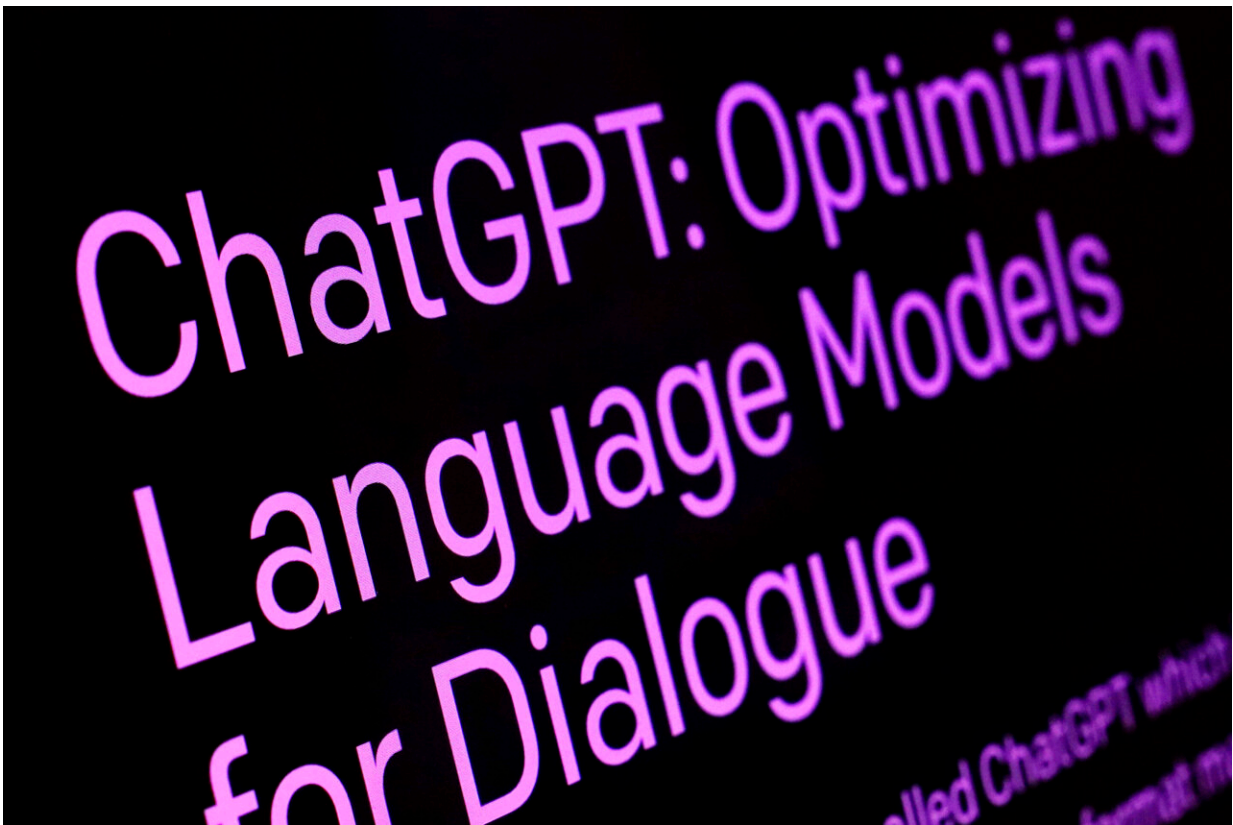


# What can ChatGPT maker's new AI model GPT-4 do?

March 15 2023, by Kelvin Chan

---



Text from the ChatGPT page of the OpenAI website is shown in this photo, in New York, Feb. 2, 2023. The company behind the ChatGPT chatbot has on Wednesday, March 15 rolled out its latest artificial intelligence model, GPT-4, in a new advance for the technology that's caught the world's attention. Credit: AP Photo/Richard Drew, File

The company behind the [ChatGPT chatbot](#) has rolled out its latest artificial intelligence model, GPT-4, in the next step for a technology that's caught the world's attention.

The new system can figure out [tax deductions](#) and [answer questions](#) like a Shakespearan pirate, for example, but it still "hallucinates" facts and makes reasoning errors.

Here's a look at San Francisco-based startup OpenAI's latest improvement on the generative AI models that can spit out readable text and [unique images](#):

## WHAT'S NEW?

OpenAI says GPT-4 "exhibits human-level performance." It's much more reliable, creative and can handle "more nuanced instructions" than its predecessor system, GPT-3.5, which ChatGPT was built on, OpenAI [said in its announcement](#).

In an online demo Tuesday, OpenAI President Greg Brockman ran through some scenarios that showed off GPT-4's capabilities that appeared to show it's a radical improvement on previous versions.

He demonstrated how the system could quickly come up with the proper income tax deduction after being fed reams of tax code—something he couldn't figure himself.

"It's not perfect, but neither are you. And together it's this amplifying tool that lets you just reach new heights," Brockman said.

## WHY DOES IT MATTER?

Generative AI technology like GPT-4 could be the future of the internet, at least according to Microsoft, which has [invested at least \\$1 billion in OpenAI](#) and made a splash by integrating AI chatbot tech into its Bing browser.

It's part of a new generation of machine-learning systems that can converse, [generate readable text on demand](#) and produce novel images and video based on what they've learned from a vast database of digital books and online text.

These new AI breakthroughs have the potential to transform the internet search business long dominated by Google, which is trying to [catch up with its own AI chatbot](#), and numerous professions.

"With GPT-4, we are one step closer to life imitating art," said Mirella Lapata, professor of natural language processing at the University of Edinburgh. She referred to the TV show "Black Mirror," which focuses on the dark side of technology.

"Humans are not fooled by the AI in 'Black Mirror' but they tolerate it," Lapata said. "Likewise, GPT-4 is not perfect, but paves the way for AI being used as a commodity tool on a daily basis."

## **WHAT EXACTLY ARE THE IMPROVEMENTS?**

GPT-4 is a "large multimodal model," which means it can be fed both text and images that it uses to come up with answers.

In one example posted on OpenAI's website, GPT-4 is asked, "What is unusual about this image?" Its answer: "The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi."

GPT-4 is also "steerable," which means that instead of getting an answer in ChatGPT's "classic" fixed tone and verbosity, users can customize it by asking for responses in the style of a Shakespearean pirate, for instance.

In his demo, Brockman asked both GPT-3.5 and GPT-4 to summarize in one sentence an article explaining the difference between the two systems. The catch was that every word had to start with the letter G.

GPT-3.5 didn't even try, spitting out a normal sentence. The newer version swiftly responded: "GPT-4 generates groundbreaking, grandiose gains, greatly galvanizing generalized AI goals."

## HOW WELL DOES IT WORK?

ChatGPT can write silly poems and songs or quickly explain just about anything found on the internet. It also gained notoriety for results that could be way off, such as confidently providing a [detailed but false account](#) of the Super Bowl game days before it took place, or even [being disparaging to users](#).

OpenAI acknowledged that GPT-4 still has limitations and warned users to be careful. GPT-4 is "still not fully reliable" because it "hallucinates" facts and makes reasoning errors, it said.

"Great care should be taken when using language model outputs, particularly in high-stakes contexts," the company said, though it added that hallucinations have been sharply reduced.

Experts also advised caution.

"We should remember that language models such as GPT-4 do not think in a human-like way, and we should not be misled by their fluency with

language," said Nello Cristianini, professor of artificial intelligence at the University of Bath.

Another problem is that GPT-4 does not know much about anything that happened after September 2021, because that was the cutoff date for the data it was trained on.

## **ARE THERE SAFEGUARDS?**

OpenAI says GPT-4's improved capabilities "lead to new risk surfaces" so it has improved safety by training it to [refuse requests](#) for sensitive or "disallowed" information.

It's less likely to [answer questions](#) on, for example, how to build a bomb or buy cheap cigarettes.

Still, OpenAI cautions that while "eliciting bad behavior" from GPT is harder, "doing so is still possible."

© 2023 The Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed without permission.

Citation: What can ChatGPT maker's new AI model GPT-4 do? (2023, March 15) retrieved 23 April 2024 from <https://techxplore.com/news/2023-03-chatgpt-maker-ai-gpt-.html>

<p>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.</p>
--