WRITTEN BY ELIZABETH MINOR

‘AUTONOMOUS’ WEAPONS: BUILDING TOWARDS INTERNATIONAL REGULATION FROM CONTRIBUTIONS TO THE CCW

Article36

WWW.ARTICLE36.ORG
INFO@ARTICLE36.ORG
@ARTICLE36

Article 36 is a specialist non-profit organisation, focused on reducing harm from weapons.

KEY MESSAGES

- There are two key problems within the discussions on autonomous weapons that states need to solve: firstly, whether some systems within the scope of discussion are fundamentally unacceptable; and secondly, how human control can be maintained over the rest, in order to adequately uphold both legal obligations and more profound moral and ethical principles.
- An effective structure for international legal regulation would prohibit certain configurations – such as systems that target people, and those that can’t be meaningfully controlled – and require positive obligations for meaningful human control over others, within a broad scope of sensor-based weapons systems that employ a particular *process* to apply force: that of matching sensor inputs to a “target profile” of characteristics following a system’s activation, emplacement, or deployment.
- Within states’ contributions to the Convention on Certain Conventional Weapons (CCW) this year, much substance has been elaborated that could support different elements within this approach. This could be brought together and developed as the groundwork towards a strong regulatory framework.

KEY RECOMMENDATIONS

- Notwithstanding the consensus that might (or might not) be reached within the CCW on commonalities and recommendations, the key to achieving an effective international response will be in developing the content and substance of its building blocks, which states and others should continue to do.
- There is space for states that find commonalities in each other’s positions on different elements of a regulatory structure to work together to build the content and substance of these shared understandings, for example through further collectively endorsed submissions to the CCW.

For states considering the issue of "emerging technology in the area of lethal autonomous weapon systems (LAWS)" within and beyond the Convention on Certain Conventional Weapons (CCW), there are two key problems to solve:

- firstly, whether some of the real or hypothetical weapon systems or configurations within the scope of these discussions are fundamentally unacceptable and must be ruled out of states' arsenals and practices; and
- secondly, how human control can be meaningfully maintained over the rest of the systems within this discussion's scope, in order to adequately uphold both legal obligations and more profound moral and ethical principles.

Following international discussions to understand the subject matter, a move towards clear common approaches and answers to these questions – which would provide the building blocks for effective international regulation – now requires the elaboration of more detailed positions and proposals from the states and others engaged in this debate.

Written submissions to the CCW this year (including national 'commentaries' on the Guiding Principles[1]), as well as discussion during September's Group of Governmental Experts (GGE), have provided a useful means for states to start doing this. 2020's work has allowed states to build on their understandings of the subject to give more form and content to concepts such as how human control over weapons can be maintained, and to explore answers to questions such as what gaps in interpretation or regulation there are in international law in this area. Much useful material has been produced in this regard – including specific proposals on the structure a future instrument for regulation could take, and interventions highlighting common ground in substance that could already provide the basis for this.

To support discussion, this short paper sketches out areas where we see useful content developing based on the outline of an international regulatory structure that we believe could effectively address the problem of increasing 'autonomy' in weapons systems. This paper is based on a reading of states' written submissions and impressions from September's GGE discussions – but is not intended to be a comprehensive analysis, nor one that reflects all points of content raised.

AN EFFECTIVE STRUCTURE FOR THE REGULATION OF SENSOR-BASED WEAPON SYSTEMS

States still have significantly different conceptualisations of the subject matter under discussion, from concepts of loops and automation/autonomy to "AI weapons".

For Article 36[2] (and the Campaign to Stop Killer Robots[3]), the discussion should encompass a broad scope of systems that employ a particular *process* to apply force: that of matching sensor inputs to a "target profile" of characteristics following a system's activation, emplacement, or deployment. With such systems, the exact time, place, and object to which force will be applied will not be known in advance. It is from this *uncertainty* that many concerns in this discussion arise, from issues of control to moral acceptability. Though many states have a significantly narrower conceptualisations of the subject matter under discussion than this broad scope, all definitions used at the CCW fall within it. We believe that considering regulation in relation to this broad range of systems, rather than narrowing the scope at this point, would be most effective for ensuring that the full range of concerns are addressed.

In our opinion, the most productive way forward will be to **consider applying legal obligations to this broad scope of systems**, centring prohibitions and regulations on human action and control regarding their use, as well as the core value of human dignity. Within this scope, certain systems should be **prohibited** as straightforwardly unacceptable, and the others should be subject to **positive obligations** on their design and use to ensure they remain under meaningful human control when used.

It is our position that the **targeting of people** through systems within this scope should be prohibited because this violates human dignity.[4] Systems must also be prohibited **that cannot be meaningfully controlled** by their users – for example, because the complexity of their functioning means that the range of outcomes they produce would not be sufficiently understood. A **structure of components to ensure meaningful human control** is needed for the remaining systems within this scope, to be applied on a case by case basis within individual attacks and operations. As some states have already noted, principles and practices of control might draw substantially from how states already manage uncertainty with less advanced sensor-based weapon systems.

CONTENT IN STATES' COMMENTARIES TOWARDS DEVELOPING AN EFFECTIVE REGULATORY APPROACH[5]

Dispersed across the 30 state commentaries and working papers submitted to the CCW in 2020,[6] there is already substance to support different elements within the approach outlined above, which could be developed and brought together towards a strong framework along the lines that we would consider effective. There has also been increasing acknowledgement of the potential value of a structure of both positive and negative obligations.[7]

Adopting a broad scope: Though the commentaries themselves reflect different understandings of the scope of the subject matter at hand, during September's CCW meeting at least one country as well as the International Committee of the Red Cross (ICRC) highlighted considering a broader range of systems or not adopting a definition that is too limited,[8] with others noting that "lethality" or causing immediate death might not be the key limit defining the work.[9] We believe it is important that the scope for regulation is not narrowed at this point.

Prohibiting anti-personnel use of systems/targeting people: A number of countries in their commentaries expressed opposition to human life and death "decisions" being carried out by machines[10] – recognised as a point of convergence by some in September's meeting[11] – and/or suggested restrictions could be made on the types of targets systems could apply force to.[12] Such positions should be

further explored and developed with respect to considering a prohibition on targeting people.

Prohibiting systems that cannot be meaningfully controlled: Several commentaries emphasised the need for the users of weapon systems to understand[13] how these will function in practice, with some linking this explicitly to legal compliance,[14] and/or to IHL-adjacent concepts of predictability, foreseeability,[15] explainability[16] or effectuating intent[17] through technology. Some expressed concern at systems that might "evolve"[18] or "set their own goals"[19] or highlighted that system design should ensure that sufficient human understanding of system functions is possible.[20] These suggestions can be linked to the need to prohibit systems whose complexities of functioning mean that their effects cannot be sufficiently predicted/foreseen or understood by their operators. This would be one element of prohibiting systems that cannot be meaningfully controlled.

Building the elements of a human control obligation: One key piece of common ground within states' commentaries and statements – frequently acknowledged during September's CCW GGE – is that further collective work is needed to "determine the type and extent of human involvement or control necessary" to ensure compliance with international law (including for many states both international humanitarian law, international human rights law and international criminal law) and to respond to ethical concerns.[21] It is useful and significant that, in general, states consider this to be the core area for substantial work, proposals and agreement – as is the consensus that human involvement is implicated in legal compliance.[22]

For Article 36, focusing on the *time, place, and target* to which force will be applied in an individual attack using a sensor-based system should be the key building blocks for constructing regulation for human control, to address the core issue of uncertainty about the point of application of force. Control over the components of the 'process' a system will apply (its target profiles) and control over the 'context' within which that process will operate (the area of time and space) are key.

Many state commentaries elaborated on what the key elements for control could or should be. For example, several mentioned applying **temporal and spatial limits**[23] to the use of systems and/or **controls on what contexts**[24] systems could be used in. Some commentaries could be read as implying that such limits could be necessary to ensure **sufficient proximity of force application to legal judgments**.[25] A number of commentaries also raised **limiting the task or operation**[26] undertaken with a system. This is an important concept to develop in relation to, for example, considering limiting the number of force applications (e.g. munitions fired) a system could carry out following activation.

Points raised regarding **understanding what systems might apply force to in practice**[27] (including 'false positives'[28]) and **limiting types of targets**[29] to enhance control and limit unintended consequences, included one proposal to place **limits on systems' target profiles depending on the operational environment.**[30]

There was recognition in several commentaries and again during September's discussion that the exact requirements for adequate levels of control might **vary depending on the tool and context.**[31] We believe therefore that placing discussions in the context of considering **individual attacks** (as at least one commentary did[32]) is helpful, in order to focus on human action—rather than the generalities and technicalities of systems.

Many commentaries suggested **technical elements such as contact/recall/supervision/self-destruction**[33] as a component of human control. Though such elements may be important in the control of systems, we believe that priority should first be given to agreeing what the generally applicable principles of human control should be, rather than the technical mechanisms that might support the implementation of these. For example, self-destruction might support the need to limit the time of operation of a system or the proximity of force application to a legal decision: it is not a principle of control in itself. This ordering of priority is important because regulation in this area – as in the rest of the law – should be addressed to human action and decision making. Through a core focus on human action, regulation can be more 'future proof' to future system developments. Additionally, decision-makers should be cautious of giving *undue* significance to what, currently, may be mainly imagined technical fixes to the problems posed by (also) hypothetical weapons systems.

MOVING FORWARD

Within states' commentaries and statements to the CCW this year, there has been a move towards elaborating more detailed points of substance regarding how problems with increasing 'autonomy' in weapons systems could be addressed, including through an international framework. From our perspective, there is already content that can be developed and elaborated, as well as some common ground, in some of the key areas and boundary lines that an instrument should address. Notwithstanding the consensus that might (or might not) be reached within the CCW on commonalities and recommendations, there is space for states that share common ground on the substance of a strong response to the issues to work together and develop these perspectives, including through further joint inputs to the CCW. The key to developing an effective international response will be in this substantial work.

ENDNOTES

1 See "Annex III, Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System," Revised Draft Final Report, CCW/MSP/2019/CRP.2/Rev.1, 15 November 2019, https://www.unog.ch/80256EDD006B8954/(httpAssets)/815F8EE33B64DADDC-12584B7004CF3A4/$file/CCW+MSP+2019+CRP.2+Rev+1.pdf, p10.

2 See Richard Moyes, "Autonomy in weapons systems: mapping a structure for regulation through specific policy questions," Article 36, 2019, http://www.article36.org/wp-content/uploads/2019/11/regulation-structure.pdf, and Richard Moyes, "Target profiles as a basis for rule-making in discussions on autonomy in weapons systems," Article 36, 2019, http://www.article36.org/wp-content/uploads/2019/08/Target-profiles.pdf.

3 See Campaign to Stop Killer Robots, "Key elements of a treaty on fully autonomous weapons,"2019, http://www.article36.org/wp-content/uploads/2019/11/regulation-structure.pdf.

4 See Maya Brehm, "Targeting people," Article 36, 2019, http://www.article36.org/wp-content/uploads/2019/11/targeting-people.pdf.

5 In this section, references are given to papers and statements given by states, to indicate examples of which commentaries/statements referred to content and themes in the areas discussed in this paper. These references will not necessarily mean that those countries support in whole, in part (or in any way) our proposals or analysis – nor that their intentions in their contributions necessarily align with our interpretations of how these contributions could or should be developed towards an effective regulatory structure

6 Available at UNODA meetings place, "Convention on Certain Conventional Weapons – Group of Governmental Experts on Lethal Autonomous Weapons Systems: Documents" https://meetings.unoda.org/section/group-of-governmental-experts-gge-on-emerging-technologies-in-the-area-of-lethal-autonomous-weapons-systems-laws-documents-4929-documents-4947/ and shared via email

7 For example, the group of states that submitted a Joint 'Commentary' (Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico, and New-Zealand) have acknowledged this in their discussion of structure, and also for example the commentary from the Non-Aligned Movement mentions a need for "prohibitions and regulations"

8 Statements to the GGE by Switzerland

9 Statements by for example Brazil, Germany, Ireland, Mexico

10 For example, commentaries and working papers from Cuba, the Non-Aligned Movement, Panama and Portugal

11 For example, a statement by Chile that was welcomed by several delegations on points of commonality amongst many that could provide a basis for a way forward included stating as a general understanding that states should "prohibit the design, development, or deployment of weapons or weapons systems that make life or death decisions"

12 Limiting types of targets (though what these types might be was not necessarily discussed) was raised in, for example, the commentary papers of Spain, Sweden and the Joint 'Commentary' submitted by Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico, and New-Zealand

13 Understanding systems was highlight in commentaries by Finland and the US, for example

14 For example, commentaries by Austria, Germany, Netherlands

15 More than 10 commentaries mention concepts around predictability/foreseeability

16 For example Germany, Spain

17 Commentaries of Israel and the US describe weapons as tools for effectuating intent

18 Commentary of Switzerland

19 Commentary of Sweden

20 Various commentaries, for example by Spain

21 See Chair's paper, Commonalities in national commentaries on guiding principles https://documents.unoda.org/wp-content/uploads/2020/09/Commonalities-paper-on-operationalization-of-11-Guiding-Principles.pdf

22 Expressed in Guiding Principle (c) (above note 1)

23 At least 14 commentaries

24 For example, commentaries by Austria, Portugal

25 Content in the commentary of the Netherlands, and the Joint 'Commentary', could suggest such a reading

26 For example commentaries of Austria, Portugal

27 The Joint 'Commentary' for example discusses the need for "reliability and predictability in the identification, selection and engagement of targets". The Non-Aligned Movement and South Africa discuss as an issue whether systems would hit a single target object or risk indiscriminate effects

28 Commentary of the US

29 See above note 11

30 Commentary of Switzerland

31 For example commentaries of Finland, Germany, Israel, Japan, Netherlands, Sweden, Switzerland

32 Commentary of France

33 At least 14 commentaries

Acknowledgements:

The project under which research and publication of this paper was undertaken is part funded by a grant from Irish Aid. The ideas, opinions and comments herein are entirely the responsibility of Article 36 and do not necessarily represent or reflect Irish Aid policy.

Design: bb-studio.co.uk