IN DEPTH

Losing control of our humanity

Noel Sharkey / Maya Brehm
Chairman of the International Committee for Robot Arms Control and Emeritus Professor of Robotics and Artificial Intelligence, University of Sheffield, UK. / Geneva Consultant for the UK NGO, Article 36.

One of the greatest emerging challenges to world security in the 21st century is the development of autonomous weapons also called ‘fully autonomous weapons’, ‘killer robots’, ‘lethal autonomous robots’ or ‘lethal autonomous weapons systems’. These are weapon systems that once activated, select targets and attack them without further human intervention.

It is becoming progressively more difficult to find any new technological artefact that is not controlled by computer chips. The technologies of violence are no exception: computer devices are becoming ubiquitous for most modern weapons and guidance control systems. Currently almost all of these weapons are under ‘supervisory control’, where human control is simply mediated by a computer program.

Some states already use a number of weapon systems that intercept high-speed inanimate objects such as incoming missiles, artillery shells, mortar grenades or saturation attacks automatically. Examples include C-RAM, Phalanx, NBS Mantis and Iron Dome. These systems complete their detection, evaluation and response process within a matter of seconds and thus render it extremely difficult for human operators to exercise meaningful supervisory control other than switch them on and off. So far, such systems have been deployed in relatively uncluttered environments, devoid of civilians.

But there is an ever-increasing push by several states to develop distance weapons that could move outside the reach of human supervisory control. The US has conducted advanced testing on a number of autonomous weapons platforms such as X-47b – a

fast subsonic autonomous jet that can now take off and land on aircraft carriers, the Crusher – a 7 ton autonomous ground robot, and an autonomous hunting submarine. The Chinese are working on the Anjain supersonic autonomous air-to-air combat vehicle. The Russians are developing an autonomous Skat jet fighter. Israel has the autonomous Guardium ground robot and the UK is in advanced testing of the Mantis – a fully autonomous intercontinental combat aircraft.

Crude target recognition

A major problem with autonomous weapons is that identifying and selecting targets requires well-defined target recognition software. Yet current automatic target recognition methods used by the military are not fit for purpose except in narrowly restricted and highly uncluttered environments. Currently three main methods are used:

1 – Shape detection makes it possible to recognise a tank in an uncluttered environment, such as a sandy desert plain. Despite decades of research, it has proved extremely difficult to distinguish between a truck and a tank or any vehicle amongst clutter, such as trees.

2 – Thermal imaging detects heat radiating from an object and shows its movement. But it would be difficult with this method to distinguish a tank from a school bus.

3 – Radar detection is used by loitering munitions to detect enemy radar signals and bomb them. It is assumed that the target is an anti-aircraft installation, otherwise radar detection doesn't determine its legitimacy.

The targeting limitations of these methods are severe, even after decades of research. Thus the idea of developing autonomous weapons, outside of narrow restrictions, that could comply with the legal requirements on the use of lethal force under international human rights law and international humanitarian law is speculative and cannot be guaranteed. The technical problems may or may not be solved by some future discovery. Importantly, though, even an improved ability to recognise targets does not allow machines to assess whether a target is legitimate and whether the attack as a whole is permissible.

The appropriateness and legality of an attack is context-dependent and tends to be assessed on a case-by-case basis. Letting states carry on with developments in the hope that it will all work out poses a grave risk that autonomous weapons will be deployed irrespective of whether they are legally compliant or not. The only redress is a comprehensive, pre-emptive prohibition on the development, production and use of such systems.

“ Even an improved ability to recognise targets does not allow machines to assess whether a target is a legitimate target of attack and whether the attack as a whole is permissible ”

Taking action

In April 2013, an international civil society Campaign to Stop Killer Robots was launched calling for a pre-emptive ban on the development production and use of fully autonomous weapon systems. The campaign does not seek to ban autonomous vehicles or robots of any kind. Its scope is clearly focused on preventing the automation of the kill decision.

A month later, Christof Heyns, UN special rapporteur on extrajudicial, summary or arbitrary executions, called for a global moratorium on the use and development of lethal autonomous robots; a breathing space for nations to consider the implications of the development of such weapons. He concluded: “If used, they could have far-reaching effects on societal values, including fundamentally on the protection and the value of life and on international stability and security.” Like many, Heyns believes that delegating the decision to kill to machines may cross a fundamental moral line.

Following growing pressure from civil society to address the challenges raised by autonomous weapons systems, in November 2013, France put forward a proposal to the 117 states parties to the Convention on Certain Conventional Weapons for an expert

discussion meeting. The mandate was adopted and will take place in May 2014.

These expert discussions provide an impetus for states to formulate urgently needed national positions on this matter. So far, only the USA has published an official policy statement. DoD Directive 3000.09 provides guidelines on autonomous weapons. While pushing their further development, it requires that a human be 'in-the-loop', for the time being, when decisions are made about lethal force. The UK government has also asserted that all weapons under current policy will remain *'under human control'.*

Despite these affirmations, it remains unclear what exactly is meant by 'human control' or 'in-the-loop'. It could mean something as simple as pressing a button to initiate an attack or even programming a weapon for a mission. Clearly, that would be very different from the kind of human control considered appropriate in relation to existing weapons systems.

" There can be no guarantee to predict that LAWS can be used in compliance with international law. "

Engaged human control

An examination of scientific research on human supervisory control allows us to develop a classification consisting of five types of control:

1. Human engages with and selects a target and initiates any attack

2. Program suggests alternative targets and human chooses which to attack

3. Program selects target and human must approve before attack.

4. Program selects target and human has restricted time to veto

5. Program selects target and initiates attack without human involvement

For level 1 control it is critically important to understand that there are a strict requirements for *engaged human control*: a human commander (or operator) has full contextual and situational awareness of the target area at the time of a specific attack and is able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack. There must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack. There must also be a means for the rapid suspension or abortion of the attack.

Level 2 control might be acceptable if shown to meet the requirement of engaged human control. A human in control of the attack would have to be in a position to assess whether an attack is necessary and appropriate, whether all (or indeed any) of the suggested alternatives are permissible objects of attack, and to select the target which may be expected to cause the least civilian harm.

Level 3 is unacceptable. This type of control has been experimentally shown to create what is known as *automation bias* in which human operators come to accept computer generated solutions as correct and disregard or don't search for contradictory information.

Level 4 is also unacceptable. It does not promote target identification and a short time to veto would reinforce automation bias and leave no room for doubt or deliberation. As the attack will take place *unless* a human intervenes, this undermines well-established presumptions under international humanitarian law in favour of civilian character and status.

In the case of level 5 control there is no human involvement in the target selection and attack. As argued above, such weapons systems could not comply with international law.

Conclusion

There can be no guarantee that autonomous weapons can be used predictably in compliance with international law. It is impossible to predict how complex software will

react in all circumstances. Conflict regions are notoriously replete with unanticipated and changing events and the technology can be 'gamed', jammed, hacked or spoofed by an adaptive enemy. Clearly, therefore, there are strong moral, legal and security reasons for the pre-emptive prohibition, under international law of the development, production and use of autonomous weapons.

When states declare that there will always be 'human control' or a 'human in-the-loop' for automated weapons systems, we need to ask whether that control meets the requirements for engaged human control. We should also expect states to explain how they ensure that this requirement is met in their weapons review processes.

Photo: Official U.S. Navy flickr Page. Modified. Link to license.