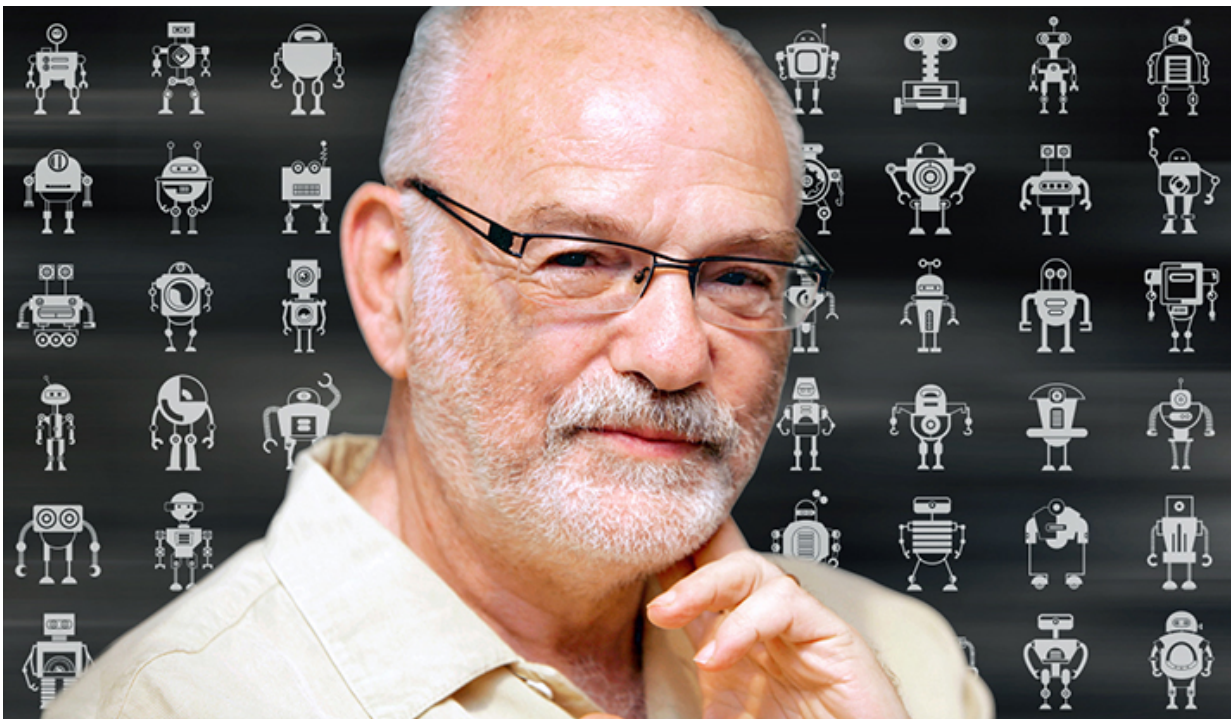


Shaping tomorrow's smart machines—Q&A with bioethicist Wendell Wallach

February 16 2016, by Jim Shelton



As intelligent machines continue to make their way into all sectors of society, a growing number of scientists, ethicists, policymakers, and business executives are converging on the idea that more thought must be given to underlying issues of machines and morality.

Already there are semi-autonomous technologies in use in military, manufacturing, [health care](#), and service industry settings. We have cars that avoid collisions and drones that may deliver packages someday. The question now is: What guiding principles should be employed, as smarter devices begin to take a more prominent role in security, [public safety](#), and other complex matters?

Wendell Wallach, a lecturer at Yale's Interdisciplinary Center for Bioethics and chair of the center's technology and ethics study group, has explored the issue for more than a decade. He is the author of "Moral Machines: Teaching Robots Right From Wrong" and "A Dangerous Master: How to Keep Technology from Slipping Beyond Our Control."

On Feb. 13, Wallach gave a presentation and press briefing on the topic of artificial intelligence (AI) at the American Association for the Advancement of Science's annual meeting, in Washington, D.C. During the press briefing, Wallach made three policy recommendations: directing 10% of AI/robotics research funding to studying and adapting to the societal impact of intelligent machines; creating an oversight and governance coordinating committee for AI/robotics; and a presidential order declaring that lethal autonomous weapons systems are in violation of international humanitarian law.

Wallach recently received a grant from inventor Elon Musk to develop a series of workshops with the Hastings Center, bringing together a variety of stakeholders around the topic of artificial intelligence and ethics.

YaleNews spoke with Wallach prior to his AAAS presentation.

Can we build machines that make moral decisions?

It is certainly possible to build machines that factor values and moral considerations into the choices and actions they take, particularly when

they function within a very limited context. However, making explicit moral judgments in many different contexts depends upon a clear and full understanding of the situation at hand. This will require that intelligent machines have consciousness and other capabilities that AI researchers do not yet know how to implement within computers and robots.

How far has this technology advanced, just in the past few years?

Recent breakthroughs using a technique called "deep learning" have demonstrated solutions to long-standing roadblocks in machine perception and learning. These problems had stymied AI researchers for decades. But it is still unclear how far the new techniques will take researchers toward the holy grail of artificial intelligence and what other breakthroughs or roadblocks exist on the near horizon. In other words, there has been a great leap forward. Present-day systems can perform remarkable tasks. But smart machines are still quite primitive as far as demonstrating the intelligence and adaptive capabilities that wise, caring, and creative humans possess.

Which emerging technologies interest you the most?

Biotechnologies, AI/robotics, and neuroscience are of particular interest to me, and will all have a dramatic societal impact over the coming decades. I am also fascinated by technologies for mitigating the effects of global climate change (geoengineering), nanotechnologies, and approaches to develop new sources of energy.

CRISPR/Cas9, a new tool for quickly editing DNA, will alone facilitate altering the human genome and the ability to create new organisms and biological products. The benefits of CRISPR and other forms of

synthetic biology, along with advances in AI, are truly transformative, but are also accompanied by serious risks and dangers. Addressing those risks, and managing and adapting to the societal impact of emerging technologies have been my primary focus.

What has our development of intelligent machines taught us about human decision-making processes and ethical systems?

Building intelligent machines has forced scholars to think comprehensively about the many skills and capabilities that come into play in making appropriate decisions, including, but not limited to: emotional intelligence, social skills, the ability to deduce the beliefs and intentions of others, having a body and being embodied in the world, the capacity to recognize the meaning of words and symbols, the capacity to discern essential from inessential information, and an aptitude to be sensitive to moral considerations. Reason alone is not sufficient to produce intelligent machines capable of acting appropriately in a world inhabited by other people, animals, and an environment worthy of care and consideration.

You've spent years advocating the need for public discussion about what decisions we want machines to make for us, and the principles guiding such decisions. Has there been enough discussion?

By no means! Indeed, the few scholars advocating for responsible innovation in AI/robotics have been in the wilderness until the reawakened concern about superintelligence emerging from recent breakthroughs using deep learning approaches. However, I am hopeful that we'll make significant progress over the next few years and decade

toward shaping the development of AI. Nevertheless, serious questions about our ability to ensure that AI systems will be truly beneficial have yet to be answered. The public dialogue as to what we want and will accept in the development of smart machines has just begun.

What are the opportunities for making the world better?

I reject notions of inevitability, naïve techno-optimism, techno-pessimism, and simplistic techno-solutionism. Humanity needs to be vigilant if it wants to exact the benefits of technological possibilities while minimizing the harms. There are inflection points, windows of opportunity, where we can shape the trajectory of a new technology. A little adjustment early on can take us toward a very different destination. But these windows open and close very quickly. For example, there is an opportunity today to restrict the use of autonomous military weapons that make life and death decisions, but if we don't enact an international ban soon that opportunity will be lost. If there is no ban, the dangers in the development of AI will increase exponentially. Technological unemployment—the downward pressure new technologies, particularly robots, place on wage and job growth—is another area that requires attention now.

If human morality has an impact on intelligent machines, does it also work in reverse? Will machines have an impact on human values?

Machines are already having an impact on human values. Indeed, the very fact that we can create [intelligent machines](#) feeds into a scientific tendency to mechanize and pathologize human nature. On the other hand, the difficulty of developing machines capable of complex decision-making, and particularly moral decision-making, underscores what

remarkable creatures we humans are.

Provided by Yale University

Citation: Shaping tomorrow's smart machines—Q&A with bioethicist Wendell Wallach (2016, February 16) retrieved 28 June 2024 from <https://phys.org/news/2016-02-tomorrow-smart-machinesqa-bioethicist-wendell.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.