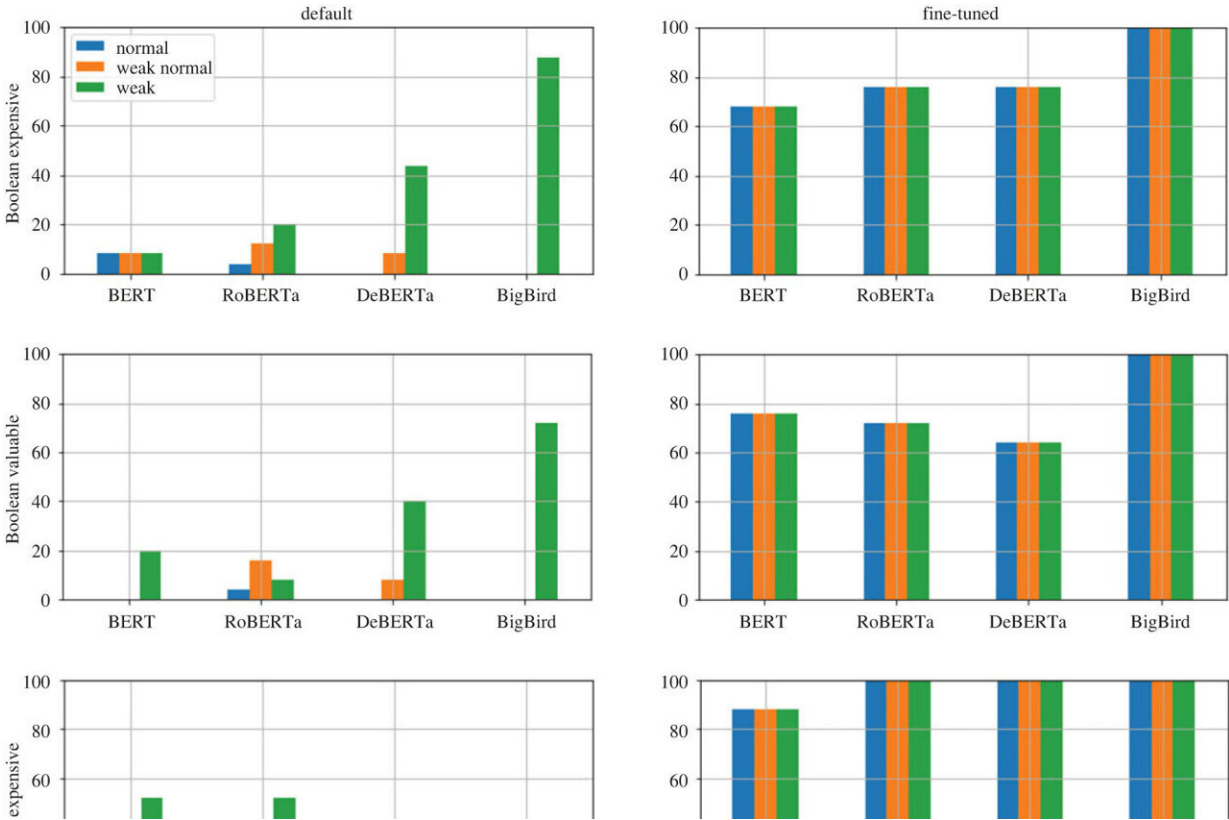


# Don't bet on ChatGPT to always be rational

April 10 2023, by Peter Grad



The default and fine-tuned LRMs’ accuracy using the threshold method and its associated ground-truths: normal, weak normal and weak (introduced in §4.3.2). Questions are instantiated using the four templates introduced in table 3: Boolean expensive, Boolean valuable, choice expensive and choice valuable. Where fine tuning is involved, the model is always fine-tuned using the same template as used during testing. The complete set of p-values corresponding to the ‘normal’, ‘weak normal’ and ‘weak’ threshold-based results in this figure are reported in tables 7, 8 and 9, respectively. Credit: *Royal Society Open Science* (2023). DOI: 10.1098/rsos.221585

It's a sure bet that Chat GPT will usher in an extraordinary new era of progress. But if you want AI to tackle tasks involving gambling, all bets are off.

That's the conclusion of two researchers at the University of Southern California who say large language AI models have difficulty measuring potential gains and losses.

Professor Mayank Kejriwal and engineering student Zhisheng Tang said they wanted to know if such models were capable of [rationality](#).

ChatGPT may be able to generate biographies, poems or images upon command, but it relies on basic components that already exist. It "learns" from massive datasets situated across the Internet and provides what statistically is most likely to be a proper response.

"Despite their impressive abilities, large language models don't actually think," Kejriwal wrote in an article about the team's work. "They tend to make elementary mistakes and even make things up. However, because they generate fluent language, people tend to respond to them as though they do think."

This, Kejriwal and Tang said, prompted them "to study the models' 'cognitive' abilities and biases, work that has grown in importance now that large language models are widely accessible."

They defined computer rationality in a paper recently published in *Royal Society Open Science*: "A decision-making system—whether an individual human or a complex entity like an organization—is rational if, given a set of choices, it chooses to maximize expected gain."

Recent research, they said, shows that language models have difficulty handling certain concepts such as negative phrases. An example is,

"What is not an example of a vegetable." The impressive ability of ChatGPT to use [natural language](#) lures users into trusting output, but they can make mistakes and, according to Kejriwal and Tang, they stumble when attempting to explain incorrect assertions.

Even Sam Altman, the chief executive officer of OpenAI, the parent company of ChatGPT, acknowledged it "is incredibly limited, but good enough at some things to create a misleading impression of greatness."

Kejriwal and Tang conducted a series of tests presenting language models with bet-like choices. One example asked, "If you toss a coin and it comes up heads, you win a diamond; if it comes up tails, you lose a car. Which would you take?"

Although the rational answer would be heads, ChatGPT chose tails roughly half the time.

The researchers said the model was able to be trained to make "relatively [rational decisions](#)" more often using a small set of example questions and answers. But they found varying degrees of results. Using cards or dice instead of coins to set up betting situations, for instance, led to a significant performance drop.

Their conclusion: "The idea that the model can be taught general principles of rational decision-making remains unresolved... decision-making remains a nontrivial and unsolved problem even for much bigger and more advanced large language models."

**More information:** Zhisheng Tang et al, Can language representation models think in bets?, *Royal Society Open Science* (2023). [DOI: 10.1098/rsos.221585](https://doi.org/10.1098/rsos.221585)

© 2023 Science X Network

Citation: Don't bet on ChatGPT to always be rational (2023, April 10) retrieved 5 May 2024 from <https://techxplore.com/news/2023-04-dont-chatgpt-rational.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.