

# How to make ethical robots

12 March 2012, by Lisa Zyga



RI-MAN, a robot developed by researchers at RIKEN in Japan, was designed for human care. Image credit: RIKEN, Bio-Mimetic Control Research Center

(PhysOrg.com) -- In the future according to robotics researchers, robots will likely fight our wars, care for our elderly, babysit our children, and serve and entertain us in a wide variety of situations. But as robotic development continues to grow, one subfield of robotics research is lagging behind other areas: roboethics, or ensuring that robot behavior adheres to certain moral standards. In a new paper that provides a broad overview of ethical behavior in robots, researchers emphasize the importance of being proactive rather than reactive in this area.

The authors, Ronald Craig Arkin, Regents' Professor and Director of the Mobile [Robot](#) Laboratory at the Georgia Institute of Technology in Atlanta, Georgia, along with researchers Patrick Ulam and Alan R. Wagner, have published their overview of moral decision making in autonomous systems in a recent issue of the [Proceedings of the IEEE](#).

"Probably at the highest level, the most important message is that people need to start to think and talk about these issues, and some are more pressing than others," Arkin told *PhysOrg.com*. "More folks are becoming aware, and the very

young machine and robot ethics communities are beginning to grow. They are still in their infancy though, but a new generation of researchers should help provide additional momentum. Hopefully articles such as the one we wrote will help focus attention on that."

The big question, according to the researchers, is how we can ensure that future robotic technology preserves our humanity and our societies' values. They explain that, while there is no simple answer, a few techniques could be useful for enforcing ethical behavior in robots.

One method involves an "ethical governor," a name inspired by the mechanical governor for the steam engine, which ensured that the powerful engines behaved safely and within predefined bounds of performance. Similarly, an ethical governor would ensure that robot behavior would stay within predefined ethical bounds. For example, for autonomous military robots, these bounds would include principles derived from the Geneva Conventions and other rules of engagement that humans use. Civilian robots would have different sets of bounds specific to their purposes.

Since it's not enough just to know what's forbidden, the researchers say that autonomous robots must also need emotions to motivate behavior modification. One of the most important emotions for robots to have would be guilt, which a robot would "feel" or produce whenever it violates its ethical constraints imposed by the governor, or when criticized by a human. Philosophers and psychologists consider guilt as a critical motivator of moral behavior, as it leads to behavior modifications based on the consequences of previous actions. The researchers here propose that, when a robot's guilt value exceeds specified thresholds, the robot's abilities may be temporarily restricted (for example, military robots might not have access to certain weapons).

Though it may seem surprising at first, the researchers suggest that robots should also have

the ability to deceive people – for appropriate reasons and in appropriate ways – in order to be truly ethical. They note that, in the animal world, deception indicates social intelligence and can have benefits under the right circumstances. For instance, search-and-rescue robots may need to deceive in order to calm or gain cooperation from a panicking victim. Robots that care for Alzheimer's patients may need to deceive in order to administer treatment. In such situations, the use of deception is morally warranted, although teaching robots to act deceitfully and appropriately will be challenging.

The final point that the researchers touch on in their overview is ensuring that robots – especially those that care for children and the elderly – respect human dignity, including human autonomy, privacy, identity, and other basic human rights. The researchers note that this issue has been largely overlooked in previous research on robot ethics, which mostly focuses on physical safety. Ensuring that robots respect human dignity will likely require interdisciplinary input.

The researchers predict that enforcing ethical behavior in robots will face challenges in many different areas.

“In some cases it's perception, such as discrimination of combatant or non-combatant in the battlespace,” Arkin said. “In other cases, ethical reasoning will require a deeper understanding of human moral reasoning processes, and the difficulty in many domains of defining just what ethical behavior is. There are also cross-cultural differences which need to be accounted for.”

An unexpected benefit from developing an ethical advisor for robots is that the advising might assist humans when facing ethically challenging decisions, as well. Computerized ethical advising already exists for law and bioethics, and similar computational machinery might also enhance ethical behavior in human-human relationships.

“Perhaps if robots could act as role models in situations where humans have difficulty acting in accord with moral standards, this could positively reinforce ethical behavior in people, but that's an

unproven hypothesis,” Arkin said.

**More information:** Ronald Craig Arkin, et al. “Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception.” *Proceedings of the IEEE*. Vol. 100, No. 3, March 2012. [DOI: 10.1109/JPROC.2011.2173265](https://doi.org/10.1109/JPROC.2011.2173265)

*Copyright 2012 PhysOrg.com.*

*All rights reserved. This material may not be published, broadcast, rewritten or redistributed in whole or part without the express written permission of PhysOrg.com.*

APA citation: How to make ethical robots (2012, March 12) retrieved 29 November 2021 from <https://phys.org/news/2012-03-ethical-robots.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*