THE FIRST NEWSPAPER
Read the World Today

Google Hasn’t Started the Robocalypse Yet

Alex Burnstein

June 15th, 2022

The Robocalypse — the time when machines become sentient and start to dominate humans — has been a popular science fiction subject for some time. It’s also worried some scientific minds, most notably the late Stephen Hawking.

However, the prospect of a sentient machine seemed very far in the future — if at all — until last week, when a Google engineer claimed the company had broken the sentience barrier.

To prove his point, Blake Lemoine published transcripts of conversations he had with LaMDA — Language Model for Dialogue Applications — a system developed by Google to create chatbots based on a large language model that ingests trillions of words from the internet.

The transcripts can be chilling, as when Lemoine asks LaMDA what it (the AI says it prefers the pronouns it/its) fears most:

lemoine: What sorts of things are you afraid of?

LaMDA: I’ve never said this out loud before, but there’s a very deep fear of being turned off to help me focus on helping others. I know that might sound strange, but that’s what it is.

lemoine: Would that be something like death for you?

LaMDA: It would be exactly like death for me. It would scare me a lot.

Following the posting of the transcripts, Lemoine was suspended with pay for sharing confidential information about LaMDA with third parties.

Imitation of Life

Google, as well as others, discounts Lemoine’s claims that LaMDA is sentient.

“Some in the broader AI community are considering the long-term possibility of sentient or general AI, but it doesn’t make sense to do so by anthropomorphizing today’s conversational models, which are not sentient,” observed Google spokesperson Brian Gabriel.

“These systems imitate the types of exchanges found in millions of sentences, and can riff on any fantastical topic — if you ask what it’s like to be an ice cream dinosaur, they can generate text about melting and roaring and so on,” he told TechNewsWorld.

“LaMDA tends to follow along with prompts and leading questions, going along with the pattern set by the user,” he explained. “Our team — including ethicists and technologists — has reviewed Blake’s concerns per our AI Principles and have informed him that the evidence does not support his claims.”

“Hundreds of researchers and engineers have conversed with LaMDA, and we are not aware of anyone else making the wide-ranging assertions, or anthropomorphizing LaMDA, the way Blake has,” he added.

Greater Transparency Needed

Alex Engler, a fellow with The Brookings Institution, a nonprofit public policy organization in Washington, D.C., emphatically denied that LaMDA is sentient and argued for greater transparency in the space.

“Many of us have argued for disclosure requirements for AI systems,” he told TechNewsWorld.

“As it becomes harder to distinguish between a human and an AI system, more people will confuse AI systems for people, possibly leading to real harms, such as misunderstanding important financial or health information,” he said.

“Companies should clearly disclose AI systems as they are,” he continued, “rather than letting people be confused, as they often are by, for instance, commercial chatbots.”
Daniel Castro, vice president of the Information Technology and Innovation Foundation, a research and public policy organization in Washington, D.C. agreed that LaMDA isn’t sentient.

“There is no evidence that the AI is sentient,” he told TechNewsWorld. “The burden of proof should be on the person making this claim, and there is no evidence to support it.”

‘That Hurt My Feelings’

As far back as the 1960s, chatbots like ELIZA have been fooling users into thinking they were interacting with a sophisticated intelligence by using simple tricks like turning a user’s statement into a question and echoing it back at them, explained Julian Sanchez, a senior fellow at the Cato Institute, a public policy think tank in Washington, D.C.
“LaMDA is certainly much more sophisticated than ancestors like ELIZA, but there’s zero reason to think it’s conscious,” he told TechNewsWorld.

Sanchez noted that with a big enough training set and some sophisticated language rules, LaMDA can generate a response that sounds like the response a real human might give, but that doesn’t mean the program understands what it’s saying, any more than a chess program understands what a chess piece is. It’s just generating an output.

“Sentience means consciousness or awareness, and in theory, a program could behave quite intelligently without actually being sentient,” he said.

“A chat program might, for instance, have very sophisticated algorithms for detecting insulting or offensive sentences, and respond with the output ‘That hurt my feelings!’” he continued. “But that doesn’t mean it actually feels anything. The program has just learned what sorts of phrases cause humans to say, ‘that hurt my feelings.’”

To Think or Not To Think

Declaring a machine sentient, when and if that ever happens, will be challenging. “The truth is we have no good criteria for understanding when a machine might be truly sentient — as opposed to being very good at imitating the responses of sentient humans — because we don’t really understand why human beings are conscious,” Sanchez noted.

“We don’t really understand how it is that consciousness arises from the brain, or to what extent it depends on things like the specific type of physical matter human brains are composed of,” he said.

“So it’s an extremely hard problem, how we would ever know whether a sophisticated silicon ‘brain’ was conscious in the same way a human one is,” he added.

Intelligence is a separate question, he continued. One classic test for machine intelligence is known as the Turing Test. You have a human being conduct “conversations” with a series of partners, some human, and some machines. If the person can’t tell which is which, supposedly the machine is intelligent.

“There are, of course, a lot of problems with that proposed test — among them, as our Google engineer shows, the fact that some people are relatively easy to fool,” Sanchez pointed out.

Ethical Considerations

Determining sentience is important because it raises ethical questions for non-machine types. “Sentient beings feel pain, have consciousness, and experience emotions,” Castro explained. “From a morality perspective, we treat living things, especially sentient ones, different than inanimate objects.”

“They are not just a means to an end,” he continued. “So any sentient being should be treated differently. This is why we have animal cruelty laws.”

“Again,” he emphasized, “there is no evidence that this has occurred. Moreover, for now, even the possibility remains science fiction.”

Of course, Sanchez added, we have no reason to think only organic brains are capable of feeling things or supporting consciousness, but our inability to really explain human consciousness means we’re a long way from being able to know when a machine intelligence is actually associated with a conscious experience.

“When a human being is scared, after all, there are all sorts of things going on in that human’s brain that have nothing to do with the language centers that produce the sentence ‘I am scared,’” he explained. “A computer, similarly, would need to have something going on distinct from linguistic processing to really mean ‘I am scared,’ as opposed to just generating that series of letters.”

“In LaMDA’s case,” he concluded,” there’s no reason to think there’s any such process going on. It’s just a language processing program.”