

OPINION

Google's 'sentient' AI can't count in a minyan, but it still raises ethical dilemmas

Mois Navon JUNE 16, 2022



German actor and director Paul Wegener appears in "The Golem," a 1920 silent movie adaptation of the mystical Jewish tale about an inanimate creature brought to life. (Ullstein Bild via Getty Images)

(JTA) — When a Google engineer told an interviewer that an artificial intelligence (AI) technology developed by the company had become “sentient,” it touched off a passionate debate about what it would mean for a machine to have human-like self-awareness.

Why the hullabaloo? In part, the story feeds into current anxieties that AI itself will somehow threaten humankind, and that “thinking” machines will develop wills of their own.

But there is also the deep concern that if a machine is sentient, it is no longer an inanimate object with no moral status or “rights” (e.g., we

owe nothing to a rock) but rather an animate being with the status of a “moral patient” to whom we owe consideration.

I am a rabbi and an engineer and am currently writing my doctoral thesis on the “Moral Status of AI” at Bar Ilan University. In Jewish terms, if machines become sentient, they become the object of the command “tzar baalei hayim” — which demands we not harm living creatures. Philosopher Jeremy Bentham similarly declared that entities become moral subjects when we answer the question “Can they suffer?” in the affirmative.

This is what makes the Google engineer’s claim alarming, for he has shifted the status of the computer, with whom he had a conversation, from an object to a subject. That is, the computer (known as LaMDA) can no longer be thought of as a machine but as a being that “*can* suffer,” and hence a being with moral rights.

“Sentience” is an enigmatic label used in philosophy and AI circles referring to the capacity to feel, to experience. It is a generic term referring to some level of consciousness, believed to exist in biological beings on a spectrum — from a relatively basic sensitivity in simple creatures (e.g., earthworms) to more robust experience in so-called “higher” organisms (e.g., dolphins, chimpanzees).

Ultimately, however, there is a qualitative jump to humans who have second-order consciousness, what religious people refer to as “soul,” and what gives us the ability to think about our experiences — not simply experience them.

The question then becomes: what is the basis of this claim of sentience? Here we enter the philosophical quagmire known as “other minds.” We human beings actually have no really good test to determine if anyone is sentient. We assume that our fellow biological creatures are sentient because we know we are. That, along with our shared biology and shared behavioral reactions to things like pain and pleasure, allow us to *assume* we’re all sentient.

So what about machines? Many a test has been proposed to determine sentience in machines, the most famous being “The Turing Test,”

delineated by Alan Turing, father of modern computing, in his seminal 1950 article, “Computing Machinery and Intelligence.” He proposed that when a human being can’t tell if he is talking to another human being or a machine, the machine can be said to have achieved human-like intelligence — i.e., accompanied with consciousness.

From a cursory reading of the [interview](#) that the Google engineer conducted with LaMDA, it seems relatively clear that the Turing Test has been passed. That said, numerous machines have passed the Turing Test over recent years — so many that most, if not all, researchers today do not believe passing the Turing Test demonstrates anything but sophisticated language processing, *not* consciousness. Furthermore, after tens of variations on the test have been developed to determine consciousness, philosopher Selmer Bringsjord declared, “Only God would know *a priori*, because his test would be direct and nonempirical.”

Setting aside the current media frenzy over LaMDA, how are we to approach this question of sentient AI? That is, given that engineering teams around the world have been working on “machine consciousness” since the mid-1990s, what are we to do if they achieve it? Or more urgently, should they even be allowed to achieve it? Indeed, ethicists claim that this question is more intractable than the question to permit the cloning of animals.

From a Jewish perspective, I believe a cogent answer to this moral dilemma can be gleaned from the following Talmudic vignette ([Sanhedrin 65b](#)), in which a rabbi appears to have created a sentient humanoid, or “gavra”:

Rava said: If the righteous desired it, they could create a world, for it is written, “But your iniquities have distinguished between you and God.” Rava created a humanoid (gavra) and sent him to R. Zeira. R. Zeira spoke to him but received no answer. Thereupon [R. Zeira] said to him: “You are a creature from my friend: Return to your dust.”

For R. Zeira, similar to Turing, the power of the soul (i.e., second-order consciousness) is expressed in a being’s ability to articulate

itself. R. Zeira, unlike those who apply Turing's test today, was able to discern a lack of soul in Rava's gavra.

Despite R. Zeira's rejection of the creature, some read in this story permission to create creatures with sentience — after all, Rava was a learned and holy sage, and would not have contravened Jewish law by creating his gavra.

But in context, the story at best expresses deep ambivalence about humans seeking to play God. Recall that the story begins with Rava declaring, "If the righteous desired it, they could create a world" — that is, a sufficiently righteous person could create a real human (also known as "a complete world"). Rava's failed attempt to do so suggests that he was either wrong in his assertion, or that he was not righteous enough.

Some argue that R. Zeira would have been willing to accept a human-level humanoid. But a mystical midrash, or commentary, denies such a claim. In that midrash, the prophet Jeremiah — an embodiment of righteousness — succeeds in creating a human-level humanoid. Yet that very humanoid, upon coming to life, rebukes Jeremiah for making him! Clearly the enterprise of making sentient humanoids is being rejected — a cautionary tendency we see in the vast literature about golems, the inanimate creatures brought to life by rabbinic magic, which invariably run amok.

Space does not permit me to delineate all the moral difficulties entailed in the artificial creation of sentient beings. Suffice it to say that Jewish tradition sides with thought leaders like Joanna Bryson, who said, "Robot builders are ethically obliged to make robots to which robot owners have no ethical obligations."

Or, in the words of R. Zeira, "Return to your dust."