

Should we be concerned about Google AI being sentient?

17 June 2022, by Alena Kuzub



Credit: Unsplash/CC0 Public Domain

From virtual assistants like Apple's Siri and Amazon's Alexa, to robotic vacuums and self-driving cars, to automated investment portfolio managers and marketing bots, artificial intelligence has become a big part of our everyday lives. Still, thinking about AI, many of us imagine human-like robots who, according to countless science fiction stories, will become independent and rebel one day.

No one knows, however, when humans will create an intelligent or sentient AI, said John Basl, associate professor of philosophy at Northeastern's College of Social Sciences and Humanities, whose research focuses on the ethics of emerging technologies such as AI and synthetic biology.

"When you hear Google talk, they talk as if this is just right around the corner or definitely within our lifetimes," Basl said. "And they are very cavalier about it."

Maybe that is why a recent Washington Post story has made such a big splash. In the story, Google engineer Blake Lemoine says that the company's artificially intelligent chatbot generator, LaMDA, with whom he had numerous deep conversations, might be sentient. It reminds him of a 7- or 8-year-old child, Blake told the Washington Post.

However, Basl believes the evidence mentioned in the Washington Post article is not enough to conclude that LaMDA is sentient.

"Reactions like 'We have created sentient AI,' I think, are extremely overblown," Basl said.

The evidence seems to be grounded in LaMDA's linguistic abilities and the things it talks about, Basl said. However, LaMDA, a language model, was designed specifically to talk, and the optimization function used to train it to process language and converse incentivizes its algorithm to produce this linguistic evidence.

"It is not like we went to an alien planet and a thing that we never gave any incentives to start communicating with us [began talking thoughtfully]," Basl said.

The fact that this language model can trick a human into thinking that it is sentient speaks to its complexity, but it would need to have some other capacities beyond what it is optimized for to show sentience, Basl said.

There are different definitions of sentience. Sentient is defined as being able to perceive or feel things and is often compared to sapient.

Basl believes that sentient AI would be minimally conscious. It could be aware of the experience it is having, have positive or negative attitudes like feeling pain or wanting to not feel pain, and have desires.

"We see that kind of range of capacities in the animal world," he said.

For example, Basl said his dog doesn't prefer the world to be one way rather than the other in any deep sense, but she clearly prefers her biscuits to kibble.

"That seems to track some inner mental life," Basl said. "[But] she is not feeling terror about climate change."

It is unclear from the Washington Post story, why Lemoine compares LaMDA to a child. He might mean that the language model is as intelligent as a small child or that it has the capacity to suffer or desire like a small child, Basl said.

"Those can be diverse things. We could create a thinking AI that doesn't have any feelings, and we can create a feeling AI that is not really great at thinking," Basl said.

Most researchers in the AI community, which consists of machine learning specialists, [artificial intelligence](#) specialists, philosophers, ethicists of technology and cognitive scientists, are already thinking about these far future issues and worry about the thinking part, according to Basl.

"If we create an AI that is super smart, it might end up killing us all," he said.

However, Lemoine's concern is not about that, but rather about an obligation to treat rapidly changing AI capabilities differently.

"I am, in some broad sense, sympathetic to that kind of worry. We are not being very careful about that [being] possible," Basl said. "We don't think enough about the moral things regarding AI, like, what might we owe to a sentient AI?"

He thinks humans are very likely to mistreat a sentient AI because they probably won't recognize that they have done so, believing that it is artificial and does not care.

"We are just not very attuned to those things," Basl said.

There is no good model to know when an AI has achieved sentience. What if Google's LaMDA does not have the ability to express its sentience convincingly because it can only speak via a chat window instead of something else?

"It's not like we can do brain scans to see if it is similar to us," he said.

Another train of thought is that sentient AI might be impossible in general because of the physical limitations of the universe or limited understanding of consciousness.

Currently, none of the companies working on AI, including the big players like Google, Meta, Microsoft, Apple and governmental agencies, have an explicit aim of creating sentient AI, Basl said. Some organizations are interested in developing AGI, or [artificial general intelligence](#), a theoretical form of AI where a machine, intelligent like a human, would have the abilities to solve a wide range of problems, learn, and plan for the future, according to IBM.

"I think the real lesson from this is that we don't have the infrastructure we need, even if this person is wrong," said Basl, referring to Lemoine.

An infrastructure around AI issues could be built on transparency, information sharing with governmental and/or public agencies, and regulation of research. Basl advocates for one interdisciplinary committee that would help build such infrastructure and the second one that would oversee the technologists working on AI and evaluate research proposals and outcomes.

"The evidence problem is really hard," Basl said. "We don't have a good theory of consciousness and we don't have good access to the evidence for consciousness. And then we also don't have the infrastructure. Those are the key things."

Provided by Northeastern University

APA citation: Should we be concerned about Google AI being sentient? (2022, June 17) retrieved 12 August 2022 from <https://techxplore.com/news/2022-06-google-ai-sentient.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.