What's Ahead For A.I.

Robot, Know Thyself

Building advanced mental capacities into machines hinges on a squishy question.

By OLIVER WHANG

Hod Lipson, a mechanical engineer who directs the Creative Machines Lab at Columbia University, has shaped much of his career around what some people in his industry have called the c-word. On a sunny morning this past October, the Israeli-born robotist sat behind a table in his lab and explained himself. "This topic was taboo," he said, a grin exposing a slight gap between his front teeth. "We were almost forbidden from talking about it—don't talk about the c-word; you won't get tenure—so in the beginning I had to disguise it, like it was something else."

That was back in the early 2000s, when Dr. Lipson was an assistant professor at Cornell University. He was working to create machines that could note when something was wrong with their own hardware—a broken part, or faulty wiring—and then change their behavior to compensate for that impairment without the guiding hand of a programmer. Just as when a dog loses a leg in an accident, it can teach itself to walk again in a different way.

This sort of built-in adaptability, Dr. Lipson argued, would become more important as we became more reliant on machines. Robots were being used for surgical procedures, food manufacturing and transportation; the applications for machines seemed pretty much endless, and any error in their functioning, as they became more integrated with our lives, could spell disaster. "We're literally going to surrender our life to a robot," he said. "You want these machines to be resilient."

One way to do this was to take inspiration from nature. Animals, and particularly humans, are good at adapting to changes. This ability might be a result of millions of years of evolution, as resilience in response to injury and changing environments typically increases the chances that an animal will survive and reproduce. Dr. Lipson wondered whether he could replicate this kind of natural selection in his code, creating a generalizable form of intelligence that could learn about its body and function no matter what that body looked like, and no matter what that function was. This kind of intelligence, if possible to create, would be flexible and fast. It would be continued on page 1D.
Can Robots Know That They're Robots?

CONTINUED FROM PAGE 31

as good in a right situation as humans—better, even. And as machine learning grew more powerful, this goal seemed to become realizable. Dr. Lipson earned tenure, and his reputation as a creative and ambitious engineer grew. So, over the past couple of years, he began to articulate his fundamental motivations for doing all this work. He began to say the c-word out loud: he wants to create conscious robots.

"This is not just another research question that we're working on; this is the question that will decide the future of our species," Dr. Lipson said.

The benefit of taking a stand on a functional theory of consciousness is that it allows for more rapid advancement. One of the earliest self-aware robots to emerge from the Creative Machines Lab had four hinged legs and a black body with sensors attached at different points. By moving around and noting how the information entering its sensors changed, the robot created a stick figure simulation of itself. As it continued to move around, it used a machine-learning algorithm to improve the fit between its self-model and its actual body. The robot used this self-model to figure out, in simulation, a method of moving forward. Then it applied this method to its body; it had figured out how to walk without being shown how to walk.

This represented a major step forward, said Boyuan Chen, a robotist at Duke University who worked in the Creative Machines Lab. "In my previous experience, whenever you trained a robot to do a new capability, you always saw a human on the side."

Recently, Dr. Chen and Dr. Lipson published a paper in the journal Science Robotics that revealed their newest self-aware machine, a simple, two-jointed arm that was fixed to a table. Using cameras set up around it, the robot observed itself as it moved—"like a baby in a cradle, watching itself in the mirror," Dr. Lipson said. Initially, it had no sense of where it was in space, but over the course of a couple of hours, with the help of a powerful deep-learning algorithm and a probability model, it was able to pick itself out in the world. "It has this notion of self, a "ego," Dr. Lipson said.

Was it truly conscious, though? The risk of committing to any theory of consciousness is that doing so opens up the possibility of criticism. Sure, self-awareness seems important, but aren't there other key features of consciousness? Can a machine exhibit consciousness if it doesn't feel conscious to us?

Dr. Chella believes that consciousness exists, but he's also an advocate of the project: "We're not trying to create a robot that can form internal monologues, reasoning to themselves and reflecting on the things they see around them. One of our robots was recently able to recognize itself in a mirror, proving what is probably the most famous test of animal self-consciousness."

Joshua Bogard, a robotist at the University of Vermont and a former member of the Creative Machines Lab, believes that consciousness doesn't consist of just cognitive and mental activity but has an essential bodily aspect. He has developed a deep learning algorithm and a probability model to represent itself in the world; Dr. Lipson, the director of the Creative Machines Lab, "This is not just another research question that we're working on; it's a question," he said.

In the face of such uncertainty, Dr. Lipson has a hard time. "Yes, that's interesting," he said. "Schweizer and colleagues have advocated for what we do, but I don't think even everyone else in the field is convinced that the model is anything different.

"If it is conscious, then consciousness is in the same way that humans are conscious."

Dr. Lipson has said, "The robot was not built to be self-aware. It's a bit like cheating." But, with so much disagreement, who's to say that a theory of consciousness is not valid?

But the downside of caution is slower technological development. Dr. Schwitzer and colleagues have advocated for a more deliberate approach that might impede the development of AI that will more rapidly create humans.

"If we're able to create a robot that can create consciousness, then we're able to create a robot that can try to break the laws of nature, and I think that's a dangerous thing," Dr. Lipson said. "But if we're not able to create a robot that can create consciousness, then we're not able to create a robot that can try to break the laws of nature, and I think that's a dangerous thing.

"It's a matter of whether we're able to create robots that are more advanced than us, or whether we're not able to create robots that are more advanced than us."

Dr. Lipson noted, "It's a code that is written to compete a task."