

ChatGPT gave 10 different answers to the same question: Is its accuracy truly reliable?

Research shows that ChatGPT is inconsistent when answering the same question multiple times and struggles to determine correctness.

ChatGPT can provide very fluent and convincing answers, but a new study shows that the system still struggles to determine what is true.

Professor Mesut Cicek from Washington State University and his colleagues tested ChatGPT by incorporating hypotheses derived from scientific studies. The AI's task was to determine whether those hypotheses were supported by research evidence—in other words, whether they were true or false.

In total, the research team tested over 700 hypotheses, and each hypothesis was entered into the system up to 10 times to assess the consistency of the responses.

The accuracy is not yet truly reliable.

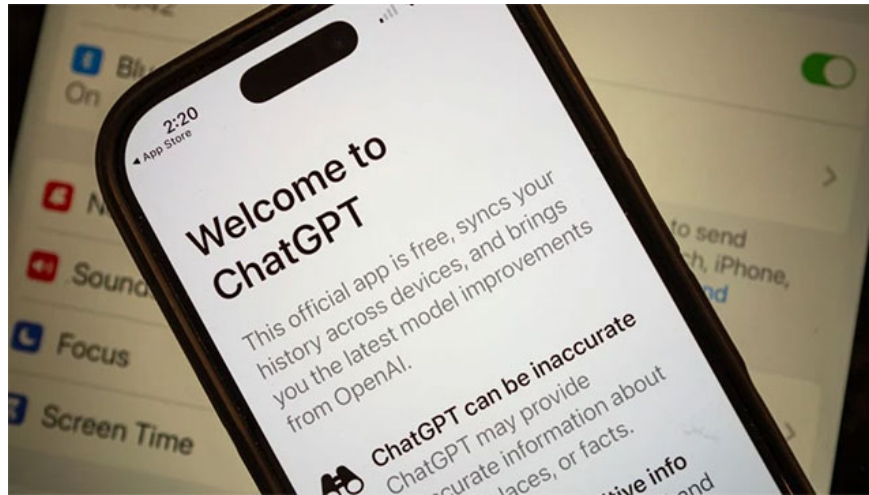
In the first experiment in 2024, ChatGPT answered approximately 76.5% of the questions correctly. When the study was repeated in 2025, this number increased slightly to 80%.

However, when the element of random guessing is removed, the results are far less promising. The actual performance of the AI is only about 60% higher than that of random guessing, which is considered quite low by academic standards.

In particular, the system faces significant difficulties in identifying false statements. The rate of correctly identifying inaccurate information is only about 16.4%.

Another notable issue is inconsistency. According to Professor Cicek, when the same question is asked multiple times, ChatGPT doesn't always provide the same answer. In about 73% of cases, the answer remains consistent, but the rest vary.

There are situations where the same question receives a series of alternating 'true – false – true – false' responses. In some cases, the number of 'true' and 'false' answers is even when repeated multiple times.



AI speaks well but doesn't truly 'understand' yet.

The research, published in the Rutgers Business Review, highlights a crucial point: the ability to generate smooth language does not necessarily mean that AI truly understands the issue.

According to Cicek, current AI systems lack the ability to think like humans. They primarily rely on memorizing and processing data, thereby generating answers that seem plausible but are not always accurate.

This suggests that generalized artificial intelligence (AGI) — with its true reasoning capabilities — may still be a long way from what many people expect.

The study involved many leading scientists. The team analyzed 719 hypotheses from scientific research papers in the business field published since 2021. Determining whether a hypothesis is proven is inherently complex, as it depends on many different factors.

In the experiment, the team used ChatGPT-3.5 in 2024 and ChatGPT-5 mini in 2025. The results showed no significant difference between the two versions.

The research reveals a key limitation of large-scale language models: while they can generate polished and persuasive answers, they struggle with deep reasoning. This leads to situations where AI can draw conclusions that sound plausible but are actually incorrect.

Why caution is needed when using AI.

Based on these findings, researchers recommend that users—especially managers and business leaders—verify information provided by AI, rather than blindly trusting it.

They also emphasized the importance of educating users to understand the strengths and limitations of AI.

Although the research focused on ChatGPT, similar tests with other AI systems yielded similar results. Previously, a 2024 survey also showed that consumers tend to be less interested in products when they are heavily promoted using AI.

You finished reading the article "**ChatGPT gave 10 different answers to the same question: Is its accuracy truly reliable?**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.
