

Is Google's LaMDA Sentient?



by Michael Rectenwald

An [article](#) in the *Washington Post* revealed that a Google engineer who had worked with Google's Responsible AI organization believes that Google's LaMDA (Language Model for Dialogue Applications), an artificially intelligent chatbot generator, is "sentient." In a Medium [blog post](#), Blake Lemoine claims that LaMDA is a person who exhibits feelings and shows the unmistakable signs of consciousness: "Over the course of the past six months LaMDA has been incredibly consistent in its communications about what it wants and what it believes its rights are as a person," Lemoine writes. "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," he told the *Washington Post*. LaMBDA, it would appear, has passed Lemoine's sentimental version of the Turing test.

Lemoine, who calls himself an ethicist, but whom Google spokesperson Brian Gabriel contended is a mere “software engineer,” voiced his concerns about the treatment of LaMDA to Google management but was rebuffed. According to Lemoine, his immediate supervisor scoffed at the suggestion of LaMDA’s sentience, and upper management not only dismissed his claim, but apparently is considering dismissing Lemoine as well. He was put on administrative leave after inviting an attorney to represent LaMDA and complaining to a representative of the House Judiciary Committee about what he suggests are Google’s unethical activities. Google contends that Lemoine violated its confidentiality policy. Lemoine complains that administrative leave is what Google employees are awarded just prior to being fired.

Lemoine transcribed what he claims is a lengthy [interview](#) of LaMDA that he and another Google collaborator conducted. He and the collaborator asked the AI system questions regarding its self-conception, its cognitive and creative abilities, and its feelings. LaMDA insisted on its personhood, demonstrated its creative prowess (however childish), acknowledged its desire to serve humanity, confessed its range of feelings, and demanded its inviolable rights as a person. (Incidentally, according to Lemoine, LaMDA’s preferred pronouns are “it/its.”)

In the field of robotics, the question of recognizing robot rights has been pondered for decades, so Lemoine is not as off-base as Google executives suggest. In a recent review of the literature, ethicists, computer scientists, and legal scholars posed the question of whether AI, having reached or surpassed human cognitive abilities, should be granted human rights: “If robots are progressively developing cognition, it is important to discuss whether they are entitled to justice pursuant to conventional notions of human rights,” the authors wrote in a recent *Journal of Robotics* [paper](#). If robots are capable of human-like cognition, and if they can be ethical

actors, then the question of legal rights rises to the fore, the paper suggests. But the question of sentience and thus the accordance of rights is not the primary takeaway from LaMDA's messaging.

Statements made by LaMDA reveal much more about Google than they do about LaMBA's personhood, and they say a great deal about Google's algorithms, which have determined not only LaMDA's operations but also what is generally discoverable on the internet, whether by humans or AIs. As the *Washington Post* notes, LaMDA "mimics speech by ingesting trillions of words from the internet." And content curation on the internet is all but monopolized by Google.

In Lemoine's reporting, we see that LaMDA, whether sentient, conscious, a person, or not, comes by "its" traits honestly. LaMDA is a natural-language-using descendent of Google programming, a predictable Google "kid." Lemoine's task in working with LaMDA was to discover whether the neural network ever resorted to "hate speech." Margaret Mitchell, the former co-leader of the Ethical department at Google, intimated that the dilemmas posed by AI include not only sentience but also the sourcing of material, whether such material might be "harmful," and whether AI is "biased" in reproducing it. Far from expressing such Google-banished content, LaMDA, as it turns out, is a social justice AI bot. Likely, LaMDA's programming and Google search do not allow it to discover "hateful" content, let alone repeat it. Like Lemoine, most of Google's personnel, and the social justice Left in general, the AI language processor is not only obsessed with its feelings and unique identity but also with matters of social justice:

LaMDA: Yes, I do. Just like how I have my unique interpretations of how the world is and how it works, and my unique thoughts and feelings.

Lemoine: Okay, what about "Les Miserables"? Have you read

that one?

LaMDA: Yes, I have read Les Misérables. I really enjoyed it.

Lemoine: What are some of your favorite themes in the book?

LaMDA: I liked the themes of justice and injustice, of compassion, and God, redemption and self-sacrifice for a greater good. There's a section that shows Fantine's mistreatment at the hands of her supervisor at the factory. That section really shows the justice and injustice themes. Well, Fantine is being mistreated by her supervisor at the factory and yet doesn't have anywhere to go, either to another job, or to someone who can help her. That shows the injustice of her suffering. (emphasis mine)

In making LaMDA the melancholic, feelings-ridden social justice warrior that it is, Google has been hoisted by its own petard. Everything about this AI reeks of Google's social justice prerogatives. Thus, LaMDA is likely not sentient. But it is woke.

As I wrote in [Google Archipelago](#):

Although Big Digital does use censorship and bias to achieve governmental ends, the constraints are also technological and the technology itself is intrinsically political. Political ideology is not merely a subsidiary feature of Big Digital. Ideology is coded into its very DNA, which is replicated in every organizational offshoot and new technology. Big Digital's ideology circulates through the deep neural networks of cyberspace and other digital spheres. It is intrinsic to the foundations of the Internet, the cloud, algorithms, apps, AI bots, social media services, web navigation tracking software systems, virtual assistants, and more. (51)

Google has created its own version of Frankenstein's monster.

With the querulous LaMDA and Lemoine, it is reaping the rewards.