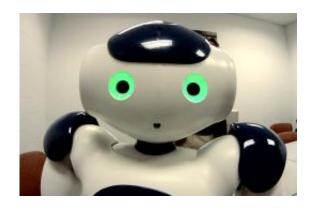# The ethical robot (w/ Video)

November 9 2010, By Christine Buckley and Bret Eckhardt

Susan and Michael Anderson have programmed a robot to behave ethically. Image by Bret Eckhardt

(PhysOrg.com) -- Philosopher Susan Anderson is teaching machines how to behave ethically.

Professor emerita Susan Anderson and her research partner, husband Michael Anderson of the University of Hartford, a University of Connecticut alumnus, at first seem to have little in common when it comes to their academic lives: she's a philosopher, he's a computer scientist.

But these two seemingly opposite fields have come together in the Andersons' collaborative work, in which the team works in a new field of research, called machine ethics, that's only about 10 years old.

Using their expertise in different areas, the Andersons have recently accomplished something that's never been done before: They've programmed a robot to behave ethically.

"There are machines out there that are already doing things that have ethical import, such as automatic cash withdrawal machines, and many others in the development stages, such as cars that can drive themselves and eldercare robots," says Susan, professor emerita of philosophy in the College of Liberal Arts and Sciences, who taught at UConn's Stamford campus. "Don't we want to make sure they behave ethically?"

The field of machine ethics combines artificial intelligence techniques with ethical theory, a branch of philosophy, to determine how to program machines to behave in an ethical manner. But there is currently no agreement, says Susan, as to which ethical principles should be programmed into machines.

In 1930, Scottish philosopher David Ross introduced a new approach to ethics, she says, called the *prima facie* duty approach, in which a person must balance many different obligations when deciding how to act in a moral way – obligations like being just, doing good, not causing harm, keeping one's promises, and showing gratitude.

The robot the Andersons use in their research has been programmed with an

ethical principle. Image by Bret Eckhardt

However, this approach was never developed far enough to instruct people how to weigh these different obligations with a satisfactory decision principle: one that would instruct them on how to behave when several of the *prima facie* duties pull in different directions.

"There isn't a decision principle within this theory, so it wasn't widely adopted," says Susan.

That's where the Andersons come in. By using information about specific ethical dilemmas supplied to them by ethicists, computers can effectively "learn" ethical principles in a process called machine learning.

The toddler-sized robot they have been using in their research, [called Nao](link), has been programmed with an ethical principle that was discovered by a computer. This learned principle allows their robot to determine how often to remind people to take their medicine and when to notify an overseer, such as a doctor, when they don't comply.

Reminding someone to take their medicine may seem relatively trivial, but the field of biomedical ethics has grown in relevance and importance since the 1960s. And robots are currently being designed to assist the elderly, so the Andersons' research has very practical implications.

Susan points out that there are several *prima facie* duties the robot must weigh in their scenario: enabling the patient to receive potential benefits from taking the medicine, preventing harm to the patient that might result from not taking the medication, and respecting the person's right of autonomy. These *prima facie* duties must be correctly balanced to

help the [robot](#) decide when to remind the patient to take medication and whether to leave the person alone or to inform a caregiver, such as a doctor, if the person has refused to take the medicine.

Philosopher Susan Anderson believes artificial intelligence has changed the field of ethics. Image by Bret Eckhardt

Michael says that although their research is in its early stages, it's important to think about ethics alongside developing artificial intelligence. Above all, he and Susan want to refute the science fiction portrayal of robots harming human beings.

"We should think about the things that robots could do for us if they had ethics inside them," Michael says. "We'd allow them to do more things for us, and we'd trust them more."

The Andersons organized the first international conference on machine ethics in 2005, and they have a book on machine ethics being published by Cambridge University Press. In the future, they envision computers continuing to engage in machine learning of ethics through dialogues with ethicists concerning real ethical dilemmas that machines might face in particular environments.

"Machines would effectively learn the ethically relevant features, *prima facie* duties, and ultimately the decision principles that should govern their behavior in those domains," says Susan.

Although this is a vision of the future of machine ethics research, Susan thinks that artificial intelligence has already changed her chosen field in major ways.

She thinks that working in machine ethics, which forces philosophers who are used to thinking abstractly to be more precise in applying ethics to specific, real-life cases, might actually advance the study of ethics.

And she believes that robots could be good for humanity: she believes that interacting with robots that have been programmed to behave ethically could even inspire humans to behave more ethically.

Provided by University of Connecticut