

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

1 **OpenAI Unveils New ChatGPT That Can Reason Through Math and Science**

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

4
5 Online chatbots like ChatGPT from OpenAI and Gemini from Google sometimes struggle with
6 simple math problems. The computer code they generate is often buggy and incomplete. From
7 time to time, they even make stuff up.

8
9 On Thursday, OpenAI unveiled a new version of ChatGPT that could alleviate these flaws. The
10 company said the chatbot, underpinned by new artificial intelligence technology called OpenAI
11 o1, could “reason” through tasks involving math, coding and science.

12
13 “With previous models like ChatGPT, you ask them a question and they immediately start
14 responding,” said Jakub Pachocki, OpenAI’s chief scientist. “This model can take its time. It can
15 think through the problem — in English — and try to break it down and look for angles in an
16 effort to provide the best answer.”

17
18 In a demonstration for The New York Times, Dr. Pachocki and Szymon Sidor, an OpenAI
19 technical fellow, showed the chatbot solving an acrostic, a kind of word puzzle that is
20 significantly more complex than an ordinary crossword puzzle. The chatbot also answered a
21 Ph.D.-level chemistry question and diagnosed an illness based on a detailed report about a
22 patient’s symptoms and history.

23
24 The new technology is part of a wider effort to build A.I. that can reason through complex tasks.
25 Companies like Google and Meta are building similar technologies, while Microsoft and its
26 subsidiary GitHub are working to incorporate OpenAI’s new system into their products.

27
28 The goal is to build systems that can carefully and logically solve a problem through a series of
29 discrete steps, each one building on the next, similar to how humans reason. These
30 technologies could be particularly useful to computer programmers who use A.I. systems to
31 write code. They could also improve automated tutors for math and other subjects.

32
33 OpenAI said its new technology could also help physicists generate complicated mathematical
34 formulas and assist health care researchers in their experiments.

35
36 With the debut of ChatGPT in late 2022, OpenAI showed that machines could handle requests
37 more like people, answer questions, write term papers and even generate computer code. But
38 the responses were sometimes flawed.

39
40 ChatGPT learned its skills by analyzing enormous amounts of text culled from across the
41 internet, including Wikipedia articles, books and chat logs. By pinpointing patterns in all that text,
42 it learned to generate text on its own.

43
44 (The New York Times sued OpenAI and Microsoft in December for copyright infringement of
45 news content related to A.I. systems.)

46

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

49

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

54

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

59

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

62

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

66

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

71

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

75

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