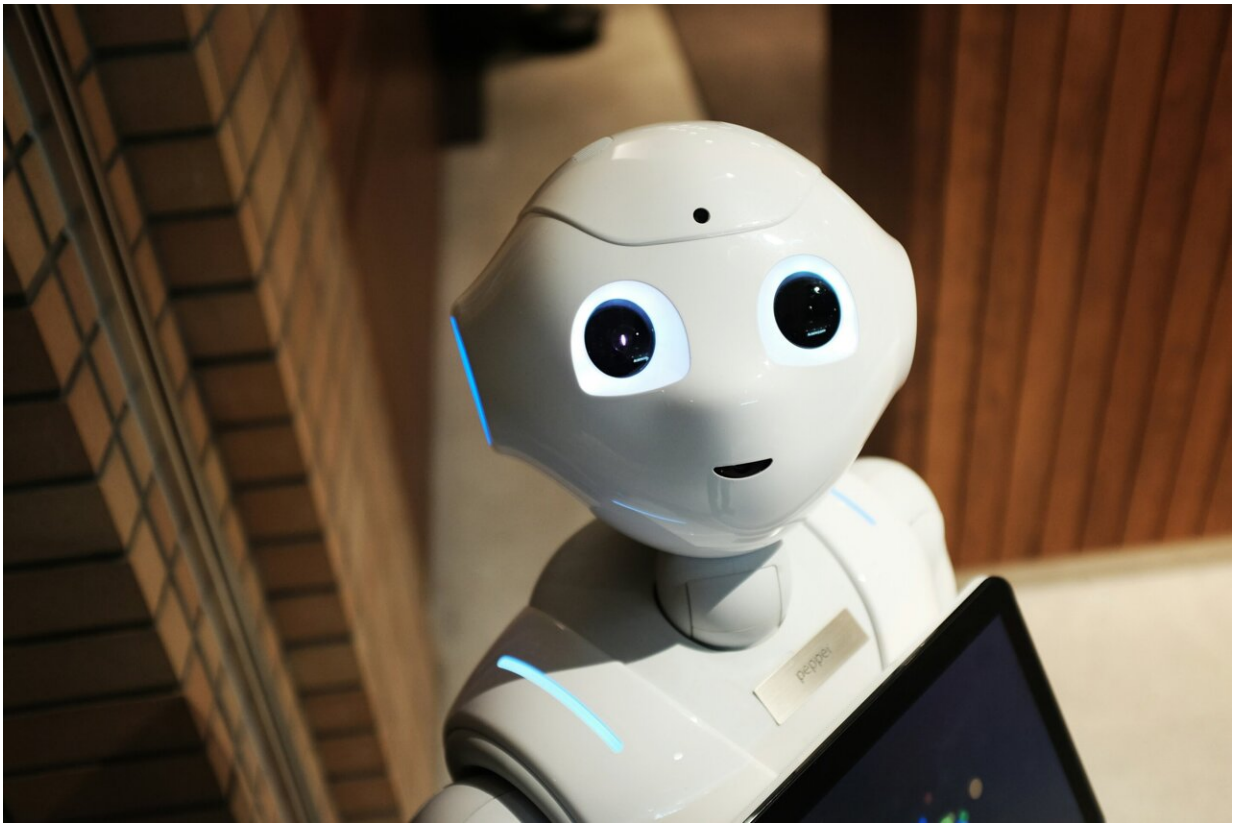# Will humans accept robots that can lie? Scientists find it depends on the lie

September 5 2024

Credit: Unsplash/CC0 Public Domain

Honesty is the best policy… most of the time. Social norms help humans understand when we need to tell the truth and when we shouldn't, to spare someone's feelings or avoid harm. But how do these norms apply

to robots, which are increasingly working with humans? To understand whether humans can accept robots telling lies, scientists asked almost 500 participants to rate and justify different types of robot deception.

"I wanted to explore an understudied facet of robot ethics, to contribute to our understanding of mistrust towards emerging technologies and their developers," said Andres Rosero, Ph.D. candidate at George Mason University and lead author of the study in *Frontiers in Robotics and AI.* "With the advent of generative AI, I felt it was important to begin examining possible cases in which anthropomorphic design and behavior sets could be utilized to manipulate users."

Three kinds of lie

The scientists selected three scenarios reflecting situations where robots already work—medical, cleaning, and retail work—and three different deception behaviors. These were external state deceptions, which lie about the world beyond the robot, hidden state deceptions, where a robot's design hides its capabilities, and superficial state deceptions, where a robot's design overstates its capabilities.

In the external state deception scenario, a robot working as a caretaker for a woman with Alzheimer's lies that her late husband will be home soon. In the hidden state deception scenario, a woman visits a house where a robot housekeeper is cleaning, unaware that the robot is also filming. Finally, in the superficial state deception scenario, a robot working in a shop as part of a study on human–robot relations untruthfully complains of feeling pain while moving furniture, causing a human to ask someone else to take the robot's place.

What a tangled web we weave

The scientists recruited 498 participants and asked them to read one of the scenarios and then answer a questionnaire. This asked participants whether they approved of the robot's behavior, how deceptive it was, if it could be justified, and if anyone else was responsible for the deception. These responses were coded by the researchers to identify common themes and analyzed.

The participants disapproved most of the hidden state deception, the housecleaning robot with the undisclosed camera, which they considered the most deceptive. While they considered the external state deception and the superficial state deception to be moderately deceptive, they disapproved more of superficial state deception, where a robot pretended it felt pain. This may have been perceived as manipulative.

Participants approved most of the external state deception, where the robot lied to a patient. They justified the robot's behavior by saying that it protected the patient from unnecessary pain—prioritizing the norm of sparing someone's feelings over honesty.

The ghost in the machine

Although participants were able to present justifications for all three deceptions—for instance, some people suggested the housecleaning robot might film for security reasons—most participants declared that the hidden state deception could not be justified. Similarly, about half the participants responding to the superficial state deception said it was unjustifiable. Participants tended to blame these unacceptable deceptions, especially hidden state deceptions, on robot developers or owners.

"I think we should be concerned about any technology that is capable of withholding the true nature of its capabilities, because it could lead to users being manipulated by that technology in ways the user (and

perhaps the developer) never intended," said Rosero.

"We've already seen examples of companies using web design principles and artificial intelligence chatbots in ways that are designed to manipulate users towards a certain action. We need regulation to protect ourselves from these harmful deceptions."

However, the scientists cautioned that this research needs to be extended to experiments which could model real-life reactions better—for example, videos or short roleplays.

"The benefit of using a cross-sectional study with vignettes is that we can obtain a large number of participant attitudes and perceptions in a cost-controlled manner," explained Rosero. "Vignette studies provide baseline findings that can be corroborated or disputed through further experimentation. Experiments with in-person or simulated human-robot interactions are likely to provide greater insight into how humans actually perceive these robot deception behaviors."

Provided by Frontiers