

Computer scientists fear robots might one day outsmart us

By JOHN MARKOFF
The New York Times

Updated Saturday, July 25, 2009 at 09:42 PM

A robot that can open doors and find electrical outlets to recharge itself. Computer viruses no one can stop. Predator drones, which, though controlled remotely by humans, come close to a machine that can kill autonomously.

Impressed and alarmed by advances in artificial intelligence, a group of computer scientists is debating whether there should be limits on research that might lead to loss of human control over computer-based systems that carry a growing share of society's workload, from waging war to chatting with customers on the phone.

Their concern is that further advances could create profound social disruptions and have dangerous consequences.

As examples, the scientists pointed to technologies as diverse as experimental medical systems that interact with patients to simulate empathy, and computer worms and viruses that defy extermination and could thus be said to have reached a "cockroach" stage of machine intelligence.

While the computer scientists agreed we are a long way from HAL, the computer that took over the spaceship in "2001: A Space Odyssey," they said there was legitimate concern that technological progress would transform the work force by destroying a widening range of jobs and force humans to learn to live with machines that increasingly copy human behaviors.

The researchers — leading computer scientists, artificial-intelligence researchers and roboticists who met at the Asilomar Conference Grounds on Monterey Bay in California — discounted the possibility of highly centralized superintelligences and the idea that intelligence might spring spontaneously from the Internet. But they agreed robots that can kill autonomously are either here or will be soon.

Exploiting avenue

They focused on the specter that criminals could exploit artificial-intelligence systems as soon as they were developed. What could a criminal do with a speech-synthesis system that could masquerade as a human being? What happens if artificial-intelligence technology is used to mine personal information from smartphones?

The researchers also discussed possible threats to human jobs, such as self-driving cars, software-based personal assistants and service robots in the home. Last month, a service robot developed by Willow Garage in Silicon Valley proved it could navigate the real world.

A report from the conference, which took place in private Feb. 25, is to be issued this year. Some attendees discussed the meeting for the first time with other scientists this month and in interviews.

The conference was organized by the Association for the Advancement of Artificial Intelligence (AAAI).

The meeting was organized by Eric Horvitz, a Microsoft researcher who is president of the association.

Horvitz said he believed computer scientists must respond to the notions of superintelligent machines and artificial-intelligence systems run amok.

The idea of an "intelligence explosion" in which smart machines would design even more intelligent machines was proposed by mathematician I.J. Good in 1965. Later, in lectures and science-fiction novels, computer scientist Vernor Vinge popularized the notion of a moment when humans will create smarter-than-human machines, causing such rapid change that the "human era will be ended." He called this shift the Singularity.

This vision, embraced in movies and literature, is seen as plausible and dangerous by some scientists such as William Joy, co-founder of Sun Microsystems. Other technologists, notably Raymond Kurzweil, have extolled the coming of ultrasmart machines, saying they will offer advances in life extension and wealth creation.

"Something new has taken place in the past five to eight years," Horvitz said. "Technologists are replacing religion, and their ideas are resonating in some ways with the same idea of the Rapture."

The Kurzweil version of technological utopia has captured imaginations in Silicon Valley. This summer an organization called the Singularity University began offering courses to prepare a "cadre" to shape the advances and help society cope with the ramifications.

"My sense was that sooner or later we would have to make some sort of statement or assessment, given the rising voice of the technorati and people very concerned about the rise of intelligent machines," Horvitz said.

"Loss of human control"

The AAI report will try to assess the possibility of "the loss of human control of computer-based intelligences." It will also wrestle, Horvitz said, with socioeconomic, legal and ethical issues, and probable changes in human-computer relationships. How would it be, for example, to relate to a machine that is as intelligent as your spouse?

Horvitz said the panel was looking for ways to guide research so technology improved society rather than moved it toward a technological catastrophe.

The meeting on artificial intelligence could be pivotal to the future of the field. Paul Berg, who received a Nobel Prize for chemistry in 1980, said it was important for scientific communities to engage the public before alarm becomes unshakable.

"If you wait too long and the sides become entrenched like with GMO," he said, referring to genetically modified foods, "then it is very difficult. It's too complex, and people talk right past each other."

Tom Mitchell, a professor of artificial intelligence and machine learning at Carnegie Mellon University, said the February meeting had changed his thinking. "I went in very optimistic about the future of AI and thinking that Bill Joy and Ray Kurzweil were off in their predictions," he said. But, he added, "The meeting made me want to be more outspoken about these issues and in particular be outspoken about the vast amounts of data collected about our personal lives."

Despite his concerns, Horvitz said he was hopeful that artificial-intelligence research would benefit humans, and perhaps compensate for human failings. He recently demonstrated a voice-based system that he designed to ask patients about their symptoms and to respond with empathy. When a mother said her child was having diarrhea, the face on the screen said, "Oh, no, sorry to hear that."

A physician told him afterward the response was wonderful. "That's a great idea," Horvitz said he was told. "I have no time for that."



KEN CONLEY / WILLOW GARAGE

This robot plugs itself in when it needs a charge. Some scientists worry these helpers might turn into masters.



Some researchers worry robots of the future might not be as congenial as C-3PO of "Star Wars" fame.