Article36

BRIEFING PAPER | MAY 2014

Key areas for debate on autonomous weapons systems

Memorandum for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)

Geneva, 13-16 May 2014

Article 36 is a UK-based not-for-profit organisation working to prevent the unintended, unnecessary or unacceptable harm caused by certain weapons.
www.article36.org

Article 36 is a founding member of the Campaign to Stop Killer Robots.
www.stopkillerrobots.org

Ahead of the CCW's informal meeting of experts on lethal autonomous weapons systems, this paper suggests key areas for discussion and critical questions in efforts to address the concerns regarding autonomous weapons systems (AWS). The paper is an updated version of the paper 'Structuring debate on autonomous weapons systems' produced in November 2013.

In current practice, there is an expectation that human control is exercised over when, where and how weapons are used, and over their effects. This is implicit in existing international law governing the use of force. Increasingly autonomous weapons systems may erode what we have come to expect in terms of human control over weapons, and there is the possibility that weapons systems are developed that operate without meaningful human control. As a first principle for addressing concerns regarding AWS, states should, therefore, formulate as an explicit legal requirement that there be meaningful human control over individual attacks.

The CCW meeting of experts offers an important opportunity for government delegations to:

- Reaffirm that meaningful (or sufficient or appropriate) human control must be exercised over the use of weapons, and express concern over future weapons that could operate without meaningful human control;
- Explain how human control is exercised over existing weapons systems, especially those termed 'automatic' or 'semi-autonomous', and, where applicable, explain how present practice informs states' policy orientation toward autonomous weapons in the future;
- Where applicable, explain how 'human control', or its equivalent, is defined in relevant national policies;
- Support the development of an explicit prohibition, under international law, of weapons systems operating without meaningful human control over individual attacks.

Why we need to discuss autonomous weapons systems now

Fully autonomous weapons systems have not yet been deployed and used. But several States Parties to the CCW, including China, Israel, Russia, the United Kingdom and the United States, may be developing capacities that would enable greater combat autonomy for machines. Action is therefore needed to promote a common understanding of what should be considered acceptable when it comes to using armed force by means of AWS, how such activities should be internationally regulated, and where the line should be drawn against unacceptable AWS.

The terminology around this emerging weapons technology is not yet settled. Although sometimes referred to as 'lethal autonomous robots' (LARs), 'lethal autonomous weapons systems' (LAWS) or 'killer robots', concerns around AWS are not limited to the killing of human beings, but extend to any infliction of harm by means of an AWS, including incapacitation or injury of human beings, and material damage to the human and natural environment.

The term 'autonomous' is used by engineers to designate systems (such as a self-driving car) that can operate without direct human control or supervision in dynamic, unstructured, open environments based on feedback information from a variety of sensors. An autonomous weapons system (AWS)[1] is best understood as being composed of disparate soft – and hardware elements that work together – including sensors, algorithmic targeting and decision-making mechanisms, and the weapon itself.[2]

Identifying and attacking targets based on algorithmic 'decisions' made by a computer raises a host of ethical, legal, security, safety and other concerns that require urgent attention.[3] The remainder of this paper sets out some lines of inquiry along which States Parties to the CCW could structure their consideration of this issue.

Meaningful human control

The exercise of control over the use of weapons and concomitant responsibility and accountability for consequences are fundamental to the governance of the use of force and to the protection of the human person from the effects of weapons.

No state is likely to argue in favour of the release of AWS without any form of human control whatsoever – for example, an AWS that could roam at will, killing people without reporting back to a human operator. Likewise it is apparent that having a person 'in', 'on' or 'touching' 'the loop' of a weapons system does not in itself ensure that meaningful human control is exercised – for example, if that person simply pressed a 'fire button' every time a light came on without having any other information.

Whilst many would agree that the acceptability of weapons hinges on human control over their use and their effects, there are likely to be divergent views about the nature and extent of that control.

As some states move toward increasingly autonomous weapons systems, a key question for consideration concerns the threshold below which human control can no longer be exercised in a meaningful way.

It should be noted that whilst this paper uses the term 'meaningful human control' there are other terms that refer to the same or similar concepts. These include 'significant', 'appropriate', 'proper', or 'necessary' 'human judgement' or 'human involvement'.

A key factor in the control exercised over weapons is the information available to those responsible for weapon use, about the target, the target context and the physical effects the weapons will cause. Under IHL, those who plan or decide on an attack, should have sufficient information and control over a weapon to be able to predict how the weapon will operate and what effects it will produce in the context of an individual attack, and thus, to make the required legal judgements.

This raises the following questions:

- × What is the nature of human control to be exercised over an AWS?
- × At what point does human control over a weapons system cease to be meaningful?
- × To what extent can computer programming augment or enable 'human' control?
- × When are we no longer confident that international legal norms governing the use of force, provisions relating to the protection of the human person, and laws dealing with accountability for the consequences of the use of force can adequately be applied?

Human control over existing weapons

In order to possess this information and predict a weapon's effects, in current practice, controls, in technical, legal and policy terms, are placed on the operation of a weapon. Understanding how we govern existing weapon technologies and the controls in place, provides critical guidance for positioning ourselves in relation to emerging weapons technologies.

Consider the example of sensor-fused weapons. Sensor-fused weapons (such as Textron Defense System's CBU-97/CBU-105, the GIWS mbH manufactured SMArt 155 or the BAE Systems AB made BONUS-155) are deployed over a pre-defined target area. Once released, the weapon's sensors search for objects within that area that match a defined set of parameters (e.g. the heat signature of a combat vehicle engine). When the sensors detect a matching object, the weapon detonates to create an explosively formed projectile that will strike at this object.

In this example, the final determination of the target to attack is made by sensors and computer algorithms. The area that will be searched by the sensors varies for different weapons and this search area is positioned by a human person when they fire the weapon. After their launch, the commander has no further control over what the weapon will target within the target area – it will simply strike the first object which, according to the weapon's sensors, matches the programmed parameters of a valid target.

Another instructive example is the Brimstone, a UK anti-tank missile that has been described as a 'fully autonomous, fire-and-forget' weapon. During the search phase, Brimstone's millimetre wave radar seeker searches for targets, comparing them to a programmed target signature in its memory. The missile automatically rejects returns which do not match this programming and continues searching and comparing until it identifies a valid target or self-destructs.[4]

In these examples, once launched, the weapons operate autonomously. Critical aspects of how human control is exercised over such weapons – pertain to the programming of the target parameters and sensor mechanisms, and to the area within which and the time during which the weapon operates independently of human control.

Describing targets through 'proxy indicators'

The technology behind an algorithmic 'decision' to detect, select and attack a target is highly complex. One challenge is to ensure that the AWS correctly identifies objects as valid targets that the user wishes to attack, but that the AWS does not identify objects as valid targets that the user is not allowed to attack, or otherwise does not wish to attack. For this purpose, characteristics of objects, like their infrared emissions or shape (or potentially biometric data for persons) are used as 'proxy indicators' of a valid target.

Depending on the type and breadth of proxy indicator(s), objects that are not legitimate objects of attack can fall within the parameters of

a valid target. For instance, with regard to the sorts of sensor-fused weapons mentioned above, armoured fighting vehicles are not the only objects with engines that might be found in a combat zone. On the basis of its heat signature, a tractor, a lorry or a school bus could potentially be identified as a valid target and attacked by a sensor-fused weapon. Neither manufacturers nor states fielding such weapons have yet made publically available information on what civilian objects may match the target profiles of such weapons or what testing has been done to determine this. The lack of such information regarding existing weapons systems makes it difficult to accept claims that existing law is being fully implemented or that further authority should be given to such target identification mechanisms in the future.

It has been noted by the Peace Research Institute Oslo (PRIO), that Brimstone's ability to autonomously select targets was deemed "ill-suited to contemporary operations", especially in Afghanistan. "There, because of the conflict's complex nature, rules of engagement required that a human monitor the engagement right up until impact of the missile."[5] This challenge led to additional mechanisms being put in place to ensure human oversight of final target selection.

But the challenge of effectively and correctly identifying legitimate targets through proxy indicators is not solely of a technical nature. Under IHL, the legality of weapon use is generally assessed on a case-by-case basis, taking into account the circumstances of every individual attack. Whether an attack complies with basic rules of IHL governing the conduct of hostilities (necessity, proportionality, distinction, etc.) is strongly context-dependent. Equally context-specific assessments of necessity and proportionality are required under international human rights law for the determination of whether a particular use of force is legal.

These considerations raise the following questions:

- What characteristics are acceptable as indicators of a target of attack?
  - How are existing weapons systems programmed to identify valid targets?
  - What objects (persons) other than 'intended and legal targets' could be captured by these parameters and what research have states undertaken on this?

- How are states assessing the adequacy of these indicators?
  - Is it morally justifiable and legally acceptable to deploy an AWS without knowing what objects or persons could be attacked that are not intended and legal targets?
    - If states do not know what objects or persons could be 'wrongly' identified as valid targets, how can they assess a weapons system's compliance with international law as required under Art. 36 of 1977 Additional Protocol I?
    - How does a military commander apply relevant legal rules without knowing what objects could be targeted by an AWS in any given context?

- What does the context-dependency of legal assessments of the use of force imply for the choice and use of proxy indicators?
  - Is it morally justifiable and legally acceptable to deploy an AWS in the knowledge that a certain percentage of objects (persons) will be 'wrongly' identified as valid targets in any given context?
  - Is it consistent with the principles of humanity and of human dignity to base 'kill decisions' on a set of broad parameters that are applied mechanically without deliberative decision?

Controlling the context through space/time limitations

As noted previously, human control over existing weapons systems is also exercised in current practice through limitations, in technical, legal and policy terms, on the time during which and the space within which a weapon operates independently of human control.

In the case of the sensor-fused weapons discussed above, a human commander determines the position of the search-area within which the weapon acts independently. Similarly, Brimstone missiles can be programmed not to search for targets until they reach a given point, or only to accept targets in a designated box area.[6] If a relatively short time passes between the determination of the target area and an attack, and if the period during the which the weapon searches for targets is also short, then the commander should have a greater capacity to make reliable judgements about conditions in that target area.

The size and geographic location of the target area and the time window are important determinants of human control exercised over weapons systems. If the area is small, fixed in space, and the time window is short, a human commander should be expected to possess the necessary information to determine, at any given moment, what objects other than legitimate military objectives within that area risk being targeted by an AWS or otherwise affected by an attack, allowing her to predict the weapon's effects, apply the law, assess the risk to civilians, balance military and humanitarian considerations, and if necessary, suspend or cancel an attack.

Existing practices limiting the space and/or time within which a weapon operates independently raise important questions for the future management of AWS:

- What are the constraints in technical, legal and policy terms on the independent operation of existing weapons systems (such as naval or land based missile defence systems, sensor-fused weapons, unmanned remote-controlled weapons systems, or sentry robots)?
- What characterises environments within which existing weapons systems are permitted to operate independently?
- What criteria are used to assess the acceptability of limitations on the independent operation of weapons systems?
- What does the context-dependency of legal assessments of the use of force imply for time/space limitations on the independent operation of AWS?

Conclusion

Deploying AWS that operate outside of meaningful human control is neither ethically nor legally acceptable. The United Kingdom, for example, has already publicly stated "that the operation of weapons systems will always be under human control."[7] However, the key is to explain how this 'human control' is understood and to delineate the nature of human control that must be present for use of the weapon to be acceptable.

The questions brought out in this paper can provide entry points to thinking about how human control is currently managed in the operation of weapons systems. To understand how meaningful human control over future weapons systems can be ensured, states should start by explaining in detail how it is ensured over systems that are already in their arsenals, or that they are developing. In the absence of such an explanation, it would seem difficult to make assertions about the capacity of AWS to be used in accordance with legal requirements or to have a sufficient basis to evaluate the lawfulness of future AWS, as required under international law.

AWS are seen primarily as a concern of the future, but the discussion about their prospective management should not primarily be informed by hypothetical scenarios. Judgements about what is considered acceptable should not be based on producers' or users' claims about what is technically feasible. The parameters identified in this paper must be kept tight enough to ensure meaningful human control over AWS. Otherwise, there is a danger that what we consider meaningful human control today is gradually eroded. Claims of greater sophistication in the combination of sensors and programming could result in weapons systems being allowed to operate independently from human control over ever wider areas for longer periods of time as this becomes technically feasible or militarily expedient.

However, before working through the details of such questions it would seem important that states accept the underpinning principle – that human control is required, that it must be meaningful not formulaic, and that it must be applied to individual attacks. Along these lines, UN Special Rapporteur Christof Heyns, has argued that an initial step "would be to take a collective decision now, before such weapons are deployed, that humans, whether in the narrow or wider loop, should retain meaningful control over each decision to launch a potentially deadly attack – and to ensure that this line is not crossed."[8]

The linking of meaningful human control to individual attacks is significant because it is in relation to individual attacks that existing rules of international humanitarian law apply – it is over individual attacks that commanders must make legal judgements. The boundaries of what constitutes an individual attack should therefore be an important element of future discussion. States should be very wary of adopting a line of thinking that sees weapons as making legal judgements. Future weapons may bring further complexity in target identification, which will affect a military commanders' ability to predict the outcome of an attack, but it must be clearly acknowledged that the responsibility for legal judgements remains with the person or person(s) who plan or decide upon an attack.

The debate on AWS is only just beginning, but already there is a sense among the public that giving machines the power to target and kill human beings crosses a moral line – a line that many people instinctively recognise.[9] A debate on AWS in the CCW provides participating states and organisations with an important opportunity to shape our orientation toward the role of computers and machines in human violence, with broad implications for future warfare.

END NOTES

1 Note that the concern here is with the weaponization of increasingly autonomous systems. It is not with robotics and related fields of science, or civilian applications of this technology.

2 The components of an AWS – the sensors, the weapon, and algorithmic tracking and targeting mechanisms – need not be directly attached to each other or co-located, but merely connected through communications links. For instance, a computer located almost anywhere in the world could receive information from a surveillance drone, and use that information to initiate and direct a strike from a weapon system at yet another location, all without human intervention or supervision. For more detail, see, P. Asaro, 'On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making', International Review of the Red Cross, vol. 94, no. 886, 2012.

3 These are discussed in more detail elsewhere, see, e.g. Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, UN doc. A/HRC/23/47, 9 April 2013.

4 B. Handy (Ed.), Royal Air Force Aircrafts & Weapons, 2003.

5 N. Marsh, 'Defining the Scope of Autonomy', PRIO Policy Brief 02, 2014.

6 Ibid.

7 See, e.g., Lord Astor of Hever, Parliamentary Under Secretary of State, Defence, L Deb, 26 March 2013, c959.

8 Christof Heyns, Speech delivered at the 'Conference on Autonomous Weapons – Law, Ethics, Policy' on 24 – 25 April 2014 hosted by the European University Institute in Florence, Italy.

9 A survey of a representative sample of 1000 Americans conducted earlier this year by Dr. C. Carpenter of the University of Massachusetts Amherst (http://bit.ly/19iMIST) showed that across the board, 55% of Americans opposed autonomous weapons (nearly 40% were 'strongly opposed'). Of those who did not outright oppose fully autonomous weapons, only 10% 'strongly favored' them; 16% 'somewhat favored' and 18% were 'not sure'. It is interesting to note that military personnel, veterans and those with family in the military were more strongly opposed to autonomous weapons than the general public, with the highest opposition among active duty troops.