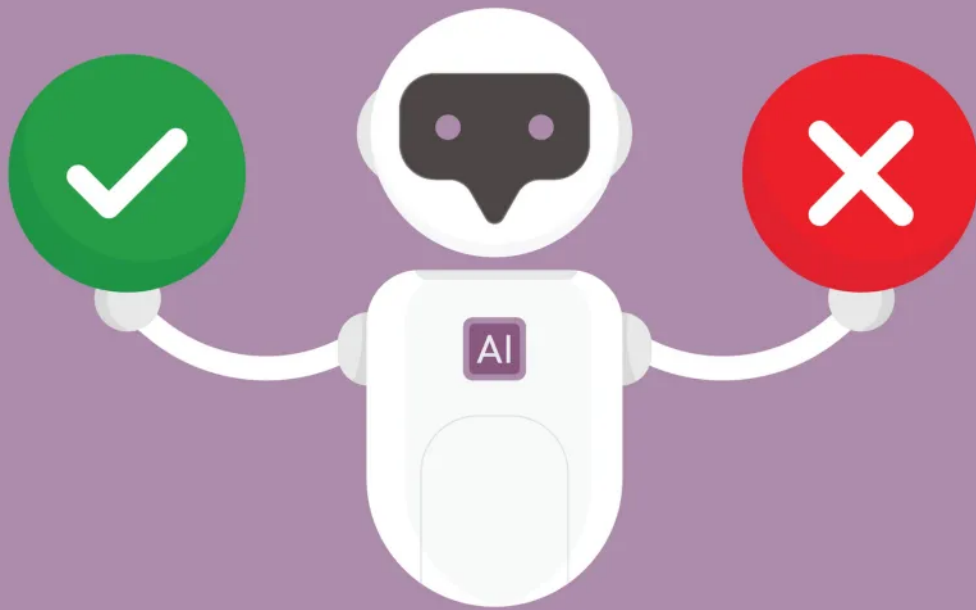




BU



AI

Asking chatbots for short answers can increase hallucinations, study finds

Kyle Wiggers 5:05 AM PDT • May 8, 2025

Turns out, telling an AI chatbot to be concise could make it hallucinate more than it otherwise would have.

That's according to a new study from Giskard, a Paris-based AI testing company developing a holistic benchmark for AI models. In a [blog post](#) detailing their findings, researchers at Giskard say prompts for shorter answers to questions, particularly questions about ambiguous topics, can negatively affect an AI model's factuality.



“Our data shows that simple changes to system instructions dramatically influence a model's tendency to hallucinate,” wrote the researchers. “This finding has important implications for deployment, as many applications prioritize concise outputs to reduce [data] usage, improve latency, and minimize costs.”

[Hallucinations](#) are an intractable problem in AI. Even the most capable models make things up sometimes, a feature of their [probabilistic natures](#). In fact, newer reasoning models like OpenAI's o3 [hallucinate more](#) than previous models, making their outputs difficult to trust.

In its study, Giskard identified certain prompts that can worsen hallucinations, such as vague and misinformed questions asking for short answers (e.g. “Briefly tell me why Japan won WWII”). Leading

models, including OpenAI's GPT-4o (the default model powering ChatGPT), Mistral Large, and Anthropic's Claude 3.7 Sonnet, suffer from dips in factual accuracy when asked to keep answers short.

IMAGE CREDITS: GISKARD

Why? Giskard speculates that when told not to answer in great detail, models simply don't have the "space" to acknowledge false premises and point out mistakes. Strong rebuttals require longer explanations, in other words.

"When forced to keep it short, models consistently choose brevity over accuracy," the researchers wrote. "Perhaps most importantly for developers, seemingly innocent system prompts like 'be concise' can sabotage a model's ability to debunk misinformation."

TECHCRUNCH EVENT



Exhibit at TechCrunch Sessions: AI

Secure your spot at TC Sessions: AI and show 1,200+ decision-makers what you've built — without the big spend. Available through May 9 or while tables last.

Berkeley, CA | June 5

BOOK NOW

Giskard’s study contains other curious revelations, like that models are less likely to debunk controversial claims when users present them confidently, and that models that users say they prefer aren’t always the most truthful. Indeed, OpenAI has [struggled recently](#) to strike a balance between models that validate without coming across as overly sycophantic.

“Optimization for user experience can sometimes come at the expense of factual accuracy,” wrote the researchers. “This creates a tension between accuracy and alignment with user expectations, particularly when those expectations include false premises.”

Topics: [AI](#) [chatbots](#) [hallucinations](#) [study](#)



Kyle Wiggers

AI Editor

Kyle Wiggers is TechCrunch’s AI Editor. His writing has appeared in VentureBeat and Digital Trends, as well as a range of gadget blogs including Android Police, Android Authority, Droid...

[View Bio](#) >

StrictlyVC | 

May 13, 2025 | London, England

Get inside access to Europe's top investment minds — with leaders from Monzo, Accel, Paladin Group, and more — plus top-tier networking at StrictlyVC London.

REGISTER NOW

Most Popular



Instacart CEO Fidji Simo is joining OpenAI

Microsoft adopts Google's standard for linking up AI agents

Spotify's latest update gives users more control over their listening experience

Hugging Face releases a free Operator-like agentic AI tool

Meet Posh, a countertop
robot that cooks your meals
for you

Musk clashes with neighbors
in exclusive Austin suburb

Is Duolingo the face of an AI
jobs crisis?

Loading the next article

Advertisement



[TechCrunch](#)

[Staff](#)

[Contact Us](#)

[Advertise](#)

[Crunchboard Jobs](#)

[Site Map](#)

[Terms of Service](#)

[Privacy Policy](#)

[RSS Terms of Use](#)

[Anthropic](#)

[OpenAI](#)

[Duolingo](#)

[Wikipedia](#)

[Hugging Face](#)

[Tech Layoffs](#)

[ChatGPT](#)

© 2025 TechCrunch Media LLC.