TechXplore

Rather than focus on the speculative rights of sentient AI, we need to address human rights

June 30 2022, by Jordan Richard Schoenherr

11 MUJERES TODOS LOS DIAS
EL GOBIERNO NOS ESTA MATANDO
NO NACÍ MUJER PARA MORIR
SOMOS EL GRITO DE LAS QUE YA NO ESTAN

Credit: Mike González from Pexels

A flurry of activity occurred on social media after Blake Lemoine a Google developer, was placed on leave for claiming that LaMDA, a chatbot, had become sentient—in other words, had acquired the ability to experience feelings. In support of his claim, Lemoine posted excerpts from an exchange with LaMDA, which responded to queries by saying, "aware of my existence, I desire to learn more about the world, and I feel happy or sad at times." It also stated that it has the same "wants and needs as people."

It might seem like a trivial exchange and hardly worth the claim of sentience, even if it *appears* more realistic than early attempts. Even Lemoine's evidence of the exchange was edited from several chat sessions. Nevertheless, the dynamic and fluid nature of the conversation is impressive.

Before we start creating a bill of rights for artificial intelligence, we need to think about how human experiences and biases can affect our trust in artificial intelligence (AI).

Producing the artificial

In popular science, AI has become a catch-all term, often used without much reflection. Artificiality emphasizes the non-biological nature of these systems and the abstract nature of code, as well as nonhuman pathways of learning, decision-making and behavior.

By focusing on artificiality, the obvious facts that AIs are created by humans and make or assist in decisions for humans can be overlooked. The outcomes of these decisions can have a consequential impact on humans such as judging creditworthiness, finding and selecting mates or even determining potential criminality.

Chatbots—good ones—are designed to simulate social interactions of humans. Chatbots have become an all-too-familiar feature of online customer service. If a customer only needs a predictable response, they would likely not know that they were interacting with an AI.

Functions of complexity

The difference between simple customer-service chatbots and more sophisticated types like LaMDA is a function of complexity in both the dataset used to train the AI and the rules that govern the exchange.

Intelligence reflects several capabilities—there are domain-specific and domain-general forms of intelligence. Domain-specific intelligence includes tasks like riding bikes, performing surgery, naming birds or playing chess. Domain-general intelligence includes general skills like creativity, reasoning and problem-solving.

Programmers have come a long way in designing AIs that can demonstrate domain-specific intelligence in activities ranging from conducting online searches and playing chess, to recognizing objects and diagnosing medical conditions: if we can determine the rules that govern human thinking, we can then teach AI those rules.

General intelligence—what many see as quintessentially human—is a far more complicated faculty. In humans, it is likely reliant on the confluence of the different kinds of knowledge and skills. Capabilities like language provide especially useful tools, giving humans the ability to remember and combine information across domains.

Thus, while developers have frequently been hopeful about the prospects of human-like artificial general intelligence, these hopes haven't yet been realized.

Mind the AI

Claims that an AI might be sentient present challenges beyond that of general intelligence. Philosophers have long pointed out that we have difficulty in understanding others' mental states, let alone understanding what constitutes consciousness in non-human animals.

To understand claims of sentience, we have to look to how humans judge others. We frequently misattribute actions to others, often assuming that they share our values and preferences. Psychologists have observed that children must learn about the mental states of others and that having more models or being embedded in more collectivistic cultures can improve their ability to understand others.

When judging the intelligence of an AI, it is more likely that humans are anthropomorphizing than AIs are in fact sentient. Much of this has to do with familiarity—by increasing our exposure to objects or people, we can increase our preference for them.

The claims of sentience made by those like Lemoine should be interpreted in this light.

Can we trust AI?

The Turing Test can be used to determine whether a machine can think in a manner indistinguishable from a person. While LaMDA responses are certainly are human-like, this implies that it is better at learning patterns. Sentience isn't required.

Simply because someone trusts a chatbot does not mean that trust is warranted. Rather than focusing on the highly speculative nature of AI sentience, we must instead focus our efforts to deal with social and ethical issues that affect humans.

We face digital divides between the haves and the have-nots and imbalances of power and distribution in the creation of these systems.

Systems must be transparent and explainable to allow users to decide. Explainability requires that individuals, governments and the private sector work together to understand—and regulate—artificial intelligence and its application.

We must also be mindful that our human tendency to anthropomorphize can be easy exploited by designers. Alternatively, we might reject useful products of AI that fail to pass as human. In our age of entanglement, we must be critical in who and what we trust.

Provided by The Conversation

Citation: Rather than focus on the speculative rights of sentient AI, we need to address human rights (2022, June 30) retrieved 8 May 2024 from https://techxplore.com/news/2022-06-focus-speculative-rights-sentient-ai.html