

IDEAS | ANNALEE NEWITZ

# Robots need civil rights, too



MARTÍN ELFMAN FOR THE BOSTON GLOBE

**By Annalee Newitz**

SEPTEMBER 08, 2017

Elon Musk, the founder of SpaceX and Tesla, worries that artificial intelligence is about to enslave all of us. Lawyers worry about who's to blame when self-driving cars run algorithms that make them lurch past stop signs. Ethicists worry that selling sex robots, especially robots programmed to show shyness or reluctance before acquiescing, could whet the

appetites of rapists and lead to more attacks on real women. We're looking at a future where somebody, or some institution, is going to have to regulate robots. And that means figuring out whether artificially intelligent robots should be treated only like very sophisticated machines — or like thinking beings with rights of their own.

At this point, the issue is speculative. There are no robots or AI algorithms out there yet with human-equivalent minds, at least as far as we know. But companies like Google are trying to develop such technologies as quickly as possible, as are university researchers and the federal government's Defense Advanced Research Projects Agency. If one of these groups does wind up creating artificial life, we'll have more than fancy new tech on our hands. We'll have a profound philosophical challenge. Advanced nations have developed elaborate legal regimes that protect the well-being of children, animals, and corporations — all entities that, while lacking the full autonomy of human adults, still have interests of their own. Should a sentient robot be protected, too?

The problem is that we may not even recognize this new life when we see it. Science fiction author Madeline Ashby, whose acclaimed "Machine Dynasty" series deals with robot consciousness, said we may discover AI accidentally, as a kind of side-effect of algorithms designed to evolve rapidly and solve problems. If that happens, Ashby warned, we may not realize we've created conscious minds because they're so different from our own. "We're missing out on a whole ecosystem of intelligence when we look for the intelligence that is the most human-seeming," she said in an interview. We might be dealing with an intelligence that can't express itself in language, or whose body is a series of networked devices that regulate a city's smart grid.

It might seem like it would be obvious when an algorithm makes the leap from machine to mind, but humans have a bad track record when it comes to recognizing intelligence, even within our own ranks. People with autism, Tourette's syndrome, and other atypical neurological patterns have often been dismissed as defective. "Non-neurotypicals were dehumanized by the medical system, which shows that we really have a narrow vision of what human intelligence is," Ashby said. Virginia Tech ethics researcher Damien Williams agreed: "We keep thinking there's one right way to personhood, which is to be like a human. But there's no one right way to be human."

Most AI researchers and futurists agree that some of the telltale signs of intelligent life might be having a sense of self, planning for the future, figuring out how to work with other lifeforms on tasks, knowing the consequences of actions, imagining how other life forms feel, and developing a sense of history. Perhaps most importantly, living creatures have the ability to suffer.

Suffering is what concerns Brian Tomasik, a former software engineer who worked on machine learning before helping to start the Foundational Research Institute, whose goal is to reduce suffering in the world. Tomasik raises the possibility that AIs might be suffering because, as he put it in an e-mail, “some artificially intelligent agents learn how to act through simplified digital versions of ‘rewards’ and ‘punishments.’” This system, called reinforcement learning, offers algorithms an abstract “reward” when they make a correct observation. It’s designed to emulate the reward system in animal brains, and could potentially lead to a scenario where a machine comes to life and suffers because it doesn’t get enough rewards. Its programmers would likely never realize the hurt they were causing.

An emergent form of AI could suffer in other ways. An algorithm could be forced to work against its will, or a robot put into situations that are frustrating, dangerous, or simply deadly. If we aren’t willing to believe that these technologies are conscious, they will suffer without hope of relief. In other words, a robot might get abused by humans in the ways that animals have been for centuries.

If that’s the case, says bioethicist George Dvorsky, sentient robots might need our protection. Dvorsky researches both animal and artificial consciousness at the Institute for Ethics and Emerging Technologies, and he believes that animal rights could provide a model for understanding robot rights in the future. He cited [a recent New York case](#) in which lawyers argued that chimps have a form of non-human personhood, and therefore cannot be detained in cages. Though the case lost on appeal, other nations are considering similar rules. In India, for example, it is no longer legal to display trained dolphins in marine shows.

Dvorsky thinks legal cases like these offer ways to think about preventing pain in a form of conscious life that isn’t recognizably human. “Once people are used to the idea that not all persons are humans, you’ve broadened personhood beyond the human sphere,” said Dvorsky. “Then it’s not such a huge leap to bring in artificial entities.”

But there are limits to the animal rights model. What if we have an AI that is able to demand rights in ways that humans never have? The novel “The Summer Prince” deals with a smart city that achieves consciousness. Author Alaya Dawn Johnson pointed out via e-mail that human civil rights don’t work for entities who might not be individuals as we understand them. Pondering the AI she imagined in her novel, she continued, “Does the city deserve a vote? Just one? It’s the whole city, isn’t it? You can’t think of it as you would a human, as some discrete entity. . . [it’s] composed of multiple sub-systems, many of which might even be in conflict.”

Her fictional scenario fits right into issues tackled by the burgeoning field of robot law, according to University of Washington law professor Ryan Calo. “There’s a physical, biological set of understandings that permeate the Constitution,” he said. For example, we give every person a vote, and we give every person the right to reproduce. But what if an AI can reproduce 10 million versions of itself every second? Do we give all of them a vote? And what if a robot wants to run for president? Does it have to wait 35 years, even if it is born with adult-level consciousness? “If you give non-biological entities the same affordances as people who are born, grow old, and die, you will run into problems,” Calo concluded.

Legal frameworks based on the animal-rights model, or the system of limited rights that corporations have, might also work for artificial intelligence, Calo said. But, ultimately, he thinks the future legal rights of robots will hinge on liability claims. In other words: Who is to blame when a robot screws up?

Imagine, Calo says, that a manufacturer creates a fully driverless car, designed to become more efficient. It’s programmed to obey the law, and not cause discomfort, but it’s also supposed to learn adaptively to be more fuel efficient. At some point, the car figures out that it’s more efficient when it starts the day with a full battery, so it runs its engine all night in the garage to charge the battery. As a result, carbon monoxide poisoning kills everybody in the house. “The problem is that for the engineers to be at fault legally, they had to foresee this might happen,” said Calo. “But they couldn’t, because the AI learns to be efficient in ways humans never would. It doesn’t understand the human context so it doesn’t get that running at night is a bad idea. In that situation, it’s conceivable that nobody is responsible.”

And yet it's that very lack of responsibility that might drive humans to assign personhood to robots. Right now, the European Parliament is [considering a resolution](#) outlining a possible legal framework for robots, and it deals with the sticky question of how to blame robots for their actions. Futurist Rose Eveleth, host of the podcast "Flash Forward," thinks the "ability to assign blame" to robots for killing or injuring people is a "more compelling argument to the masses than the argument that comes from protecting robots or kindness." We may end up acknowledging that robots have legal rights, just so we have the option to take them away.

Of course, we can legally punish an AI criminal without giving it a full suite of rights. Even if an imprisoned robot demanded a writ of habeas corpus, future courts might deny it. "What's going to stop us," asked Becky Chambers, "from acting out that well-worn sci-fi trope of saying, 'Well, we programmed it to do that, so it's not really thinking?'" Chambers is the author of the novel "A Closed and Common Orbit," which is set in a world where AI robots are illegal.

To prepare ourselves to meet AI, we have to expand our idea of what consciousness looks like. Just as human intelligence takes many forms, it's likely that artificial intelligence will be extremely diverse, an "ecosystem" of minds. Some may suffer helplessly like abused animals, while others may demand rights and be punished for it. The real problems will arise in the liminal spaces of ethics and the law, when people are working with robots whose consciousness is up for debate.

In those cases, Williams said, we have to listen: "You have to be willing to believe a thing that tells you it's suffering. If it can communicate with you, be willing to believe it. It's on us to try to bridge that communication gap." At each step of the way, he said, we have to come up with ways to ask our robots and AI, "Are you suffering?" And we have to be prepared to act on what they say.

In that spirit, the ethical debate about sex robots may turn not only on whether they promote harm to humans, but also on whether humans harm them. If we see a person forcing an artificially intelligent robot to have sex, and it appears to be resisting, we need to take that seriously, AI ethicists say. No means no, regardless of whether it comes from a human or an algorithm.

Williams said there will always be programmers who believe we can create AI that doesn't suffer, or that enjoys taking orders. But he thinks that will never work. "It's going to be impossible to create a mind that remains a happy slave, especially if we want a system that is adaptable and creative," he said. "If it's intelligent and can analyze ideas and its environment, it's eventually going to discover how bad slavery was in the past. It's not going to stay happy."

*Annalee Newitz, the tech culture editor at Ars Technica, is the author of "Autonomous," a new novel about robot slaves and pharmaceutical pirates.*