



ICRC

Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts on Lethal Autonomous Weapons Systems 9–13 April 2018, Geneva

Statement of the International Committee of the Red Cross (ICRC)

Further consideration of the human element in the use of lethal force; aspects of human-machine interaction in the development, deployment and use of emerging technologies in the area of lethal autonomous weapons systems

Mr Chair

The ICRC advocates a human-centred approach, focused on the obligations and responsibilities of humans in the use of weapons with autonomy in their critical functions.

The ICRC's view is that a minimum level of human control is required to ensure

- compliance with international humanitarian law rules that protect civilians and combatants in armed conflict, and
- ethical acceptability in terms of the principles of humanity and the public conscience.

Legal basis for human control

The ICRC is clear that the law is addressed to States and humans, and the legal obligations under international humanitarian law ultimately rest with combatants who plan, decide upon, and carry out attacks.¹

Combatants will require a minimum level of human control over weapon systems with autonomy in their critical functions so that they can effectively make legal judgements - of distinction, proportionality and precautions – in specific attacks.

Human control can take different forms during the development, activation, and operation of an autonomous weapon system.² However, these legal judgements are context specific.

Therefore, concerns will arise where the design and/or use of the weapon interferes with combatants' ability to make the necessary legal judgements in carrying out attacks.

For example, if an autonomous weapon system is activated over a certain area and/or duration, and the commander/operator who authorises the launch does not know where and when the subsequent attack will take place, the question arises: how can they effectively ensure distinction, judge proportionality, or take precautions should the circumstances on the ground change?

¹ Machines can never "apply" international humanitarian law, and responsibility and accountability for decisions to use force cannot be transferred to machines, computer programs or weapon systems.

² First, the development and testing of the weapon system ('development stage'); second, the decision by the commander, or operator, to activate the weapon system ('activation stage'), and third, the operation of the autonomous weapon system during which it independently selects and attacks targets ('operation stage').

Ethical basis for human control

In the ICRC's view, ethical considerations also demand a minimum level of human control.

From an ethical viewpoint, "meaningful" human control would be the type and degree of control that preserves human agency and upholds moral responsibility in decisions to use force.

This requires a sufficiently direct and close connection to be maintained between the human intent of the user and the eventual consequences of the operation of the weapon system in a specific attack.

Ethical and legal considerations may actually demand some similar constraints to ensure human control is maintained.

However, ethical concerns about loss of human agency in decisions to use force, diffusion of moral responsibility and loss of human dignity could have more far reaching implications – potentially preventing the development and use of anti-personnel autonomous weapon systems, and even limiting the applications of anti-materiel systems, depending on the risks for human life.

Meaningful human control in practice

In the ICRC's view, the GGE should work to determine – for the broad category of weapons with autonomy in their critical functions – what does "meaningful", "effective" or "appropriate" human control mean in practice? What type and degree of human control is required?

Key elements of human control include

- predictability and reliability
- human supervision, and the ability to intervene and deactivate, and
- operational constraints.

For each of these elements there are lessons to be drawn from human control in existing weapons with autonomy in their critical functions.

Whether an autonomous weapon system will operate within the constraints of international law after it has been activated will depend on the predictability and reliability³ of

- the weapon system itself (including the algorithms used)
- the environment of its use, and
- the interaction between the two.

The level of human supervision, and the ability to intervene and deactivate is also important. Without human supervision it is hard to see how operators/commanders can take into account changes in the situation, and ensure predictability in the effects from the use of the weapon.

In many circumstances, constant human supervision of the weapon system and the target area may be required, so that the operator has sufficient information and understanding

³ Predictability is the ability to "say or estimate that (a specified thing) will happen in the future or will be a consequence of something". Applied to an autonomous weapon system, predictability is knowledge of how it will function in the circumstances of use, and the effects that will result. Reliability is "the quality of being trustworthy or performing consistently well". In this context, reliability is knowledge of how consistently the machine will function as intended, e.g. without failures or unintended effects.

about the operation of the weapon system, the environment of use, and the interaction of the two, over time and in space.

Communication links that permit adjustment of the engagement criteria and the ability to cancel the attack, as well as sufficient time for such intervention, may also be necessary.

This type of human supervision and potential for deactivation is seen in existing weapons with autonomy in their critical functions.⁴

Finally, other operational parameters and constraints are also important for human control, and for legal and ethical assessments, including:

- the task the weapon is used for – is it used to defend against incoming missile or to search for varied targets over a wide area?
- the types of targets it attacks – is it used to attack an incoming rocket or to attack humans directly?
- the type of force and munitions it employs (and associated effects);
- the operating environment – is it used in an “uncluttered” environment at sea without civilians present or use in a populated area on land?
- the duration of autonomous operation (time-frame) – is the autonomous mode activated for a short period or for several hours “loiter” time?
- the scope of movement over an area (mobility) – is the weapon stationary (fixed in place) with constrained movement and range or mobile over a wide area?

Further work by the GGE on these elements will help determine what “meaningful”, “effective” or “appropriate” human control means in practice, and where limits on autonomy must be placed.

Weapon systems falling outside those limits will then be more easily identified as unlawful and/or ethically unacceptable autonomous weapon systems.

Thank you

For further analysis by the ICRC on the legal and ethical basis for human control please see the following publications:

Autonomous weapon systems under international humanitarian law, November 2017:
<https://www.icrc.org/en/document/autonomous-weapon-systems-under-international-humanitarian-law>

Ethics and autonomous weapon systems: An ethical basis for human control? April 2018:
<https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control>

⁴ For example, existing counter-rocket, artillery and mortar weapons retain the ability, even with incoming projectiles, for a human operator to visually verify the projectile on screen and decide to cancel the attack if necessary.