

# CHARACTERISTICS OF AUTONOMOUS WEAPON SYSTEMS

CCW Meeting of Experts, 14 April 2015

Neil Davison, Science Adviser, Legal Division, International Committee of the Red Cross

## INTRODUCTION

The first challenge to enable coherent discussions of the legal, ethical and humanitarian implications of autonomous weapon systems, is developing some common understanding of what type of weapons we are talking about

This is something many delegations have already emphasized in their opening statements.

We do not need a definition but perhaps some general characteristics, or boundaries for the discussion, to help distinguish autonomous weapon systems from other weapons.

I will concentrate my remarks on this, with suggestions on two areas:

- 1) **General characteristics** for understanding autonomous weapon systems; and
- 2) **Contextual factors** for understanding human control.

### 1) GENERAL CHARACTERISTICS

With this in mind, I have **three suggestions** of general characteristics, or bounds for discussion, for understanding autonomous weapon systems.

#### *i) We should focus on the 'critical functions' of weapon systems*

- A simple way of understanding an **autonomous weapon system** is as a **weapon that can independently select and attack targets**. (This is consistent with most definitions).
- So, after initial activation, it is the weapon system itself that **selects** and **attacks** a target without further human intervention, validation or approval.
- It is the machine – using its sensors, programming, and onboard weapon(s) – that takes on tasks ordinarily carried out directly by humans.
- In other words, there is a shift of decision-making processes (or actions) on the use of force from man to machine.
- The functions that enable the weapon system to operate in this way are the functions directly controlling the targeting process.
- These are what the ICRC has described as the '**critical functions**' of **selecting** (i.e. search for, detect, identify, track, select) **and attacking** (i.e. intercept, use force against, neutralise, damage or destroy) **targets**.
- At a basic level it is autonomy in these critical functions that **distinguishes** autonomous weapon systems from all other weapon systems. And it is autonomy in these functions that **raises** specific legal and ethical questions.

**ii) We should focus on autonomy in the use of force and not autonomy in terms of technical sophistication**

- Autonomy in these critical functions of weapon systems, and therefore the use of force, is not determined by the technical sophistication (or degree of “intelligence”) of the weapon.
  - I’ll use an **example** to illustrate this more clearly.
  - Imagine a **robotic gun system**, which has an **autonomous mode** and that is used to protect an area or border from intruders. Now consider the following **two configurations**:
    - i) The gun is connected to a basic video sensor via a simple computer interface to detect human-like shapes. When the autonomous mode is activated, the weapon system will fire independently if it identifies a suitable target according to its sensor and programming.
    - ii) The same gun connected to a network of advanced sensors via a computer system with highly complex decision-making algorithms designed to help distinguish legitimate human targets, while constantly assessing changes in the environment. When the autonomous mode is activated, the weapon system will fire independently if it identifies a suitable target according to its sensors and programming.
  - The first configuration could be built today. The second configuration is not technically feasible today, and may never be feasible.
  - Some people might describe the first weapon as ‘automated’ and the second as ‘autonomous’. But these technical distinctions are not the most relevant issue.
  - More important is the **defining characteristic** they share: Both these weapon systems, once activated, can select and attack targets without human intervention.
  - This common characteristic means they **raise the same legal questions** under international humanitarian law (IHL), e.g.:
    - Can the weapon **distinguish** a lawful target from an unlawful one?
    - Can the weapon make judgements of **proportionality** regarding military advantage and expected incidental civilian casualties, or damage to civilian objects?
    - Can the weapon take **precautions** and cancel an attack if the situation changes?
    - And who is **accountable** if the weapon causes a violation of IHL?
  - This common characteristic also means they **raise the same ethical questions**, e.g.:
    - Is it ethically acceptable, under the principles of humanity and dictates of public conscience, for the weapon to select and attack targets, in particular human targets, without human intervention?
  - So, to reiterate, we must focus on autonomy in terms of selecting and attacking targets (critical functions), and not autonomy in terms of technical sophistication or “intelligence”.
- iii) We should exclude other autonomous functions, and non-weaponized systems**
- It seems clear that, in focusing on critical functions, we would exclude other autonomous functions, or non-weaponized autonomous robotic systems.

- A particular weapon system might have a variety of different autonomous functions – for example, take-off and landing, navigation, flying or driving, and control of sensors.
- But these autonomous functions do not directly determine the ability of the system to independently use force by selecting and attacking a target. (*A caveat is that other autonomous functions may present other risks or could be used to cause harm, e.g. flying into a building*).
- If a weapon system has other autonomous functions, such as flying, but the critical functions remain remotely operated then it would not be an autonomous weapon system.
- This type of distinction clearly separates autonomous weapon systems from those whose critical functions are operated remotely, which is how we currently understand armed drones to operate. (*NB: If the critical functions later become autonomous they the system would become an autonomous weapon system*).
- Given the recognition of the dual-use nature of robotics technology, it's also important to emphasise that autonomous robotic systems that are not weaponized, whether civilian or military in nature, would also be excluded by this focus on critical functions.
- Examples might include autonomous robots used for reconnaissance, for search and rescue, or for any other non-weapon purpose.

## 2) CONTEXT RELATED FACTORS INFLUENCING HUMAN CONTROL

- Keeping in mind these general characteristics, connected to autonomy in the critical functions of selecting and attacking targets, I will add some thoughts on human control.
- It is uncontroversial to say, as many delegations have already emphasized, that meaningful, adequate, or effective human control must be maintained over weapon systems, and over the use of force.
- But a question that remains is whether there are additional **context related factors** that influence human control over weapons and the use of force.
- One approach is to say that **any** ability of a weapon to independently select and attack targets without human intervention means there is a loss of meaningful human control, i.e. regardless of the context.
- Another approach is to ask: are there **additional factors, depending on the context**, that contribute to maintaining human control when a weapon system has autonomy in its critical functions?
- In other words, **is meaningful human control context dependent?**
- It seems that many existing weapons with autonomy in their critical functions have specific contextual constraints: **For example**, commonly mentioned air defence systems:
  - They are overseen in real-time by a human operator;
  - They have autonomous 'modes' for specific periods (not permanently autonomous);
  - They are constrained in their tasks, types of targets, and environments of use.

- With this in mind, I want to emphasise **three context related factors**, or parameters, that seem important for considering any weapon with autonomy in its critical functions:
  - i) The level of **human supervision** over the weapon
 

*Is weapon system supervised by humans in real-time? (On site? Remotely?)*

*Can the weapon system be deactivated? (On site? Remotely? By software or physical controls?)*

*In other words, is a “human on the loop”? And to what extent? Or is the weapon system effectively “out of the loop”?*
  - ii) The **freedom of action** of the weapon, which is determined by several **sub-factors**:
 

**Tasks:** *What are the limits in the tasks it carries out (e.g. defensive versus offensive roles? Narrowly or widely defined types of attacks?)*

**Targets:** *What are the limits on targets it is designed to attack? (Objects? Humans? Both?)*

**Environments:** *What are the constraints on the environments in operates in? In relatively simple, predictable environments (e.g. air defence of a ship at sea) or in complex, dynamic, and therefore unpredictable, environments (e.g. urban warfare)?*

**Time-frame:** *What are the time limits on its autonomous operation? Does it have a temporary autonomous mode? Or is it autonomous over long periods?*

**Mobility and area of operation:** *Is it fixed in one place or can it move? If it's mobile, what are the constraints on its movement over a geographical area?*
  - iii) Finally, a factor that is interconnected to the first two factors, the **technical capability of the weapon**:
 

*How **predictably** will the weapon operate in all the circumstances it may encounter?*

*How **reliably** will the weapon operate as intended in all the circumstances it may encounter?*
- I would suggest that further thinking is needed on how these different parameters affect human control over the selection and attack of targets, and therefore legal and ethical discussions.
- And this does need to be a theoretical exercise.
- A first concrete step we can take is to learn from existing weapon systems with autonomy in their critical functions, and ask:
  - How is meaningful human control maintained?
  - And through which parameters and processes (technical or operational)?
- This may help better understand the factors and parameters that are crucial for ensuring meaningful, effective, or adequate human control of the critical functions of weapon systems.
- It may also help to better identify the developments in autonomous weapon systems that raise particular concerns.