ARTICLE 36 | APRIL 2015

KILLING BY MACHINE

KEY ISSUES FOR UNDERSTANDING MEANINGFUL HUMAN CONTROL

UNDERSTANDING HUMAN CONTROL OVER WEAPONS

In current practice, there is an expectation that human control is exercised over the use of weapons. This means when, where and how weapons are used; what or whom they are used against; and the effects of their use. Increasingly autonomous weapon systems threaten to erode what we have come to expect in terms of human control over weapons. Weapon systems that operate outside of the parameters of meaningful human control are neither ethically acceptable nor legally permissible.

Article 36 argues that meaningful human control over weapon systems is required in every individual attack. States should develop new international law to make this requirement explicit. Understanding how we exercise control over existing weapons provides critical guidance for developing an adequate response to increasing autonomy in weapon systems.

In the case of weapon systems that can detect and attack target objects without direct human intervention, critical aspects of human control broadly relate to:

- The pre-programmed target parameters, the weapon's sensor-mechanism and the algorithms used to match sensor-input to target parameters.
- The geographic area within which and the time during which the weapon system operates independently of human control.

The charts on the reverse illustrate these two key themes, and how they relate to each other.

ARTICLE 36

Article 36 is a UK-based not-for-profit organisation working to prevent the unintended, unnecessary or unacceptable harm caused by certain weapons. Article 36 undertakes research, policy and advocacy and promotes civil society partnerships to respond to harm caused by existing weapons and to build a stronger framework to prevent harm as weapons are used or developed in the future. The name refers to article 36 to the 1977 Additional Protocol I to the Geneva Conventions that requires states to review new weapons, means and methods of warfare.

www.article36.org

THE CAMPAIGN TO STOP KILLER ROBOTS

Article 36 is a founding member of the *Campaign to Stop Killer Robots.*

The Campaign to Stop Killer Robots calls for a pre-emptive and comprehensive ban on the development, production, and use of fully autonomous weapons, also known as lethal autonomous robots. This should be achieved through new international law (a treaty), as well as through national laws and other measures.

www.stopkillerrobots.org

THE CONTRACTING OR ACCEDING PARTIES RESERVE TO THEMSELVES TO COME HEREAFTER TO AN UNDERSTANDING WHENEVER A PRECISE PROPOSITION SHALL BE DRAWN UP IN VIEW OF FUTURE IMPROVEMENTS WHICH SCIENCE MAY EFFECT IN THE ARMAMENT OF TROOPS, IN ORDER TO MAINTAIN THE PRINCIPLES WHICH THEY HAVE ESTABLISHED, AND TO CONCILIATE THE NECESSITIES OF WAR WITH THE LAWS OF HUMANITY.×

× 1868 Saint Petersburg Declaration

WHEN MACHINES MAKE PEOPLE THE OBJECT OF ATTACK

The technology behind an algorithmic 'decision' to detect, select and cause harm to a target is highly complex. Characteristics of objects that are amenable to being 'sensed' by a machine, including their infrared emissions or their shape, or biometric information about persons, are used as 'proxy indicators' of a target.

Rendering the world machine-sensible reduces people to objects. This is an affront to human dignity. As the UN Special Rapporteur on extrajudicial, summary or arbitrary executions has observed, grave consequences for individuals are at stake in any use of force, requiring deliberative, human decision-making that considers ethical implications. Without this, the use of force becomes dehumanized and there is a vacuum of moral responsibility. Social processes of moral and legal reasoning on the use of force cannot be encoded into machines.

Related to the use of proxy indicators are concerns about probabilistic matching programs used for identifying and verifying appropriate objects of attack. Such programs have a specified 'tolerance of error' to offset their 'efficiency' in producing a match, with their 'accuracy' of producing a valid match (identifying an appropriate object of attack). Within the tolerance of error are objects that match the system's criteria (based on proxy indicators), but that are not appropriate objects of attack. The use of a weapon system that, by its very design, harms a certain number of people and objects that are not appropriate objects of attack (e.g. a civilian, a hospital, a civilian plane) is morally and legally highly problematic.

FIG.1 (OPPOSITE)

Persons and objects other than legitimate targets may fall within the target parameters of a weapon and be made the object of attack.

A pressure-activated anti-personnel landmine can be triggered by any person or object weighing more than a specified amount. Consequently, children, other civilians, heavy animals, and civilian vehicles can be harmed by landmines, even though they are not appropriate targets of attack.

If a sensor-fuzed weapon is used in an area that contains civilian objects, it may identify civilian tractors, buses, or aircraft and military medical vehicles as valid targets on the basis that their shape or heat signatures match pre-programmed target parameters. Such vehicles are, however, a priori protected from attack under international humanitarian law.

Similarly, a weapon system identifying targets on the basis of biometric signatures or similar proxy indicators may identify children, adult civilians and soldiers hors de combat as targets, even though these persons are not legitimate targets of attack under international humanitarian law.

INFORMATION, PREDICTABILITY AND CONTROL OVER THE RISK OF HARM

Any use of force must be sufficiently contained (geographically and in time) to allow the person(s) responsible for the planning and conduct of an operation and for its consequences to make informed judgments about the utility, necessity, moral acceptability and legality of the proposed use of force.

This requires those responsible for an attack to gather and understand information about the locations in which harm can be caused by a weapon system, and what these locations contain. This is of particular importance in terms of people and objects that may be present in the location where force is applied, but should not be attacked or that must be protected against the effects of military operations..

CONTROLLING THE LIMITS OF AN 'INDIVIDUAL ATTACK'

In connection with the conduct of hostilities, assessments about the legality of the use of force, such as the determination of whether force is used in keeping with the international humanitarian law rules on distinction and proportionality, are made on the basis of individual attacks. An attack can comprise different acts of violence against the adversary, but it is spatially and temporally bounded.

Weapons that can cause harm in multiple, geographically disparate, locations and over an extended period of time without direct human involvement risk expanding the notion of attack. This brings a serious risk of undermining the exercise of meaningful human control over the effects of weapons.

The risk that persons or objects other than legitimate targets are harmed by a weapon system depends on a number of factors. This risk increases and the human control exercised over the weapon's effects decreases, if

- the weapon operates independently for a longer period of time;
- the weapon operates over a wider geographical area;
- the weapon uses broader target parameters;
- and the weapon is used where there are a greater number of persons and objects that potentially match those parameters (a 'cluttered environment').

A RESPONSIBILITY TO UNDERSTAND THE RISKS TO CIVILIANS

An important challenge is ensuring that both the parameters of a target and the algorithm used to match objects with targets are such that the weapon system

- detects (identifies as targets) objects and persons that the weapon user intends to attack and is allowed to attack, but
- does not detect (identify as targets) people and objects that the weapon user does not intend to attack or is not allowed to attack.

In addition, harm to civilians and civilian objects in the vicinity of a target needs to be avoided, or at any rate, minimized.

For example, using a weapon system that identifies vehicles as targets on the basis of infrared emissions and shape in a civilian populated area involves a higher risk that civilians and civilian objects (e.g. a school bus) are made the object of attack and of causing incidental civilian harm, compared to an area largely devoid of objects other than legitimate targets.

Users of weapons have a responsibility to understand who and what risk falling within the target parameters of their weapon systems and to prevent harm to civilians and civilian objects by placing adequate controls, both, on the development of weapons and on their use.

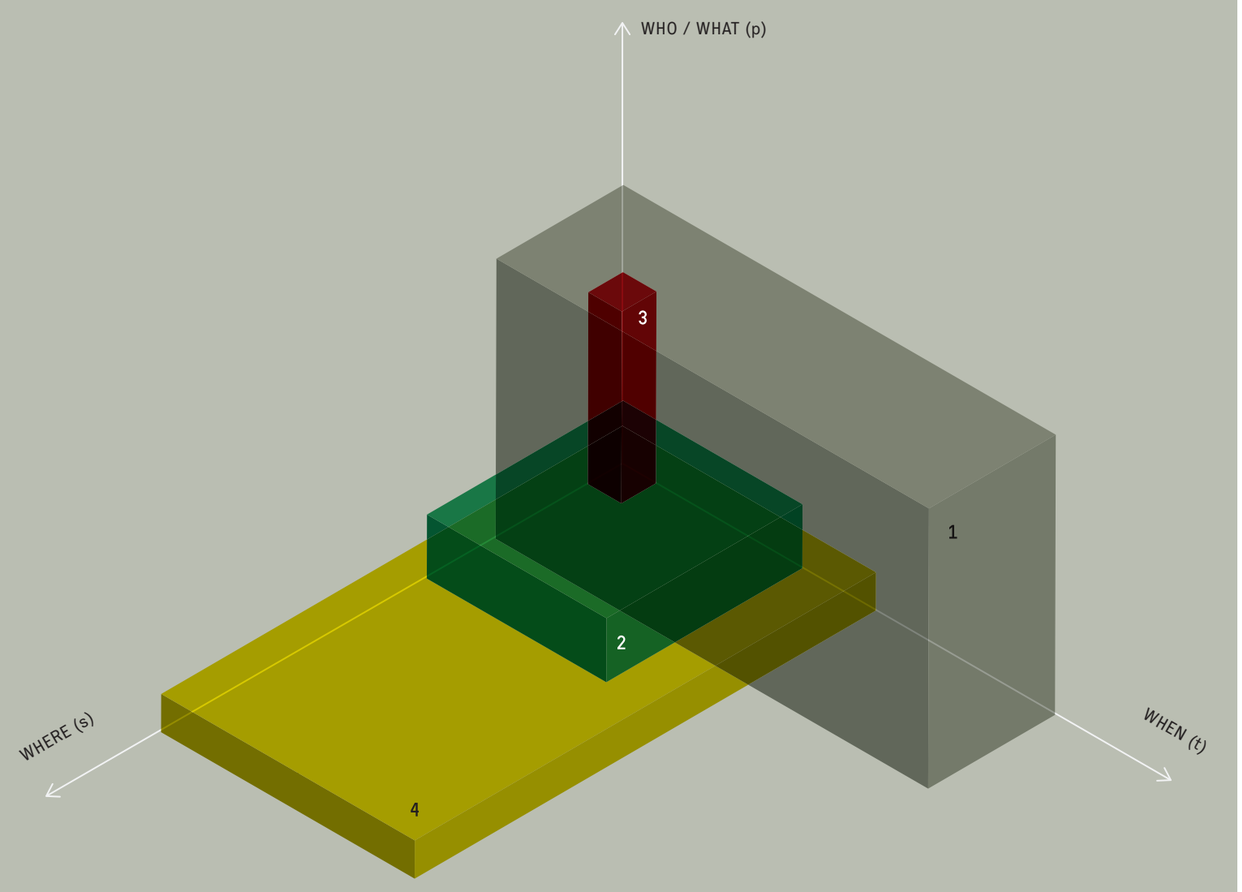

FIG.2

The volume of the boxes in the diagram represents the difficulty for a human commander in controlling the effects that can be produced by a weapon system.
The larger the box, the less control is exercised over the weapon. The time of operation of the weapon system extends along the 't' axis. The geographical space within which the weapon operates extends along the 's' axis. The vertical axis 'p' indicates who or what is at risk of being harmed, as a function of the breadth of the proxy indicators and the target environment.

Box 1: A mine field of pressure-activated **anti-personnel landmines** without a self-deactivating feature may be used in a limited geographical area (a short 's' line) but they continue to present a risk almost indefinitely (along the 't' axis). They use a very broad proxy indicator (weight over a certain amount) in an environment where many potential target objects may be present, which means that they have a high value on the 'p' axis. These weapons have been banned.

Box 2: An **anti-missile system** with a radar- and infra-red detector can operate during a short time period ('t') when turned on and off by a human operator. The system can scan a relatively large area ('s') for potential targets (missiles or aircraft, for instance), especially if the weapon platform is mobile, but the 'p' dimension can be fairly low if the proxy indicators are narrowly defined and the operating environment is relatively free from 'clutter' (scanning only the sky over the sea, not the littoral region, for instance), reducing the risk that persons or objects that are not legitimate targets are fired on by the system.

Box 3: An **air-launched missile with a millimeter-wave radar seeker** operating independently during a relatively short period of time (depending on target-search height and self-destruct setting) can scan for targets, such as armored vehicles (personnel carriers, tanks, or self-propelled guns) within a bracketed target area of limited dimensions. If used in an urban environment, such a weapon still presents a risk ('p') that civilian vehicles are made the object of attack and that civilians are harmed incidentally.

Box 4: A **weapon system using biometric data or facial recognition** to target a specific person (a 'targeted killing') could have relatively limited target parameters ('p'), but it could pose a risk of harm to the intended target, persons misidentified as the intended target, and bystanders in a vast area ('s') over an extended period of time ('t').

JUSTIFICATION IS REQUIRED FOR ENGAGING IN WAR AND IN OTHER FORMS OF VIOLENCE, AND THEREFORE, IT SEEMS TO FOLLOW, JUSTIFICATION IS REQUIRED FOR ENGAGING IN THE SEARCH FOR THE KNOWLEDGE TO PROVIDE THE MEANS TO FIGHT WARS AND COMMIT OTHER FORMS OF VIOLENCE.[×]

× John Forge

RECOMMENDATIONS

Human control over the use of weapons and their effects is essential to ensuring that the use of a weapon is morally justifiable and legal. Such control is also required for accountability over the consequences of the use of force. To demonstrate that such control can be exercised, states must show that they

- × understand the process by which a system identifies individual target objects, and
- × understand the context in space and time where an attack will take place.

Given the development of greater autonomy in weapons systems,

- + States should make it explicit that meaningful human control is required over individual attacks.
- + Weapon systems that operate without meaningful human control should be prohibited.
- + States should explain how control is applied over existing weapon systems, especially those with autonomous or automatic functions, and why they consider such systems to be acceptable and permissible.

Article 36
81 Rivington Street
London EC2A 3AY
United Kingdom
info@article36.org

www.article36.org

Article36