Want responsible robotics? Start with responsible humans

July 29 2009, by Pam Frost Gorder

(PhysOrg.com) -- When the legendary science fiction writer Isaac Asimov penned the "Three Laws of Responsible Robotics," he forever changed the way humans think about artificial intelligence, and inspired generations of engineers to take up robotics.

In the current issue of journal *IEEE Intelligent Systems,* two engineers propose alternative laws to rewrite our future with robots.

The future they foresee is at once safer, and more realistic.

"When you think about it, our cultural view of robots has always been anti-people, pro-robot," explained David Woods, professor of integrated systems engineering at Ohio State University. "The philosophy has been, 'sure, people make mistakes, but robots will be better -- a perfect version of ourselves.' We wanted to write three new laws to get people thinking about the human-robot relationship in more realistic, grounded ways."

Asimov's laws are iconic not only among engineers and science fiction enthusiasts, but the general public as well. The laws often serve as a starting point for discussions about the relationship between humans and robots.

But while evidence suggests that Asimov thought long and hard about his laws when he wrote them, Woods believes that the author did not intend for engineers to create robots that followed those laws to the letter.

"Go back to the original context of the stories," Woods said, referring to Asimov's I, [Robot](link) among others. "He's using the three laws as a literary device. The plot is driven by the gaps in the laws -- the situations in which the laws break down. For those laws to be meaningful, robots have to possess a degree of social intelligence and moral intelligence, and Asimov examines what would happen when that intelligence isn't there."

"His stories are so compelling because they focus on the gap between our aspirations about robots and our actual capabilities. And that's the irony, isn't it? When we envision our future with robots, we focus on our hopes and desires and aspirations about robots -- not reality."

In reality, engineers are still struggling to give robots basic vision and language skills. These efforts are hindered in part by our lack of understanding of how these skills are managed in the human brain. We are far from a time when humans may teach robots a moral code and responsibility.

Woods and his coauthor, Robin Murphy of Texas A&M University, composed three laws that put the responsibility back on humans.

Woods directs the Cognitive Systems Engineering Laboratory at Ohio State, and is an expert in automation safety. Murphy is the Raytheon Professor of Computer Science and Engineering at Texas A&M, and is an expert in both rescue robotics and human-robot interaction.

Together, they composed three laws that focus on the human organizations that develop and deploy robots. They looked for ways to ensure high safety standards.

Here are Asimov's original three laws:

- A robot may not injure a human being, or through inaction, allow a human being to come to harm.

- A robot must obey orders given to it by human beings, except where such orders would conflict with the First Law.

- A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

And here are the three new laws that Woods and Murphy propose:

- A human may not deploy a robot without the human-robot work system meeting the highest legal and professional standards of safety and ethics.

- A robot must respond to humans as appropriate for their roles.

- A robot must be endowed with sufficient situated autonomy to protect its own existence as long as such protection provides smooth transfer of control which does not conflict with the First and Second Laws.

The new first law assumes the reality that humans deploy robots. The second assumes that robots will have limited ability to understand human orders, and so they will be designed to respond to an appropriate set of orders from a limited number of humans.

The last law is the most complex, Woods said.

"Robots exist in an open world where you can't predict everything that's going to happen. The robot has to have some autonomy in order to act and react in a real situation. It needs to make decisions to protect itself,

but it also needs to transfer control to humans when appropriate. You don't want a robot to drive off a ledge, for instance -- unless a human needs the robot to drive off the ledge. When those situations happen, you need to have smooth transfer of control from the robot to the appropriate human," Woods said.

"The bottom line is, robots need to be responsive and resilient. They have to be able to protect themselves and also smoothly transfer control to humans when necessary."

Woods admits that one thing is missing from the new laws: the romance of Asimov's fiction -- the idea of a perfect, moral [robot](#) that sets engineers' hearts fluttering.

"Our laws are little more realistic, and therefore a little more boring," he laughed.

More information:

[doi.ieeecomputersociety.org/10.1109/MIS.2009.69](#)
[en.wikipedia.org/wiki/Three_Laws_of_Robotics](#)

Source: The Ohio State University ([news](#) : [web](#))