

Title:

OpenAI says new model GPT-4 is more creative and less likely to invent facts

Latest version can take images as inputs and improves upon many of the criticisms users had, but will still ‘hallucinate’ facts

The artificial intelligence research lab OpenAI has released GPT-4, the latest version of the groundbreaking AI system that powers [ChatGPT](#), which it says is more creative, less likely to make up facts and less biased than its predecessor.

Calling it “our most capable and aligned model yet”, OpenAI cofounder Sam Altman said the new system is a “multimodal” model, which means it can accept images as well as text as inputs, allowing users to ask questions about pictures. The new version can handle massive text inputs and can remember and act on more than 20,000 words at once, letting it take an entire novella as a prompt.

The new model is available today for users of ChatGPT Plus, the paid-for version of the ChatGPT chatbot, which provided some of the training data for the latest release.

OpenAI has also worked with commercial partners to offer GPT-4-powered services. A new subscription tier of the language learning app Duolingo, Duolingo Max, will now offer English-speaking users AI-powered conversations in French or Spanish, and can use GPT-4 to explain the mistakes language learners have made. At the other end of the spectrum, payment processing company Stripe is using GPT-4 to answer support questions from corporate users and to help flag potential scammers in the company’s support forums.

“Artificial intelligence has always been a huge part of our strategy,” said Duolingo’s principal product manager, Edwin Bodge. “We had been using it for personalizing lessons and running Duolingo English tests. But there were gaps in a learner’s journey that we wanted to fill: conversation practice, and contextual feedback on mistakes.” The company’s experiments with GPT-4 convinced it that the technology was capable of providing those features, with “95%” of the prototype created within a day.

During a demo of GPT-4 on Tuesday, Open AI president and co-founder Greg Brockman also gave users a sneak peek at the image-recognition capabilities of the newest version of the system, which is not yet publicly available and only being tested by a company called Be My Eyes. The function will allow GPT-4 to analyze and respond to images that are submitted alongside prompts and answer questions or perform tasks based on those images. “GPT-4 is not just a language model, it is also a vision model,” Brockman said, “It can flexibly accept inputs that intersperse images and text arbitrarily, kind of like a document.”

At one point in the demo, GPT-4 was asked to describe why an image of a squirrel with a camera was funny. (Because “we don’t expect them to use a camera or act like a human”.) At another point, Brockman submitted a photo of a hand-drawn and rudimentary sketch of a website to GPT-4 and the system created a working website based on the drawing.

OpenAI claims that GPT-4 fixes or improves upon many of the criticisms that users had with the previous version of its system. As a “large language model”, GPT-4 is trained on vast amounts of data scraped from the internet and attempts to provide responses to sentences and questions that are statistically similar to those that already exist in the real world. But that can mean that it makes up information when it doesn’t know the exact answer – an issue known as “hallucination” – or that it provides upsetting or abusive responses when given the wrong prompts.

By building on conversations users had with ChatGPT, OpenAI says it managed to improve – but not eliminate – those weaknesses in GPT-4, responding sensitively to requests for content such as medical or self-harm advice “29% more often” and wrongly responding to requests for disallowed content 82% less often.

GPT-4 will still “hallucinate” facts, however, and OpenAI warns users: “Great care should be taken when using language model outputs, particularly in high-stakes contexts, with the exact protocol (such as human review, grounding with additional context, or avoiding high-stakes uses altogether) matching the needs of a specific use-case.” But it scores “40% higher” on tests intended to measure hallucination, OpenAI says.

The system is particularly good at not lapsing into cliché: older versions of GPT will merrily insist that the statement “you can’t teach an old dog new tricks” is factually accurate, but the newer GPT-4 will correctly tell a user who asks if you can teach an old dog new tricks that “yes, you can”.

More on this story

More on this story

AI race is disrupting education firms – and that is just the start

UK watchdog warns chatbot developers over data protection laws

Italy’s privacy watchdog bans ChatGPT over data breach concerns

AI chatbots making it harder to spot phishing emails, say experts

Bard: how Google's chatbot gave me a comedy of errors

Will Google's rush to join chatbot party with launch of Bard backfire?

Google's Bard chatbot launches in US and UK

AI makes plagiarism harder to detect, argue academics – in paper written by chatbot

Chinese ChatGPT rival from search engine firm Baidu fails to impress

The stupidity of AI

Most viewed

Most viewed

URL:

<https://www.theguardian.com/technology/2023/mar/14/chat-gpt-4-new-model>