

CAMPAIGN TO STOP KILLER ROBOTS

Response to GGE Chair's Guiding Questions September 2021

This paper sets out the Campaign to Stop Killer Robots' response to the additional questions circulated by the Chair of the Group of Governmental Experts on 12th August 2021.

1. Definition / characteristics

A. What specific characteristics would inherently render a weapons system incapable of being used in compliance with IHL?

Responsibility for compliance with International Humanitarian Law [IHL] lies with humans and not systems. The law is addressed to humans and it is they that must ensure compliance. A weapon system is not, in itself, capable or incapable of complying with the law. Accountability under IHL primarily arises during the use of a weapon system by a human operator. It is the user who is responsible for its use and who can be held accountable for any violations of the law.

In order to comply with IHL including the principles of distinction, precaution, and proportionality, the human operator must be capable of asserting meaningful control over the use of force. Absent such control over the critical functions of a weapon system the human operator would be unable to determine the specific target, location, timing, duration and extent of the force applied. If the operator cannot dictate the effects of the weapon system or the system does not function as intended, the operator cannot ensure compliance with IHL.

There are a wide range of factors that would determine whether or not human control can be said to be meaningful. These factors include whether the system is understandable, predictable and reliable and whether the system can be limited in its geographic and temporal scope. As set out in our [recommendations for elements of a new treaty on autonomous weapons](#), it is necessary to consider decision-making, technological and operational components in order to determine whether or not meaningful human control can be exercised over a weapon system.

Weapons systems that cannot be used with meaningful human control, such as those that are unpredictable or where mission parameters can change during an operation without human approval, would be incapable of being used in compliance with IHL and therefore should be prohibited.

B. Where and how to draw the line to determine what autonomous weapons systems would be lawful or not?

The determination of what types of autonomous weapon system should be considered lawful includes consideration of laws other than IHL, including International Human Rights Law [IHRL], and

International Criminal Law [ICL]¹. In addition, ethical considerations should be central to determining where the line should be drawn.

In the view of the Campaign to Stop Killer Robots, a line should be drawn around two kinds of autonomous weapons systems.

First, if an autonomous weapon system cannot be used with meaningful human control, then it cannot be used in compliance with existing legal rules and should be prohibited. Such a system could not be used in compliance with IHL, particularly the principles of distinction and proportionality. It also would present challenges for compliance with IHRL, especially the right not to be arbitrarily deprived of life, and create an accountability gap under IHL, IHRL, and ICL. It would also raise ethical concerns related to moral responsibility, human dignity and the loss of human agency in the use of force.

Second, autonomous weapon systems that target humans should be prohibited because they raise fundamental ethical concerns in addition to legal and practical challenges. By 'autonomous weapon systems that target humans' we mean sensor-based weapons systems where the target profiles are intended to represent people. Such systems would use proxy indicators (such as weight, heat-shape, movement or specific biometrics) as a basis for encoding patterns of sensor data as representations of a human.

Autonomous weapon systems that target humans, whether civilian or combatant, would dehumanise people, harming them on the basis of processing of sensor inputs and treating them as data, sensed and sorted by a machine. In killing or wounding people based on such abstractions, these systems would undermine human dignity, a fundamental principle of international human rights law. To allow sensors, data and software to determine who lives and who dies is morally unacceptable.

In addition to this fundamental moral objection, using sensor-based systems to target people autonomously raises a range of legal and practical concerns. Such systems would either be designed to target (a) a certain person or certain groups of people or (b) all people in a location.

a. Targeting a certain person or certain groups of people

If it is claimed that an autonomous weapon system is theoretically capable of distinguishing between people to whom force may be applied and those who must be protected, then acute concerns arise regarding compliance with the legal requirement of distinction and for the protection of certain classes of persons. Under IHL, people are targetable only on a contextual, case-by-case basis. The legal status of humans in conflict can change depending on their participation in hostilities and whether as combatants they are wounded or have surrendered. Claiming that machines can make such determinations undermines the law and legal processes and dilutes human legal and moral responsibility. Additional serious problems relating to prejudice and bias would arise if systems used target profiles to identify 'targetable people'

¹ For a detailed analysis of autonomous weapon systems and international human rights law see '[Shaking the Foundations. Human Rights Impact of Autonomous Weapon Systems](#)', Human Rights Watch and Harvard Law School's International Human Rights Clinic, 2014

based on certain characteristics such as race, gender, age, and other characteristics. Historically marginalized groups would certainly face higher rates of error, as has been demonstrated in several studies.²

Furthermore, systems that are particularly complex or opaque (“black box”) would make it difficult or perhaps impossible to offer a meaningful explanation of why certain people were targeted in certain circumstances; raising additional issues of responsibility and accountability.

b. Targeting all people in a location

If systems were enabled to target all people in a particular location, they would risk having indiscriminate effects, contrary to established rules of IHL. It might be argued that use could be restricted to areas from which civilians are excluded. Yet such an approach is likely to be unreliable in practice. Further, this approach still shifts the burden of avoiding harm onto the civilian population, thus eroding the presumption of protected status and undermining the general principle of the protection of the civilian population against the effects of hostilities (as set out in preamble of the CCW.)

These ethical, legal and practical problems would be most straightforwardly addressed through a legally binding instrument that prohibits autonomous weapon systems that cannot be used with meaningful human control, and those that target humans; and regulates other autonomous weapons systems to ensure they are not used without meaningful human control.

2. Human-machine interaction / human control

A. How to move forward to clarify the quality and extent of human-machine interaction/human control that would be required to use weapons systems based on emerging technologies in the area of LAWS in compliance with IHL?

In order to clarify the quality and extent of human-machine interaction and human control, states should begin the process of negotiation of a legally binding instrument. Discussions at the CCW have produced a wealth of expert reports and opinions on the issue. As has been seen in other processes, further clarification will be best achieved if states take the next step and begin the process of negotiation. At that stage, states can develop and agree on the precise legal and technical parameters of human-machine interaction and human control that are needed to comply with international law and ethical norms.

The Campaign to Stop Killer Robots has published recommendations on the normative and operational framework for autonomous weapon systems and a [‘Key Elements of a Treaty on Fully Autonomous Weapons’](#) which includes details on the types of human-machine interaction and human control needed to ensure compliance with international law and ethical principles.

² See for instance: www.ohchr.org/Documents/Issues/Racism/SR/Call/campaigntostopkillerrobots.pdf and <https://bit.ly/SEHLACAI-AWS> (note 6).

Meaningful human control requires that a human can make a moral judgement over the acceptability of the effects of an attack, fulfil her/his obligation to apply the law, and that there is a human user who is morally and legally responsible for the effects of an attack. These principles should form the foundation when clarifying the quality and extent of human-machine interaction/human control.

B. Which elements of context would influence the quality and extent of the required human-machine interaction/human control/human judgment?

The context in which an autonomous weapon system is used will significantly influence whether or not the level of control a human operator has can be considered meaningful. The greater the 'complexity' and unpredictability of the context the greater the quality and extent of control will be required.

There are a range of factors that contribute to the complexity of the context. These include

- o Physical environment
 - Land, sea, air
 - Terrain—mountains, woods, desert, open field
- o Climate and weather
- o Time of day
- o Urban area or rural area
- o Presence of humans, especially civilians
- o Nature of target
- o Geographic range of weapon system
- o Duration of weapons system's mission
- o Speed of operation and how dynamic the environment is

Such factors are interlinked and may combine to create greater complexity of the context. It is necessary for a human operator of an autonomous weapon system to evaluate these range of factors and determine whether meaningful human control can be exercised at the time the system is to be used.

C. In what sense would limits and constraints (type of target, duration, geographical area,...) be useful to ensure compliance with IHL and how?

As set out above, autonomous weapon systems that cannot be used with meaningful human control and systems that target human beings should be prohibited.

The broad scope of autonomous weapon systems that are not prohibited must still be subject to regulations to maintain meaningful human control. Whether or not meaningful human control can be exercised will be context-based, dynamic, multidimensional, and situation-dependent. Regulations should therefore provide clarification on how meaningful human control should be assessed on a case-by-case basis.

Constraints on the time and space of the system's operation exemplify these kinds of regulations. The wider the geographic area covered by the system and the longer the duration of its mission the

less control the human operator has over the effects of an attack. Therefore, temporal and spatial limits are needed to ensure the human operator retains meaningful control and thus has the ability to comply with IHL.

As discussed under the previous question, the specific limits may depend on the context. The context in which the autonomous weapon system is used is one of the factors that will be relevant to such an assessment. As above, the greater the complexity of context, the greater the degree of control the human operator will need for it to be meaningful.

In addition to consideration of the context in which the weapon system is being used, it is necessary that the human operator is able to understand, evaluate and predict the system. The system user should understand and have meaningful information on how a system works and has worked in the past; and be able to provide a comprehensive explanation, including on the conditions that will trigger an application of force. Human control would be enhanced through an ability of the system user to evaluate the system and the context during the course of its functioning and by having mechanisms to intervene as required. The system should also enable predictability of outcomes that align with the user's intent and the foreseeable consequences of specific actions being taken, which serves to ensure human responsibility and accountability in the use of force.

Regarding constraints on types of targets, autonomous weapon systems that target humans should be prohibited. As discussed elsewhere this raises serious legal and ethical concerns and should be prohibited. If the autonomous weapon system is not specifically designed to target a human, but it is possible that a human may be harmed, for example if the weapon is targeting a vehicle or building with humans inside, then a higher degree of control will be required for it to be meaningful.

D. Some delegations have stressed that human control is context-based, dynamic, multidimensional, and situation-dependent. How to evaluate the minimum level of human control that would be required in some or all situations to ensure compliance with IHL?

The Campaign to Stop Killer Robots agrees that human control is context-based, dynamic, multidimensional and situation dependent. For these reasons, the 'minimum' level of human control that is required is inherently flexible. It is not possible to provide an exhaustive list of what is specifically required due to the infinite scenarios in which autonomy can be used in a weapon system. Instead, it is more useful to agree that, in any scenario, the human operator's control must be 'meaningful' in order to meet obligations under international law, including IHL, IHRL and ICL. This means amongst other things that a human can make a moral judgement over the acceptability of the effects of an attack, fulfil her/his obligation to apply the law, and that there is a human user who is morally and legally responsible for the effects of an attack. As set out above, clarification on the precise legal and technical parameters relating to human control that are needed will be best achieved through commencing a process of negotiating a legally binding instrument.

E. At what stage of the life-cycle should human control be exercised?

Meaningful human control should be exercised at all stages of a weapon systems' life-cycle, including in the design, verification and testing, deployment, operation and monitoring stage. However, meaningful human control is especially important at the pre-deployment stage where limits of

operation are set (time, space, type of target) and the time of use because decisions at these moments directly influence the effects of an attack and affect legal and ethical compliance.

F. Which decisions, activities, and processes across the life-cycle of a weapons system would collectively contribute towards and enable appropriate human-machine interaction/human control/human judgment? What would be the interaction between these various decisions, activities, and processes and how would they vary based on the operational context and the characteristics and capabilities of the weapons system?

As set out above, it is the human operator of the weapon that is primarily accountable for using the weapon in compliance with existing legal rules. The processes of ensuring appropriate human control throughout the research, design, development, production and training stages for an autonomous weapon system can only be effective if the criteria that the system will ultimately be required to comply with at the point of use is properly understood and incorporated throughout. This in itself is a challenge because it would be impossible to train an autonomous weapons system to incorporate the myriad of elements that can be present in a situation of conflict.³

There are different points throughout the lifecycle of the weapons systems—such as design, programming, testing, training and monitoring—that will have an effect on the level of control at the time targets are selected and engaged. Those activities must be undertaken with a clear understanding and appreciation of what characteristics the weapon system will need to have to be used lawfully at the time targets are selected and engaged. In order for effective decisions, activities and processes to be undertaken across the life-cycle of a weapon, it is therefore necessary to determine the criteria for the weapon system's lawfulness at the point of use.

Examples of decisions, activities, and processes include ensuring the weapon system is understandable, reliable and predictable; the user is trained and understands the way the weapon systems works; the weapon system is tested extensively to evaluate compliance with legal and ethical norms; and clear processes are developed on how the weapon system should be used in compliance with legal and ethical norms.

Weapons reviews under Article 36 of Additional Protocol I can also play an important role in promoting IHL compliance, but those reviews are not a panacea. These reviews provide an obligation to review weapons, means and methods of warfare against states legal obligations. We therefore need to establish the legal obligations in relation to autonomous weapon systems as a precondition for having effective review processes. They need to be done transparently and consistently, and clear international standards on autonomous weapons systems are needed to ensure reviews are uniform.

G. How would the analysis of existing weapons systems help elaborate on the range of factors that should be considered in determining the quality and extent of human-machine interaction/human control/human judgment?

³ See for instance UNIDIR's side event: [Data issues in military autonomous systems](#).

Analysis of existing weapons systems would provide examples of what types of human control are currently exercised, and where in the life-cycle of the weapons systems they occur. It could also help understand possible concerns related to the increasing levels of autonomy in weapon systems. Analysis would also illuminate whether these measures have provided meaningful control that ensures compliance with international law. Research on existing systems have already advanced discussions of autonomous weapons systems. Transparency is essential to the effectiveness of such study.

3. **Ethics**

A. How to move forward and clarify the ethical dimension of our debate, including in light of the principles of humanity and the dictates of public conscience?

Autonomous weapons systems raise a host of ethical as well as legal concerns, and many of these concerns are intertwined. IHL's Martens Clause, for example, brings ethics into law by invoking the principles of humanity and dictates of public conscience. Ethics needs to remain a key part of the CCW debate.

Delegating the decision to take a human life to machines undermines human dignity and should be recognized as a challenge to the principles of humanity. The deep ethical concerns expressed by faith leaders, scientists, governments, states, private sector tech companies, military veterans, the public, the UN Secretary General and more should be documented as evidence that these emerging weapons systems are problematic under the dictates of public conscience. In a [global survey](#) the most cited reason for opposition to autonomous weapons systems is that they would “*cross a moral line, because machines should not be allowed to kill*”.

The ethical dimension also needs to be integrated into the response we develop. The prohibitions and positive obligations proposed here would respond to the ethical challenges raised by these systems. For example, the proposed prohibition on machines targeting people would protect human dignity and avoid algorithmic bias. In addition, ethics should help us frame policy and governance for autonomous weapons systems--and other uses of AI and emerging technologies--by considering their potential impact on groups that have been historically oppressed or dominated.⁴

Ethical concerns about the means and methods of warfare have been guiding IHL since its inception hundreds of years ago. Furthermore, the international community has committed to several principles and engagements related to the management of autonomy in emerging technologies that are also relevant to this discussion, such as the OECD principles on AI and the UNESCO Recommendations on ethics and AI. CCW states parties can draw on these models as they move toward negotiating a legally binding instrument on autonomous weapons systems.

⁴ For more on this, see: CHENGETA, T. The Relevance of Ethics in AWS debates. In ICRC, College of Europe, [New Tech on the Battlefield: Friend or Foe?](#) 2020.