

Link:	<a href="https://www.nytimes.com/2024/09/12/technology/openai-chatgpt-math.html">https://www.nytimes.com/2024/09/12/technology/openai-chatgpt-math.html</a>
Medium:	The New York Times
Country:	USA
Language:	English
Date of publication:	12.09.2024
Author:	Cade Metz

## OpenAI Unveils New ChatGPT That Can Reason Through Math and Science

Driven by new technology called OpenAI o1, the chatbot can test various strategies and try to identify mistakes as it tackles complex tasks.

Online chatbots like ChatGPT from OpenAI and Gemini from Google sometimes struggle with simple math problems. The computer code they generate is often buggy and incomplete. From time to time, they even make stuff up.

On Thursday, OpenAI unveiled a new version of ChatGPT that could alleviate these flaws. The company said the chatbot, underpinned by new artificial intelligence technology called OpenAI o1, could “reason” through tasks involving math, coding and science.

“With previous models like ChatGPT, you ask them a question and they immediately start responding,” said Jakub Pachocki, OpenAI’s chief scientist. “This model can take its time. It can think through the problem — in English — and try to break it down and look for angles in an effort to provide the best answer.”

In a demonstration for The New York Times, Dr. Pachocki and Szymon Sidor, an OpenAI technical fellow, showed the chatbot solving an acrostic, a kind of word puzzle that is significantly more complex than an ordinary crossword puzzle. The chatbot also answered a Ph.D.-level chemistry question and diagnosed an illness based on a detailed report about a patient’s symptoms and history.

The new technology is part of a wider effort to build A.I. that can reason through complex tasks. Companies like Google and Meta are building similar technologies, while Microsoft and its subsidiary GitHub are working to incorporate OpenAI’s new system into their products.

The goal is to build systems that can carefully and logically solve a problem through a series of discrete steps, each one building on the next, similar to how humans reason. These technologies could be particularly useful to computer programmers who use A.I. systems to write code. They could also improve automated tutors for math and other subjects.

OpenAI said its new technology could also help physicists generate complicated mathematical formulas and assist health care researchers in their experiments.

With the debut of ChatGPT in late 2022, OpenAI showed that machines could handle requests more like people, answer questions, write term papers and even generate computer code. But the responses were sometimes flawed.

ChatGPT learned its skills by analyzing enormous amounts of text culled from across the internet, including Wikipedia articles, books and chat logs. By pinpointing patterns in all that text, it learned to generate text on its own.

(The New York Times sued OpenAI and Microsoft in December for copyright infringement of news content related to A.I. systems.)

Because the internet is filled with untruthful information, the technology learned to repeat the same untruths. Sometimes, it made things up.

Dr. Pachocki, Mr. Sidor and their colleagues have tried to reduce those flaws. They built OpenAI's new system using what is called reinforcement learning. Through this process — which can extend over weeks or months — a system can learn behavior through extensive trial and error.

By working through various math problems, for instance, it can learn which methods lead to the right answer and which do not. If it repeats this process with an enormously large number of problems, it can identify patterns. But the system cannot necessarily reason like a human. And it can still make mistakes and hallucinate.

"It is not going to be perfect," Mr. Sidor said. "But you can trust it will work harder and is that much more likely to produce the right answer."

Access to the new technology started Thursday for consumers and businesses that subscribe to the company's ChatGPT Plus and ChatGPT Teams services. The company is also selling the technology to software developers and businesses building their own A.I. applications.

OpenAI said the new technology performed better than previous technologies had on certain standardized tests. On the qualifying exam for the International Mathematical Olympiad, or I.M.O. — the premier math competition for high schoolers — its previous technology scored 13 percent. OpenAI o1, the company said, scored 83 percent.

Still, standardized tests are not always a good judge of how technologies will perform in real-world situations, and though the system might be good at a math test question, it could still struggle to teach math.

"There is a difference between problem solving and assistance," said Angela Fan, a research scientist at Meta. "New models that reason can solve problems. But that is very different than helping someone through their homework."