If your phone had feelings, would you treat it differently?

Artificial intelligence could achieve sentience in 10 years. We should prepare for these systems to have their own subjective experiences

BY BRIAN KATEMAN

Whether it's the virtual assistants in our phones, the chatbots providing customer service for banks and clothing stores, or tools like ChatGPT and Claude making workloads a little lighter, artificial intelligence has quickly become part of our daily lives. We tend to assume that our robots are nothing but machinery — that they have no spontaneous or original thought, and definitely no feelings. It seems almost ludicrous to imagine otherwise. But lately, that's exactly what experts on AI are asking us to do.

Eleos AI, a nonprofit organization dedicated to exploring the possibilities of AI sentience — or the capacity to feel — and well-being, released a report in October in partnership with the NYU Center for Mind, Ethics and Policy, titled "Taking AI Welfare Seriously." In it, they assert that AI achieving sentience is something that really could happen in the not-too-distant future — about a decade from now. Therefore, they argue, we have a moral imperative to begin thinking seriously about these entities' well-being.

I agree with them. It's clear to me from the report that unlike a rock or river, AI systems will soon have certain features that make consciousness within them more probable — capacities such as perception, attention, learning, memory and planning.

That said, I also understand the skepticism. The idea of any nonorganic entity having its own subjective experience is laughable to many because consciousness is thought to be exclusive to carbon-based beings. But as the authors of the report point out, this is more of a belief than a demonstrable fact — merely one kind of theory of consciousness. Some theories imply that biological materials are required, others imply that they are not, and we currently have no way to know for sure which is correct. The reality is that the emergence of consciousness might depend on the structure and organization of a system, rather than on its specific chemical composition.

The core concept at hand in conversations about AI sentience is a classic one in the field of ethical philosophy: the idea of the "moral circle," describing the kinds of beings to which we give ethical consideration. The idea has been used to describe whom and what a person or society cares about, or, at least, whom they ought to care about. Historically, only humans were included, but over time many societies have brought some animals into the circle, particularly pets like dogs and cats. However, many other animals, such as those raised in industrial agriculture like chickens, pigs and cows, are still largely left out.

Many philosophers and organizations devoted to the study of AI consciousness come from the field of animal studies, and they're essentially arguing to extend the line of thought to nonorganic entities, including computer programs. If it's a realistic possibility that something can become a someone who suffers, it would be morally negligent for us to not give some serious consideration to how we can avoid inflicting that pain.

An expanding moral circle demands ethical consistency and makes it difficult to carve out exceptions based on cultural or personal biases. And right now, it's only those biases that allow us to ignore the possibility of sentient AI. If we are morally consistent, and we care about minimizing suffering, that care has to extend to many other beings — including insects, microbes and maybe something in our future computers.

Even if there's just a tiny chance that AI could develop sentience, there are so many of these "digital animals" out there that the implications are huge. If every phone, laptop, virtual assistant, etc. someday has its own subjective experience, there could be trillions of entities that are subjected to pain at the hands of humans, all while many of

us function under the assumption that it's not even possible in the first place. It wouldn't be the first time people have dealt with ethical quandaries by telling themselves and others that the victims of their practices simply can't experience things as deeply as you or I.

For all these reasons, leaders at tech companies like OpenAI and Google should start taking the possible welfare of their creations seriously. This could mean hiring an AI welfare researcher and developing frameworks for estimating the probability of sentience in their creations. If AI systems evolve and have some level of consciousness, research will determine whether their needs and priorities are similar to or different from those of humans and animals, and that will inform what our approaches to their protection should look like.

Maybe a point will come in the future where we have widely accepted evidence that robots can indeed think and feel. But if we wait to even entertain the idea, imagine all the suffering that will have happened in the meantime. Right now, with AI at a promising but still fairly nascent stage, we have the chance to prevent potential ethical issues before they get further downstream. Let's take this opportunity to build a relationship with technology that we won't come to regret. Just in case.

Brian Kateman is co-founder of the Reducetarian Foundation, a nonprofit organization dedicated to reducing societal consumption of animal products. His latest book and documentary is "Meat Me Halfway."