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.
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

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 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
INFORMATION, PREDICTABILITY AND CONTROL OVER THE RISK OF HARM

Any use of force must be sufficiently contained geographically and in time to allow the possibility 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 those 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 being used in keeping with the international humanitarian law on distinction and proportionality, are made on the basis of individual attacks. An attack can comprise different acts of violence against this adversary. But it is spatially and temporally bounded.

Weapons that cause harm in multiple, geographically dispersed, locations over an extended period of time without direct human involvement, risk expanding the notion of attack. This brings a serious risk of undermining the essence of meaningful human control over the effects of weapons. The risk 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 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 these parameters (a cluttered environment).

A RESPONSIBILITY TO UNDERSTAND THE RISKS TO CIVILIANS

An important challenge is ensuring 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; and
- 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 shapes 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 incipient civilian harm, compared to an area largely devoid of objects other than legitimate targets.

Users of weapon systems have a responsibility to understand and control the risk of malfunctioning of their weapon system. Those parameters of their weapon systems and to prevent harm to civilians and civilian objects by applying adequate controls, both on the development of weapons and on their use.

FIG. 2

The volume of the boxes is the diameter of the weapon system, the dimensions of the boxes indicate the diameter of the weapon system, the smaller the boxes, the smaller the diameter of the weapon system. The boxes are arranged in a way that allows the image of the weapon system to be seen from a certain distance. The weapon system operates in a specific area, the area is defined by the outer boundaries of the boxes. The boxes represent the potential area of effect of the weapon system.

FIG. 3

The figure shows a schematic of a weapon system. The weapon system consists of several components: a sensor, a data processing unit, a control unit, and a weapon. The weapon system operates in a specific area, the area is defined by the outer boundaries of the boxes. The boxes represent the potential area of effect of the weapon system.

WHEN MACHINES MAKE PEOPLE THE OBJECT OF ATTACK

The technology behind an algorithm to decide on the attack and cause harm to a target is highly complex. Characteristics of objects that are similar to being `nominal' by machines, including their infrared emissions or their shape, or kinetic information about persons, are used as `proxy indicators' of a target.

Reducing the world machines-enabled reduces people to objects. This is an attack 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 case of force, requiring deliberate, 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 are needed to unembed machines.

Related to the use of proxy indicators are concerns about predictive matching programs used for identifying and verifying appropriate objects of attack. Such programs have a specified tolerance of error to allow 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 of object, are at least two categories of objects or persons that are not appropriate objects of attack (e.g. a civilian, a hospital, a civilian plane) is morally and legally problematic.

FIG. 4

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

A weapon system is algorithmically matched and non-terrorized objects may be present in the target area. Non-terrorized objects may be harmed.

WHO / WHAT (p)

WHERE (s)

WHEN (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.

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.