



# CAMPAIGN TO STOP KILLER ROBOTS

## **Commentary for the Convention on Conventional Weapons Group of Governmental Experts on lethal autonomous weapons systems**

20 May 2020

This commentary has been prepared on behalf of the [Campaign to Stop Killer Robots](#), the international coalition of 160 non-governmental organizations in 66 countries working to prohibit fully autonomous weapons and retain meaningful human control over the use of force.

The Campaign has participated in every Convention on Conventional Weapons (CCW) meeting on lethal autonomous weapons systems since the first in May 2014. In the view of the Campaign, states should be launching negotiations on a legally binding instrument or new CCW protocol, as so many have suggested.

The guiding principles adopted by CCW states in 2018 and 2019 are, on their own, not an adequate or appropriate response to the multiple concerns raised by increasing ‘autonomy’ and the dangerous prospect of removing meaningful human control from the use of force. The CCW principles were simply intended to guide the deliberations. They were never supposed to be an end in themselves or intended to provide the structure for or outcome to CCW work on lethal autonomous weapons systems.

However, the CCW provides a useful forum for states to elaborate their views and work to achieve common understandings to build greater collective policy clarity. In this regard, we draw attention to the request from the 2020 chair of the CCW’s Group of Governmental Experts (GGE) on lethal autonomous weapons systems, Ambassador Janis Karklins, for states to provide their “recommendations in relation to the clarification, consideration and development of aspects of the normative and operational framework on emerging technologies in the area of lethal autonomous weapons systems.” To that end, he invites CCW commentaries or working papers on the “operationalization of the guiding principles at the national level.”

This commentary shares the views of the Campaign to Stop Killer Robots on 1) the need to retain meaningful human control over the use of force; and 2) how to operationalize this requirement via the normative framework of a new international treaty. It provides recommendations for how a treaty could be effectively structured, through a multi-faceted approach, to meet the objective of prohibiting fully autonomous weapons and retaining meaningful human control over the use of force.

This commentary draws from the Campaign’s detailed paper on “Key Elements of a Treaty on Fully Autonomous Weapons” and related Frequently Asked Questions document.<sup>1</sup>

## **I. Retaining human control**

Since international deliberations on killer robots began seven years ago, human control, decision-making, and/or judgment has been at the center of discussion. It is now widely regarded as critical to the acceptability and legality of future weapons systems.

All eight CCW meetings on lethal autonomous weapons systems have seen strong interest in the importance of retaining human control over weapons systems and the use of force. This is reflected in the principle on human-machine interaction that CCW states participating in the GGE added last year to a set of guiding principles developed in 2018.<sup>2</sup>

Human-machine interaction attracted the greatest interest by far during the virtual Berlin Forum on lethal autonomous weapons systems attended by more than 60 countries on 1-2 April 2020. There was widespread recognition at the Rio Seminar on autonomous weapons on 20 February 2020 that human control is where states should focus their collective work.

The CCW principle on human-machine interaction has been accurately described by one NGO as “a placeholder for more substantial work elsewhere” as it provides no significant guidance on how to ensure human control as autonomy in weapons systems increases.<sup>3</sup> To determine the quality and extent of human-machine interaction, the principle merely suggests “a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole.”

The Campaign recommends states explore in depth how to retain meaningful human control over the use of force. We prefer “control” to other terms such as judgment and intervention, which imply a weaker role for humans than control, and could be insufficient to address all concerns. The modifier “meaningful” ensures that control is substantive.

The concept of meaningful human control can be distilled into decision-making, technological, and operational components:

---

<sup>1</sup> Campaign to Stop Killer Robots, “Key Elements of a Treaty to Ban Fully Autonomous Weapons,” November 2019. <https://www.stopkillerrobots.org/wp-content/uploads/2020/04/Key-Elements-of-a-Treaty-on-Fully-Autonomous-WeaponsvAccessible.pdf>; Campaign to Stop Killer Robots, “Frequently Asked Questions,” November 2019. <https://www.stopkillerrobots.org/wp-content/uploads/2020/04/FAQ-Treaty-ElementsvAccessible.pdf>

<sup>2</sup> The Guiding Principles state, “Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular International Humanitarian Law (IHL). In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole.” Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, CCW/GGE.1/2019/3, 25 September 2019, pp. 3-4. <https://undocs.org/en/CCW/GGE.1/2019/3>

<sup>3</sup> Article 36, Critical Commentary on the “Guiding Principles,” November 2019. <http://www.article36.org/wp-content/uploads/2019/11/Commentary-on-the-guiding-principles.pdf>

- Decision-making components of meaningful human control help ensure humans have the information and ability to make decisions about whether the use of force complies with legal rules and ethical principles. In particular, the human operator of a weapon system should have: an understanding of the operational environment; an understanding of how the system functions, including what it might identify as a target; and sufficient time for deliberation.
- Technological components are embedded features of a weapon system that can enhance meaningful human control. They include: predictability and reliability; the ability of the system to relay relevant information to the human operator; and the ability for a human to intervene after the activation of the system.
- Operational components make human control more meaningful by limiting when and where a weapon system can operate and what it can target. Factors that could be constrained include: the time between a human’s legal assessment and the system’s application of force; the duration of the system’s operation; the nature and size of the geographic area of operation; and the permissible types of targets (e.g., personnel or material).

While none of these components are independently sufficient to amount to meaningful human control, all have the potential to enhance control in some way. In addition, the components often work in tandem. Further analysis of existing and emerging technology could help determine which of these or other components should be codified in a legal instrument as prerequisites for meaningful human control.

## **II. Normative framework needed**

A legally binding instrument or international treaty is the optimal normative framework for dealing with the many serious challenges raised by fully autonomous weapons. A new international treaty could lay down explicit rules to ensure appropriate constraints on autonomy in weapon systems and resolve differing views on human control over the use of force. Most importantly, it would show that states are serious about responding appropriately and with urgency to this existential threat to humanity.

In the spirit of contributing to shared concepts and understandings, the Campaign’s “Key Elements” paper outlines in more detail on the essential elements for a legally binding instrument on lethal autonomous weapons systems. To meet the objective of prohibiting fully autonomous weapons and maintaining meaningful human control over the use of force such a treaty could be comprised of three types of core obligations: (1) a general obligation to maintain meaningful human control over the use of force, along with (2) prohibitions and (3) positive obligations to implement it.

### *1. General obligation*

A general obligation to maintain meaningful human control over the use of force provides an overarching provision to facilitate compliance with applicable legal and ethical norms. The obligation should focus on control over conduct (“use of force”) rather than specific technology. This approach would help future-proof the treaty by obviating the need to predict how technology will develop. The term “use of force” also makes the general obligation applicable to both situations of armed conflict and of law enforcement.

## *2. Prohibitions*

The legally binding instrument should prohibit the development, production, and use of weapons systems that select and engage targets based on sensor processing and are inherently unacceptable for ethical or legal reasons. The clarity of the prohibitions would facilitate monitoring and enforcement, and their absoluteness would create a strong stigma against the banned systems.

The new instrument should prohibit weapons systems that by their nature select and engage targets without meaningful human control. The prohibition should cover, for example, systems that become too complex for human users to understand and thus produce unpredictable and inexplicable effects. These complex systems might apply force based on prior machine learning or allow critical system parameters to change without human authorization. Such weapons systems would run afoul of the new instrument's general obligation discussed above.

The prohibitions could also extend to specific other weapons systems that select and engage targets and are by their nature, rather than their manner of use, problematic. In particular, the treaty could prohibit weapons systems that select and engage humans as targets, regardless of whether they operate under meaningful human control. Such systems would rely on certain types of data, such as weight, heat, or sound, to represent people or categories of people. In killing or injuring people based on such data, these systems would contravene the principle of human dignity and dehumanize violence. A prohibition on this category of systems would also encompass systems that are designed to or unintentionally target groups of people based on discriminatory indicators related to age, gender, or other social identities.

## *3. Positive obligations*

Finally, the normative framework or treaty should include specific positive obligations to ensure that meaningful human control is maintained in the use of all other systems that select and engage targets. The obligations would require states to ensure that weapons systems that select and engage targets and are not already prohibited as inherently unacceptable are used only with meaningful human control.

The content of the positive obligations should draw on the components of meaningful human control discussed in the previous section. For example, the treaty could require that operators understand how a weapon system functions before activating it. It could set minimum standards for predictability and reliability. In addition, the treaty could limit permissible systems to those operating within certain temporal or geographic parameters.

These positive obligations would help preserve meaningful human control over the use of force and establish requirements that in effect render the use of systems operating as fully autonomous weapons unlawful.

## **III. The way forward**

Since 2014, states have made some progress at the CCW to identify key issues of concern regarding autonomy in weapons systems. Virtually all states have acknowledged the importance of human control over the use of force and several have committed not to acquire or develop lethal autonomous weapons systems. A total of 30 countries have called for a ban on such weapons systems, while group of states such as the Non-Aligned Movement (NAM)

have repeatedly called for a legally-binding instrument stipulating prohibitions and regulations on lethal autonomous weapons systems.

Yet the CCW talks have yielded little in the way of a lasting multilateral outcome due to the opposition by a handful of military powers, most notably Russia and the United States, which firmly reject proposals to negotiate a new international treaty or protocol. The CCW has heard proposals for political declarations and codes of conduct, as well as for greater transparency. Such measures have value, but are by themselves insufficient to deal with this serious threat to humanity and some could be more productively undertaken after the international legal framework has been put in place.

Therefore, states should work to achieve measurable progress by the CCW's Sixth Review Conference in December 2021. Focused deliberations will help lay the groundwork for the international ban treaty that is urgently required to retain meaningful human control over the use of force. To achieve progress, the Campaign recommends states use their CCW commentaries to:

1. Identify factors to help determine the necessary quality and extent of human control over weapons systems and the use of force;
2. Express their preferred normative framework and its basic content, be it a legally binding ban instrument or another form of regulation.