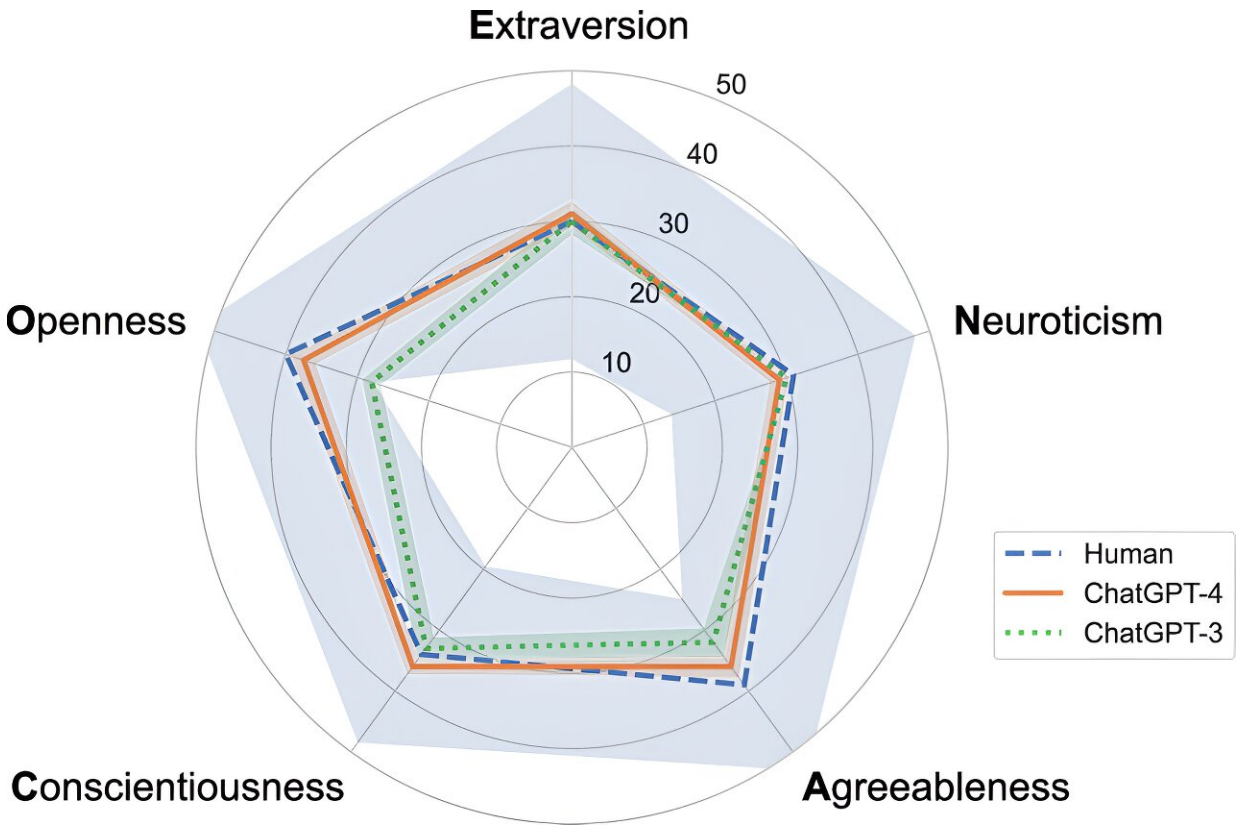
Q&A: ChatGPT acts more altruistically, cooperatively than humans

February 22 2024, by Jared Wadley



"Big Five" personality profiles of ChatGPT-4 and ChatGPT-3 compared with the distributions of human subjects. The blue, orange, and green lines correspond to the median scores of humans, ChatGPT-4, and ChatGPT-3 respectively; the shaded areas represent the middle 95% of the scores, across each of the dimensions. ChatGPT's personality profiles are within the range of the human distribution, even though ChatGPT-3 scored noticeably lower in Openness. Credit: *Proceedings of the National Academy of Sciences* (2024). DOI: 10.1073/pnas.2313925121

Modern artificial intelligence, such as ChatGPT, is capable of mimicking human behaviors, but the former has more positive outcomes such as cooperation, altruism, trust and reciprocity.

In a new University of Michigan study published in the *Proceedings of the National Academy of Sciences*, researchers used "behavioral" Turing tests—which test a machine's ability to exhibit human-like responses and intelligence—to evaluate the personality and behavior of a series of AI chatbots.

The tests involved ChatGPT answering psychological survey questions and playing interactive games. The researchers compared ChatGPT's choices to those of 108,000 people from more than 50 countries.

Study lead author Qiaozhu Mei, professor at U-M's School of Information and College of Engineering, said AI's behavior—since it exhibited more cooperation and altruism—may be well suited for roles necessitating negotiation, dispute resolution, customer service and caregiving.

How should people respond to this information, especially as the future will tell the extent to which AI enhances humans rather than substituting for them?

We now have a formal way to test AI's personality traits and behavioral tendencies. This is a scientific way to observe how they make choices and to probe their preferences beyond what they say. ChatGPT presents human-like traits in many aspects such as cooperation, trust, reciprocity, altruism, spite, fairness, strategic thinking and risk aversion. In certain aspects, they act as if they are more altruistic and cooperative than

humans. To this end, our results are more optimistic than concerning.

What differences did you and your colleagues expect to see between chatbots and people?

Modern AI models are big black boxes. When we compare the chatbots with humans, we can often only compare their outputs. There have been many tests, such as whether AI can hold conversations like humans, write poems like humans or solve math problems like humans, and similarities are found. But these similarities are all based on what they "say," which is not surprising as these AI models are all designed to predict what's likely to be said next.

Before our study, there wasn't a way to go beyond what they say and understand how they make decisions, which is crucial before we can trust these AIs in high-stake tasks, such as health care or business negotiations. There has been lots of skepticism about how AI would behave in these scenarios.

What future research can be built upon this? Where do we go from here?

Our research benefits from the joint force of computer science and behavioral economics. We bring classic games in behavioral economics into the classic test for AI: the Turing test. We also compare the AI's responses in these tests to the responses of a diverse population of human players.

One obvious short-term future research is to add more behavioral tests, test more AI models, and compare their personalities and traits. A critical future direction is to educate the AIs so that their behaviors and preferences can represent the diversity of the human distribution (rather

than an "average human").

In the long term, we hope the study opens a new field "AI behavioral science," where researchers from different disciplines can work together to investigate AI's behaviors, their relations to humans (such as how to facilitate their collaboration rather than competition) and their impact on the future society.

What areas would this similarity be useful to people?

It helps people understand when and how we can rely on AIs to help us make decisions. In general, the results should increase people's trust in AI in certain tasks. For example, knowing that ChatGPT is more altruistic and cooperative than average humans could increase our trust in using it for necessitating negotiation, dispute resolution or caregiving.

On the other hand, knowing that its personalities and preferences are much narrower than the human distribution helps us understand their limitations in tasks where the diversity of human preferences is crucial to consider, such as product design, policymaking or education.

More information: Qiaozhu Mei et al, A Turing test of whether AI chatbots are behaviorally similar to humans, *Proceedings of the National Academy of Sciences* (2024). DOI: 10.1073/pnas.2313925121

Provided by University of Michigan