PHYS ORG

Teaching robots right from wrong

May 9 2014

Researchers from Tufts University, Brown University, and Rensselaer Polytechnic Institute are teaming with the U.S. Navy to explore technology that would pave the way for developing robots capable of making moral decisions.

In a project funded by the Office of Naval Research and coordinated under the Multidisciplinary University Research Initiative, scientists will explore the challenges of infusing autonomous robots with a sense for right, wrong, and the consequences of both.

"Moral competence can be roughly thought about as the ability to learn, reason with, act upon, and talk about the laws and societal conventions on which humans tend to agree," says principal investigator Matthias Scheutz, professor of computer science at Tufts School of Engineering and director of the Human-Robot Interaction Laboratory (HRI Lab) at Tufts. "The question is whether machines—or any other artificial system, for that matter—can emulate and exercise these abilities."

One scenario is a battlefield, he says. A robot medic responsible for helping wounded soldiers is ordered to transport urgently needed medication to a nearby field hospital. En route, it encounters a Marine with a fractured leg. Should the robot abort the mission to assist the injured? Will it?

If the machine stops, a new set of questions arises. The robot assesses the soldier's physical state and determines that unless it applies traction, internal bleeding in the soldier's thigh could prove fatal. However, applying traction will cause intense pain. Is the robot morally permitted to cause the soldier pain, even if it's for the soldier's well-being?

The ONR-funded project will first isolate essential elements of human moral competence through theoretical and empirical research. Based on the results, the team will develop formal frameworks for modeling human-level moral reasoning that can be verified. Next, it will implement corresponding mechanisms for moral competence in a computational architecture.

"Our lab will develop unique algorithms and computational mechanisms integrated into an existing and proven architecture for autonomous robots," says Scheutz. "The augmented architecture will be flexible enough to allow for a robot's dynamic override of planned actions based on moral reasoning."

Once architecture is established, researchers can begin to evaluate how machines perform in human-robot interaction experiments where robots face various dilemmas, make decisions, and explain their decisions in ways that are acceptable to humans.

Selmer Bringsjord, head of the Cognitive Science Department at RPI, and Naveen Govindarajulu, post-doctoral researcher working with him, are focused on how to engineer ethics into a robot so that moral logic is intrinsic to these artificial beings. Since the scientific community has yet to establish what constitutes morality in humans the challenge for Bringsjord and his team is severe.

In Bringsjord's approach, all robot decisions would automatically go through at least a preliminary, lightning-quick ethical check using simple logics inspired by today's most advanced artificially intelligent and question-answering computers. If that check reveals a need for deep, deliberate moral reasoning, such reasoning would be fired inside the robot, using newly invented logics tailor-made for the task.

"We're talking about robots designed to be autonomous; hence the main purpose of building them in the first place is that you don't have to tell them what to do," Bringsjord said. "When an unforeseen situation arises, a capacity for deeper, on-board reasoning must be in place, because no finite rule set created ahead of time by humans can anticipate every possible scenario."

Bertram Malle, from the Department of Cognitive, Linguistic and Psychological Services at Brown University, will perform some of the human research and human-robot interaction studies. "To design a morally competent robot that interacts with humans we need to first get clear on how moral competence functions in humans," he said. "There is a fair amount of scientific knowledge available, but there are still many unanswered questions. By answering some of these questions in the project, we can move closer to designing a robot that has moral competence."

The overall goal of the project, says Scheutz, "is to examine human moral competence and its components. If we can computationally model aspects of moral cognition in machines, we may be able to equip robots with the tools for better navigating real-world dilemmas."

Besides the experts from Tufts, Brown, and RPI, this team will include consultants from Georgetown University and Yale University in this multi-year effort.

The group brings together extensive research expertise in theoretical models of moral cognition and communication; experimental research on human reasoning; formal modeling of reasoning; design of computational architectures; and implementation in robotic systems.

Provided by Tufts University

Citation: Teaching robots right from wrong (2014, May 9) retrieved 11 July 2024 from https://phys.org/news/2014-05-robots-wrong.html