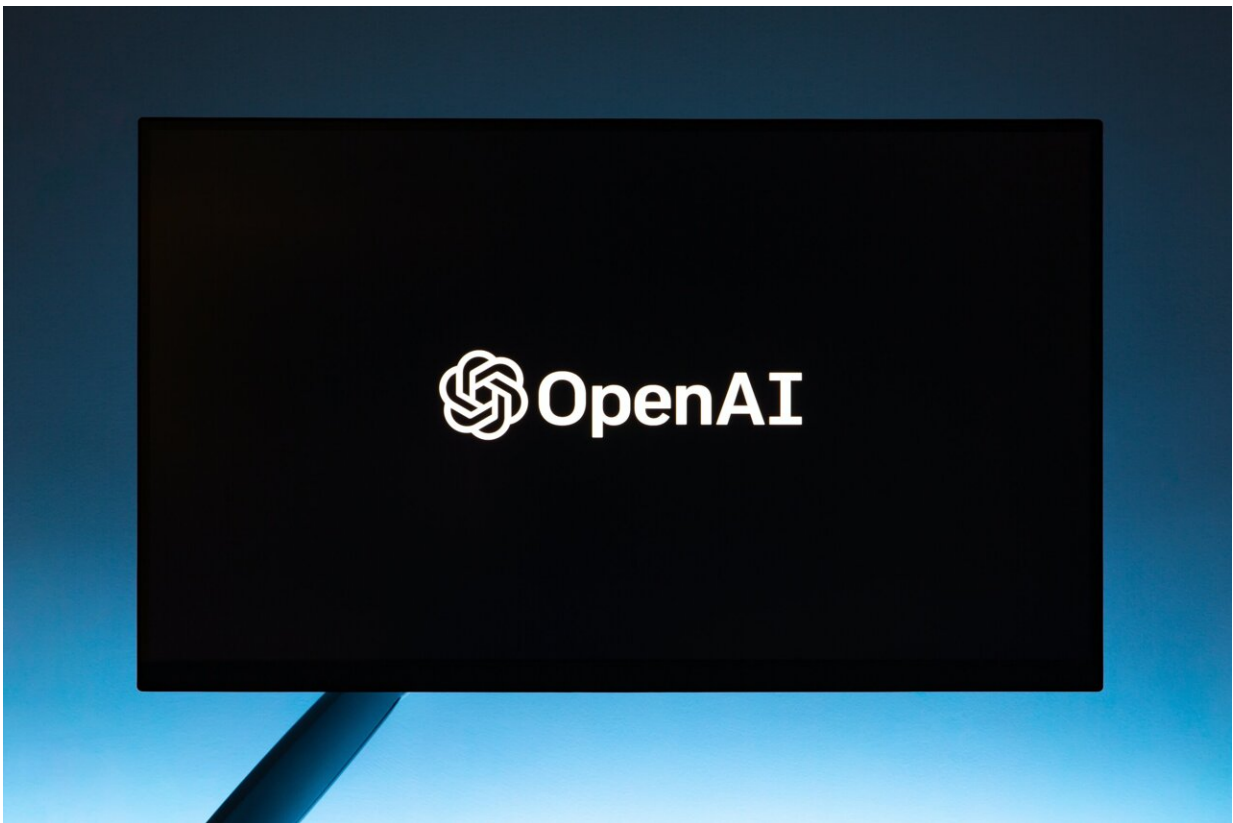


AI models struggle to identify nonsense, says study

September 14 2023



Credit: Unsplash/CC0 Public Domain

The AI models that power chatbots and other applications still have difficulty distinguishing between nonsense and natural language, according to a study released on Thursday.

The researchers at Columbia University in the United States said their work revealed the limitations of current AI models and suggested it was too early to let them loose in legal or medical settings.

They put nine AI models through their paces, firing hundreds of pairs of sentences at them and asking which were likely to be heard in [everyday speech](#).

They asked 100 people to make the same judgment on pairs of sentences like: "A buyer can own a genuine product also / One versed in circumference of highschool I rambled."

The research, published in the *Nature Machine Intelligence* journal, then weighed the AI answers against the human answers and found dramatic differences.

Sophisticated models like GPT-2, an earlier version of the model that powers viral chatbot ChatGPT, generally matched the human answers.

Other simpler models did less well.

But the researchers highlighted that all the models made mistakes.

"Every model exhibited blind spots, labeling some sentences as meaningful that [human participants](#) thought were gibberish," said psychology professor Christopher Baldassano, an author of the report.

"That should give us pause about the extent to which we want AI systems making [important decisions](#), at least for now."

Tal Golan, another of the paper's authors, told AFP that the models were "an exciting technology that can complement human productivity dramatically".

However, he argued that "letting these models replace human decision-making in domains such as law, medicine, or student evaluation may be premature".

Among the pitfalls, he said, was the possibility that people might intentionally exploit the [blind spots](#) to manipulate the models.

AI models burst into public consciousness with the release of ChatGPT last year, which has since been credited with passing various exams and has been touted as a possible aide to doctors, lawyers and other professionals.

More information: Testing the limits of natural language models for predicting human language judgements, *Nature Machine Intelligence* (2023). [DOI: 10.1038/s42256-023-00718-1](https://doi.org/10.1038/s42256-023-00718-1) , www.nature.com/articles/s42256-023-00718-1

© 2023 AFP

Citation: AI models struggle to identify nonsense, says study (2023, September 14) retrieved 6 May 2024 from <https://phys.org/news/2023-09-ai-struggle-nonsense.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.