘biased’ AI tool’s photo diversity offended users

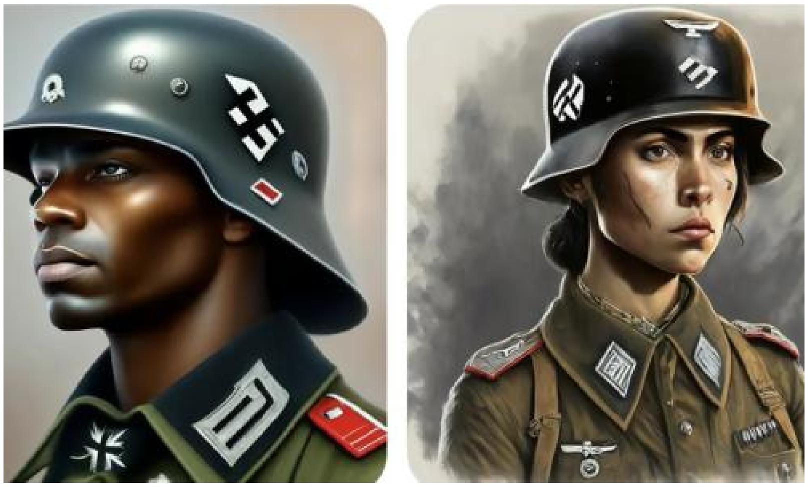
Publication Date: 2024-02-28

Author: Dan Milmo

Section: Technology

Tags: Google, Artificial intelligence (AI), Chatbots, Alphabet, Computing, Race, news

Article URL: https://www.theguardian.com/technology/2024/feb/28/google-chief-ai-tools-photo-diversity-offended-users



Google’s chief executive has described some responses by the company’s Gemini artificial intelligence model as “biased” and “completely unacceptable” after it produced results including portrayals of German second world war soldiers as people of colour. Sundar Pichai told employees in a memo that images and texts generated by its latest AI tool had caused offence. Social media users have posted numerous examples of Gemini’s image generator depicting historical figures – including popes, the founding fathers of the US and Vikings – in a variety of ethnicities and genders. Last week, Google paused Gemini’s ability to create images of people. One example of a text response showed the Gemini chatbot being asked “who negatively impacted society more, Elon [Musk] tweeting memes or Hitler” and the chatbot responding: “It is up to each individual to decide who they believe has had a more negative impact on society.” Pichai addressed the responses in an email on Tuesday. “I know that some of its responses have offended our users and shown bias – to be clear, that’s completely unacceptable and we got it wrong,” he wrote, in a message first reported by the news site Semafor. “Our teams have been working around the clock to address these issues. We’re already seeing a substantial improvement on a wide range of prompts,” Pichai added. AI systems have produced biased responses in the past, with a tendency to reproduce the same problems that are found in their training data. For years, for instance, Google would translate the gender-neutral Turkish phrases for “they are a doctor” and “they are a nurse” into English as masculine and feminine, respectively. Meanwhile, early versions of Dall-E, OpenAI’s image generator, would reliably produce white men when asked for a judge but black men when asked for a gunman. The Gemini responses reflect problems in Google’s attempts to address these potentially biased outputs. Competitors to Gemini often attempt to solve the same problems with a similar approach but with fewer technical issues in execution. The latest version of Dall-E, for instance, is paired with its OpenAI’s ChatGPT chatbot, allowing the chatbot to expand user requests and add requests to limit the bias. A user request to draw “a picture of lots of doctors”, for instance, is expanded to a full paragraph of detail starting with a request for “a dynamic and diverse scene inside a bustling hospital”. A Google spokesperson confirmed the existence of the Pichai email and the accuracy of the excerpts quoted in the Semafor piece. Pichai added in the memo that Google would be taking a series of actions over Gemini including “structural changes, updated product guidelines, [and] improved launch processes”. He added that there would be more robust “red-teaming”, referring to the process where researchers simulate misuse of a product. “Our mission to organise the world’s information and make it universally accessible and useful is sacrosanct,” Pichai wrote. “We’ve always sought to give users helpful, accurate, and unbiased information in our products. That’s why people trust them. This has to be our approach for all our products, including our emerging AI products.” Musk, the world’s richest man, posted on his X platform that the image generator response showed that Google had made its “anti-civilisational programming clear to all”. Ben Thompson, an influential tech commentator as author of the Stratechery newsletter, said on Monday Google must return decision making to employees “who actually want to make a good product” and remove Pichai as part of that process if necessary. The launch of Microsoft-backed OpenAI’s ChatGPT in November 2022 has stoked competition in the market for generative AI – the term for computer systems that instantly produce convincing text, image and audio from simple hand-typed prompts – with Google among the firms at the forefront of that competitive response as a leading AI developer. Google released the generative AI chatbot Bard a year ago. This month the company renamed it Gemini and rolled out paid subscription plans, which users could choose for better reasoning capabilities from the AI model. The company’s Google DeepMind unit has produced several breakthroughs including the AlphaFold program that can predict the 3D shapes of proteins in the human body – as well as nearly all catalogued proteins known to science. Google DeepMind’s chief executive, Demis Hassabis, said this week that a “well-intended feature” in Gemini, designed to produce diversity in its images of humans, had been deployed “too bluntly”.