Society | DOI:10.1145/3077231

Sarah Underwood

Potential and Peril

The outlook for artificial intelligence-based autonomous weapons.

The history of battle knows no bounds, with weapons of destruction evolving from prehistoric clubs, axes, and spears to bombs, drones, missiles, landmines, and systems used in biological and nuclear warfare. More recently, lethal autonomous weapon systems (LAWS) powered by artificial intelligence (AI) have begun to surface, raising ethical issues about the use of AI and causing disagreement on whether such weapons should be banned in line with international humanitarian laws under the Geneva Convention.

Much of the disagreement around LAWS is based on where the line should be drawn between weapons with limited human control and autonomous weapons, and differences of opinion on whether more or less people will lose their lives as a result of the implementation of LAWS. There are also contrary views on whether autonomous weapons are already in play on the battlefield.

Participants in the first NGO Conference of the Campaign to Stop Killer Robots in London in 2013.

PHOTO BY SHARRON WARD FOR THE CAMPAIGN TO STOP KILLER ROBOTS

Ronald Arkin, Regents' Professor and Director of the Mobile Robot Laboratory in the College of Computing at Georgia Institute of Technology, says limited autonomy is already present in weapon systems such as the U.S. Navy's Phalanx Close-In Weapons System, which is designed to identify and fire at incoming missiles or threatening aircraft, and Israel's Harpy system, a fire-and-forget weapon designed to detect, attack, and destroy radar emitters.

The Campaign to Stop Killer Robots, which was founded in 2013 by a group of regional, national, and international non-governmental organizations (NGOs), agrees that no fully autonomous weapons are yet in use, but says existing systems could soon be extended to become fully autonomous and that the window to fulfill its ambition of achieving a preemptive ban on all such systems is closing.

The campaign's key tenet is that giving machines the power to decide who lives and dies on the battlefield is an unacceptable application of technology and makes human control of any combat robot essential to ensuring humanitarian protection.

Mary Wareham, global coordinator of the Campaign to Stop Killer Robots at Human Rights Watch in Washington D.C., explains the potential of precurser weapon systems that could be extended, exampling armed drones. Says Wareham, "Remotely piloted armed drones still have a human in the loop deciding on the selection of targets and force to be used, but new generations of arms could fly autonomously and complete missions with no human control. We don't want to see these systems in action."

From a robotic perspective, Arkin also defines autonomous machines as those that have no opportunity for human intervention, but says one of the concerns around LAWS is that there is no substantive agreement on what they constitute. He does not advocate a total ban on LAWS, but suggests the need to look at specific instances of weapons and decide on a one-by-one basis whether they are viable or should be banned. He explains: "I am a proponent of a moratorium until there is more understanding of autonomous weapons. We all agree we don't want a scenario like *The Terminator*, but we do need to talk about what we do want."

Arkin argues that a better understanding of autonomous weapons could lead to the development of intelligent autonomous military systems that could be precise in hitting targets and, at the same time, reduce civilian casualties and property damage when compared to the performance of human fighters, whose behavior in the theatre of war can be inconsistent and waver between heroic and atrocious.

Says Arkin, "We need to assume more responsibility for non-combatants, and not shoot first and ask questions later. In some circumstances, autonomous weapons could comply better with international humanitarian law than humans. But if weapons can't do as well as human fighters, they should not be put in place, hence my view on a moratorium."

With countries including the U.S., U.K., China, Russia, and South Korea developing autonomous weapons, and the U.K. Ministry of Defence estimating in 2011 that AI-based systems, as opposed to complex and clever automated systems, could be achieved in five to 15 years and that fully autonomous swarms of weapons such as drones could be available by 2025, the Campaign to Stop Killer Robots goes a step further than Arkin in its call for a pre-emptive ban on all autonomous weapons. It is pressing for the ban to be enacted through the implementation of international legislation or a new protocol under the Convention on Certain Conventional Weapons (CCCW), the key U.N. vehicle promoting disarmament, aiming to protect military troops from inhumane injuries, and seeking to prevent non-combatants from accidentally being wounded or killed by certain types of arms.

The most recent weapons to be excluded from warfare under the CCCW treaty are blinding lasers, which were banned in 1995.

The campaign defines three types of robotic weapons: human-**in**-the-loop weapons, robots that can select targets and deliver force only with a human command; human-**on**-the-loop weapons, robots that can select targets and deliver force under the oversight of a human operator who can override the robots' actions; and human-**out-of**-the-loop weapons, robots that are capable of selecting targets and delivering force without any human input or interaction. While these definitions are commonly used among developers of AI-powered weapons, their definitive meanings have yet to be agreed upon.

Reporting on a February 2016 roundtable discussion on autonomous weapons, civilian safety, and regulation versus prohibition among AI and robotics developers, Heather Roff, a research scientist in the Global Security Initiative at Arizona State University with research interests in the ethics of emerging military technologies, international humanitarian law, humanitarian intervention, and the responsibility to protect, distinguishes automatic weapons from autonomous weapons. She describes sophisticated automatic weapons as incapable of learning, or of changing their goals, although their mobility and, in some cases, autonomous navigation capacities mean they could wreak havoc on civilian populations and are most likely to be used as anti-material, rather than anti-personnel, weapons.

Roff describes initial autonomous weapons as limited learning weapons that are capable both of learning and of changing their sub-goals while deployed, saying, "Where sophisticated automatic weapons are concerned, governments must think carefully about whether these weapons should be deployed in complex environments. States should institute regulations on how they can be used. But truly autonomous systems—limited learning or even more sophisticated weapons—ought to be banned. Their use would carry enormous risk for civilians, might escalate conflicts, and would likely provoke an arms race in AI."

Toby Walsh, professor of AI at the University of New South Wales, Australia, says, "There are many dangers here, not only malevolence, but also incompetence, systems designed by those with malicious intent, or systems that are badly made. Today, the military could develop, sell, and use stupid AI that hands responsibility to weapons that can't distinguish between civilians and combatants. The technology is brittle and we don't always know how it will behave, so the last place to put AI systems that are trained on data in the environment is the battlefield, which is already a chaotic place.

"The real challenge is ensuring good outcomes of AI, but unexpected outcomes could be good or bad, and that is for us to decide."

"There are many dangers here, not only malevolence, but also incompetence, systems designed by those with malicious intent, or systems that are made badly."

Other perils identified by AI researchers in this space include unilateral use of autonomous weapons to support asymmetric warfare, the potential unpredictability of weapon behavior particularly where multiple systems interact as swarms, and the unimaginable human and material destruction that could result from terrorist use of such weapons.

Looking at the ethical issues of LAWS, Eric Schwitzgebel, professor of philosophy at University of California, Riverside with research interests in philosophy of mind and moral psychology, discusses AI-based systems as objects of moral concern and questions whether AI could become sophisticated enough to be conscious.

Schwitzgebel acknowledges that such a scenario is unlikely in the short term, but says it could be possible to create an autonomous system capable of experiencing joy and suffering at a similar level to a human. If such a system were to be sent to war and "die," he suggests this may not be morally different to the case of a human who is sent to war and dies, as the system was human enough that it would not want this to happen. Similarly, Schwitzgebel notes that if a system was sent to war against its will, this would be the moral equivalent of creating slaves and sending them to war.

Says Schwitzgebel, "We haven't thought through carefully what sorts of AI systems we need and don't need to be concerned about, and the differences between them and us that would make them morally different. Hypothetically, an artificial being could be created with moral rights and the capacities of a person. This sort of AI will not be developed any time soon, but development could go in this direction and should be stopped short of getting there."

Schwitzgebel cites more immediate dangers of deploying autonomous intelligences in combat as loss of responsibility and lack of predictability. The loss of responsibility for autonomous systems in combat makes it difficult to apportion blame when something goes wrong. The question of whether the weapon designer, deployer, or indeed, any other entity should take the blame is far from answered. Schwitzgebel suggests the diffusion of blame could be consistent with governments collaborating on undesirable uses of autonomous systems.

Lack of predictability could also become a more serious threat as AI systems become more complex. "Human soldiers in warfare can be unpredictable, but within limits as military commanders have an understanding of what has happened in various conditions in the past," says Schwitzgebel. "Autonomous systems could be more unpredictable than humans, which in warfare could lead to disastrous consequences. The ethics of autonomous weapons and issues of AI and philosophy are not as widely talked about as they should be."

With many questions about the benefits and dangers of LAWS still up in the air, and no international agreements in place to provide answers, the Campaign to Stop Killer Robots and other research organizations keen to ensure a ban on their development, manufacture, and deployment, are exerting pressure on governments to adopt and implement their approach.

While the campaign is concerned that the window of time to reach agreement on a ban on LAWS is closing as autonomous weapons are being developed, progress in its favor is beginning to be made and autonomous weapons are moving up the agenda following a December 2016 United Nations review of the CCCW.

Making a small piece of history, the U.N. voted during the review to start a formal process that might lead to a ban on LAWs. Of course, there are no guarantees that the process will be successful, but as Walsh puts it: "If states hadn't voted to start the process, there would have been no chance to finish." Russia abstained from the vote.

Countries participating in the vote agreed to set up an open-ended Group of Governmental Experts that will discuss nations' concerns about LAWS and the line between autonomous and non-autonomous weapons. The group will meet for two weeks in August this year, but the expectation is that it will take multiple years to reach consensus and add a protocol to the CCCW that will ban autonomous weapons operating beyond the boundaries of international humanitarian law.

"Autonomous systems could be more unpredictable than humans, which in warfare could lead to disastrous consequences."

Further Reading

Losing Humanity: The Case against Killer Robots
November 2012, Human Rights Watch and International Human Rights Clinic at Harvard Law School
https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

Views of the International Committee of the Red Cross (ICRC) on autonomous weapon systems
April 2016, ICRC
https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system

Three in Ten Americans Support Using Autonomous Weapons
February 2017, Ipsos
http://www.ipsos-na.com/news-polls/pressrelease.aspx?id=7555

IEEE Ethically Aligned Design Document Elevates the Importance of Ethics in the Development of Artificial Intelligence (AI) and Autonomous Systems (AS)
December 2016, IEEE
http://standards.ieee.org/news/2016/ethically_aligned_design.html

Lethal Autonomous Systems and the Plight of the Non-combatant
July 2103, Ronald Arkin, Georgia Institute of Technology
http://www.cc.gatech.edu/ai/robot-lab/online-publications/aisbq-137.pdf

Campaign to Stop Killer Robots
https://www.stopkillerrobots.org/

Sarah Underwood is a technology writer based in Teddington, U.K.

ACM Member News

DETERMINING NORMS FOR CYBER WARFARE

Patrick McDaniel, a Distinguished Professor in the School of Electrical Engineering and Computer Science at Pennsylvania State University (Penn State), says that when he was 11 years old, his father brought home a TRS-80 portable computer from Radio Shack, and handed him the manual to BASIC. "Within 10 minutes I was addicted, and I have never looked back. I have bachelor's, master's, and Ph.D. degrees in computer science, and it has never even been a thought to do anything else."

McDaniel obtained his undergraduate degree at Ohio University in 1989, and his master's degree at Ball State University in 1991. He then worked to develop some of the first IP networking hardware as a project manager at Primary Access Corp. in San Diego, which was acquired by 3Com in 1995.

He later earned his Ph.D. in computer science and engineering at the University of Michigan, Ann Arbor. McDaniel spent several years as a senior research staff member at AT&T Labs in New Jersey, before joining the faculty at Penn State in 2004.

McDaniel is director of the Institute for Networking and Security Research at Penn State, and also university lead for the U.S. Army Cyber Security Research Alliance, a 10-year project to develop an understanding of how to make security-relevant decisions in cyberspace.

One area McDaniel is focused on concerns the norms for international cyber warfare. He works to help define and set standards for what is allowable; in effect, a Geneva Convention for cyber warfare. "Right now, because nothing is set up, it is really hard to go to the U.N. Security Council for sanctions when you haven't set up any norms."

—John Delaney