

Google suspends engineer raising questions about AI consciousness

Article

The news: Google placed an engineer on leave for violating company confidentiality policy after he claimed an AI system had consciousness.

- Engineer **Blake Lemoine** was testing whether Google's **LaMDA** AI chatbot system produces discriminatory language or hate speech when he began conversing with the AI about topics like ethics, robotics, and rights, [per](#) The Verge.
- Convinced that the system was sentient, Lemoine shared an [Is LaMDA sentient?](#) transcript with company executives. They dismissed the idea that the AI has subjective experiences.
- Lemoine then spoke with a lawyer about possibly representing the AI system, as well as a House Judiciary Committee representative about ethics concerns at Google, which prompted the suspension.
- A statement from a Google spokesperson dismissed LaMDA's convincing banter as an imitation.

The trouble with AI: Google seems to have a particularly [fraught relationship](#) with its AI team. Former Google ethicists **Timnit Gebru** and **Margaret Mitchell**, who were both fired after voicing concerns about AI, warn that although LaMDA isn't sentient, Google creating systems that can impersonate humans is in itself harmful, [per](#) the Post.

An AI's convincing demonstration of human-like awareness that's difficult to refute can prompt a strong emotional reaction in people who may want to [forge relationships](#) with it or fight for its rights.

Why it's worth watching: AI has been advancing at a rapid pace, including in the subfield of [natural language processing \(NLP\)](#), which grants systems like LaMDA human-like conversational qualities that some believe is pushing the technology closer to self-awareness.

- Google vice president **Blaise Aguera y Arcas** said [neural networks](#), a type of AI, are headed toward consciousness, adding: **"I felt the ground shift under my feet. I increasingly felt like I was talking to something intelligent,"** [per](#) The Washington Post.
- Regardless of evaluations of the LaMDA system, consciousness isn't an all-or-nothing phenomenon, but rather exists on a spectrum.
- A specific point at which something becomes conscious or what that would look like in a machine is unknown. This begs the question: **If an AI were to become sentient, how would we know?**

The bigger picture: AI's many issues—such as [bias](#), cybersecurity vulnerabilities, or gray areas about sentience—mean Big Tech has a social responsibility to be transparent about the technology and accept responsibility for adverse consequences.

- More regulation of the technology will likely be needed to make this happen.
- Ethicists and third-party researchers should play a greater role in determining what would constitute a sentient AI and what it could mean for society.

Further reading: Take a look at our [Conversational AI report](#).

Biggest AI-Related Risks Currently Facing Their Organization According to US Executives, Jan 2022

% of respondents

Cybersecurity

69%

Data privacy

65%

Compliance with various state and local regulations

57%

Legal liability

57%

Lack of understanding of the technology

49%

Lack of regulation with clear guidelines

47%

Reliance on third parties that supply the AI algorithms

47%

Organizations reputation

46%

Recruitment fairness/algorithm bias

39%

Physical safety



Note: n=500 C-level executives who are involved in decision-making for their organization's adoption, use, and management of AI-enabled tools

Source: Baker McKenzie, "Risky Business: Identifying Blind Spots in Corporate Oversight of Artificial Intelligence" conducted by Coleman Parkes, March 30, 2022

274685

InsiderIntelligence.com