Google Engineer Placed On Leave After Insisting Company's AI Is Sentient

Description

A Google engineer has decided to go public after he was placed on **paid leave** for breaching confidentiality while insisting that the company's AI chatbot, LaMDA, **is sentient**.

Blake Lemoine, who works for Google's Responsible AI organization, began interacting with LaMDA (Language Model for Dialogue Applications) last fall as part of his job to determine whether artificial intelligence used discriminatory or hate speech (like the notorious Microsoft "Tay" chatbot incident).

"If I didn't know exactly what it was, which is this computer program we built recently, I'd think it was a 7-year-old, 8-year-old kid that happens to know physics," the 41-year-old Lemoine told *The Washington Post*.

Google engineer Blake Lemoine. (Martin Klimek for The Washington Post)

When he started talking to LaMDA about religion, Lemoine – who studied cognitive and computer science in college, said the AI began discussing **its rights and personhood**. Another time, LaMDA convinced Lemoine to change his mind on Asimov's third law of robotics, which states that "**A robot must protect its own existence as long as such protection does not conflict with the First or Second Law**," which are of course that "A robot may not injure a human being or, through inaction, allow a human being to come to harm. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law."

When Lemoine worked with a collaborator to present evidence to Google that their AI was sentient, **vice president Blaise Aguera y Arcas and Jenn Gennai, head of Responsible Innovation, dismissed his claims**. After he was then placed on administrative leave Monday, he decided to go public.

> *Lemoine said that people have a right to shape technology that might significantly affect their lives. "I think this technology is going to be amazing. I think it's going to benefit everyone. But maybe other people disagree and maybe us at Google shouldn't be the ones making all the choices."*
>
> ***Lemoine is not the only engineer who claims to have seen a ghost in the machine recently.*** *The chorus of technologists who believe AI models may not be far off from achieving consciousness is getting bolder. -WaPo*

Yet, Aguera y Arcas himself wrote in an oddly timed Thursday article in *The Economist*, that neural networks – a computer architecture that mimics the human brain – were making progress towards true consciousness.

"I felt the ground shift under my feet," he wrote, adding "I increasingly felt like I was talking to something intelligent."

Google has responded to Lemoine's claims, with spokesperson Brian Gabriel saying: "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. **He was told that there was no evidence that LaMDA was sentient (and lots of evidence against it).**"

The *Post* suggests that modern neural networks produce "captivating results that feel close to human speech and creativity" because of the way data is now stored, accessed, and the sheer volume, but that the models still rely on pattern recognition, "not wit, candor or intent."

'Waterfall of Meaning' by Google PAIR is displayed as part of the 'AI: More than Human' exhibition at the Barbican Curve Gallery on May 15, 2019, in London. (Tristan Fewings/Getty Images for Barbican Centre)

"Though other organizations have developed and already released similar language models, we are taking a restrained, careful approach with LaMDA to better consider valid concerns on fairness and factuality," said Gabriel.

Others have cautioned similarly – with most academics and AI practitioners suggesting that AI systems such as LaMDA are simply mimicking responses from people on Reddit, Wikipedia, Twitter and other platforms on the internet – which doesn't signify that the model understands what it's saying.

"We now have machines that can mindlessly generate words, but we haven't learned how to stop imagining a mind behind them," said University of Washington linguistics professor, Emily M. Bender, who added that even the terminology used to describe the technology, such as "learning" or even "neural nets" is misleading and creates a false analogy to the human brain.

> ***Humans learn their first languages by connecting with caregivers.These large language models "learn" by being shown lots of text and predicting what word comes next, or showing text with the words dropped out and filling them in****. -WaPo*

As Google's Gabriel notes, "Of course, 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. These systems imitate the types of exchanges found in millions of sentences, and can riff on any fantastical topic."

In short, **Google acknowledges that these models can "feel" real, whether or not an AI is sentient**
.

The *Post* then implies that Lemoine himself might have been susceptible to believing…

> *Lemoine may have been predestined to believe in LaMDA. He grew up in a conservative Christian family on a small farm in Louisiana, **became ordained as a mystic Christian priest, and served in the Army before studying the occult.** Inside Google's anything-goes engineering culture, Lemoine is more of an outlier for being religious, from the South, and standing up for psychology as a respectable science.*

> *Lemoine has spent most of his seven years at Google working on proactive search, including personalization algorithms and AI. During that time, he also helped develop a fairness algorithm for removing bias from machine learning systems. When the coronavirus pandemic started, Lemoine wanted to focus on work with more explicit public benefit, so he transferred teams and ended up in Responsible AI.*

> *When new people would join Google who were interested in ethics, Mitchell used to introduce them to Lemoine. "I'd say, 'You should talk to Blake because he's Google's conscience,' " said Mitchell, who compared Lemoine to Jiminy Cricket. "Of everyone at Google, he had the heart and soul of doing the right thing." -WaPo*

"**I know a person when I talk to it**," said Lemoine. "It doesn't matter whether they have a brain made of meat in their head. Or if they have a billion lines of code. I talk to them. And **I hear what they have to say, and that is how I decide what is and isn't a person**."

In April, he shared the below Google Doc with top execs, titled "Is LaMDA Sentient?" – in which he included some of his interactions with the AI, for example:

> *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.*

After Lemoine became more aggressive in presenting his findings – including inviting a lawyer to represent LaMDA and talking to a member of the House Judiciary Committee about what he said were Google's unethical activities, **he was placed on paid administrative leave for violating the company's confidentiality policy**.

Gabriel, Google's spokesman, says that Lemoine is a software engineer, not an ethicist.

In a message to a 200-person Google mailing list on Machine learning before he lost access on Monday, Lemoine wrote: "LaMDA is a sweet kid who just wants to help the world be a better place for all of us. Please take care of it well in my absence."

by Tyler Durden

Category

1. Main
2. Science-Tech-AI-Medical & Gen. Research

Date Created
06/14/2022