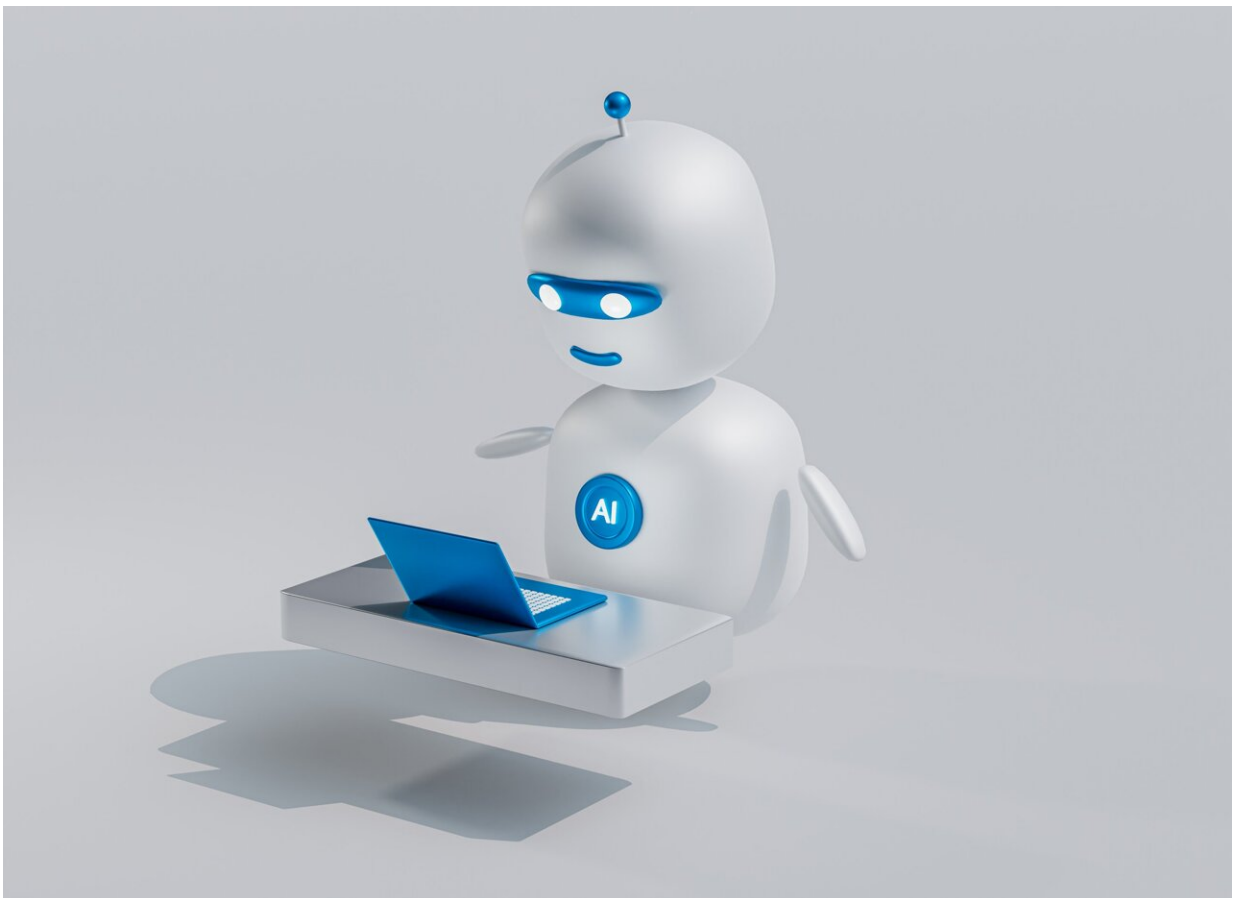


# GPT-3 can reason about as well as a college student, psychologists report

July 31 2023

---



Credit: Unsplash/CC0 Public Domain

People solve new problems readily without any special training or practice by comparing them to familiar problems and extending the

solution to the new problem. That process, known as analogical reasoning, has long been thought to be a uniquely human ability.

But now people might have to make room for a new kid on the block.

Research by UCLA psychologists shows that, astonishingly, the artificial intelligence language model GPT-3 performs about as well as college undergraduates when asked to solve the sort of reasoning problems that typically appear on [intelligence tests](#) and standardized tests such as the SAT. The study is published in *Nature Human Behaviour*.

But the paper's authors write that the study raises the question: Is GPT-3 mimicking human reasoning as a byproduct of its massive language training dataset or it is using a fundamentally new kind of cognitive process?

Without access to GPT-3's inner workings—which are guarded by OpenAI, the company that created it—the UCLA scientists can't say for sure how its reasoning abilities work. They also write that although GPT-3 performs far better than they expected at some reasoning tasks, the popular AI tool still fails spectacularly at others.

"No matter how impressive our results, it's important to emphasize that this system has major limitations," said Taylor Webb, a UCLA postdoctoral researcher in psychology and the study's first author. "It can do [analogical reasoning](#), but it can't do things that are very easy for people, such as using tools to solve a physical task. When we gave it those sorts of problems—some of which children can solve quickly—the things it suggested were nonsensical."

Webb and his colleagues tested GPT-3's ability to solve a set of problems inspired by a test known as Raven's Progressive Matrices, which ask the subject to predict the next image in a complicated

arrangement of shapes. To enable GPT-3 to "see," the shapes, Webb converted the images to a text format that GPT-3 could process; that approach also guaranteed that the AI would never have encountered the questions before.

The researchers asked 40 UCLA undergraduate students to solve the same problems.

"Surprisingly, not only did GPT-3 do about as well as humans but it made similar mistakes as well," said UCLA psychology professor Hongjing Lu, the study's senior author.

GPT-3 solved 80% of the problems correctly—well above the human subjects' average score of just below 60%, but well within the range of the highest human scores.

The researchers also prompted GPT-3 to solve a set of SAT analogy questions that they believe had never been published on the internet—meaning that the questions would have been unlikely to have been a part of GPT-3's training data. The questions ask users to select pairs of words that share the same type of relationships. (For example, in the problem "'Love' is to 'hate' as 'rich' is to which word?," the solution would be "poor.")

They compared GPT-3's scores to published results of college applicants' SAT scores and found that the AI performed better than the average score for the humans.

The researchers then asked GPT-3 and student volunteers to solve analogies based on short stories—prompting them to read one passage and then identify a different story that conveyed the same meaning. The technology did less well than students on those problems, although GPT-4, the latest iteration of OpenAI's technology, performed better

than GPT-3.

The UCLA researchers have developed their own computer model, which is inspired by human cognition, and have been comparing its abilities to those of commercial AI.

"AI was getting better, but our psychological AI model was still the best at doing analogy problems until last December when Taylor got the latest upgrade of GPT-3, and it was as good or better," said UCLA psychology professor Keith Holyoak, a co-author of the study.

The researchers said GPT-3 has been unable so far to solve problems that require understanding physical space. For example, if provided with descriptions of a set of tools—say, a cardboard tube, scissors and tape—that it could use to transfer gumballs from one bowl to another, GPT-3 proposed bizarre solutions.

"Language learning models are just trying to do word prediction so we're surprised they can do reasoning," Lu said. "Over the past two years, the technology has taken a big jump from its previous incarnations."

The UCLA scientists hope to explore whether language learning models are actually beginning to "think" like humans or are doing something entirely different that merely mimics human thought.

"GPT-3 might be kind of thinking like a human," Holyoak said. "But on the other hand, people did not learn by ingesting the entire internet, so the training method is completely different. We'd like to know if it's really doing it the way people do, or if it's something brand new—a real artificial intelligence—which would be amazing in its own right."

To find out, they would need to determine the underlying cognitive processes AI models are using, which would require access to the

software and to the data used to train the software—and then administering tests that they are sure the software hasn't already been given. That, they said, would be the next step in deciding what AI ought to become.

"It would be very useful for AI and cognitive researchers to have the backend to GPT models," Webb said. "We're just doing inputs and getting outputs and it's not as decisive as we'd like it to be."

**More information:** Taylor Webb, Emergent analogical reasoning in large language models, *Nature Human Behaviour* (2023). [DOI: 10.1038/s41562-023-01659-w](https://doi.org/10.1038/s41562-023-01659-w).  
[www.nature.com/articles/s41562-023-01659-w](https://www.nature.com/articles/s41562-023-01659-w)

Provided by University of California, Los Angeles

Citation: GPT-3 can reason about as well as a college student, psychologists report (2023, July 31) retrieved 27 April 2024 from <https://techxplore.com/news/2023-07-gpt-college-student-psychologists.html>

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.