

SENSOR-BASED TARGETING SYSTEMS: AN OPTION FOR REGULATION

WWW.ARTICLE36.ORG

INFO@ARTICLE36.ORG

@ARTICLE36

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

KEY MESSAGES

- × Central concerns in the context of ‘autonomy in weapons’ are around limiting or removing the ‘human element’ or ‘human control’ from the process by which targets are identified and force is applied.
- × Whilst autonomous weapons apply force automatically on the basis of certain patterns of sensor data (pre-encoded target profiles), related concerns would be raised by decision-support tools that use machine analysis of sensor data to ‘propose’ targets to a human operator. For the purposes of this paper, we are calling these ‘sensor-based targeting systems’.
- × Given the similar (yet not exactly the same) challenges posed by both sensor-based weapons systems and sensor-based targeting systems, discussions on autonomy in weapons systems, including discussions on a future legally binding instrument, could consider encompassing both types of systems.
- × As food for thought, we suggest that autonomous weapons systems could be defined as systems that apply force automatically on the basis of a sensor-based targeting system - and that some rules then could be applied to the former and others to the latter.

INTRODUCTION

There is an ongoing international policy debate regard possible limits on ‘autonomy’ in the context of weapons systems. Formal discussions are taking place within a Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems set up under the framework of the Convention on Certain Conventional Weapons (CCW)¹. This discussion is currently concerned with clarifying and developing various aspects of a ‘normative and operational framework’ for addressing the challenges posed by autonomy in weapons.² Whilst discussions in the CCW are constrained by its consensus-based mode of work, the group of states supporting the development of a legally binding instrument to address these challenges is growing.³ This builds on the ICRC’s recommendation in 2021 that states “adopt new legally binding rules.”⁴

Notwithstanding the fact that the broad frame of the discussions on ‘autonomy in weapons’ has started taking more distinct shape, fundamental questions of what scope of systems are subject to future legal rules are yet to be clarified and determined. As food for thought in future discussions, this short paper suggests an approach that might be useful when considering the material scope of future legally binding instrument i.e. when determining what kind of systems and weapons systems should be covered by the future treaty.

THE BASIC STRUCTURE FOR REGULATION

Article 36 calls for the development and adoption of a legally binding international treaty as the necessary response to the risks that ‘autonomous weapons’ pose. As we have set out elsewhere,⁵ the scope of such a treaty should encompass a wide range of weapons systems that use sensors to determine when, where and to ‘what’ force will be applied without this being specifically decided by a human. Within this broad scope of regulation a line should be drawn between the weapons systems that should be subject to outright prohibition and the rest. The latter should be subject to positive obligations aimed at ensuring that ‘meaningful human control’ is exerted in their use, and thus rules of international humanitarian law (IHL) can be substantively applied.

The weapons systems that are in our opinion fundamentally unacceptable and thus should be prohibited include: (i) autonomous weapons systems that use target profiles that represent people, and (ii) autonomous weapons systems that cannot be meaningfully controlled by humans (e.g. systems that set their own goals, systems where their functioning cannot be limited in time and space, or opaque systems whose functioning cannot be sufficiently understood and explained). The remaining weapons systems should be subject to restrictions applicable during their development, approval for deployment and use so that ‘meaningful human control’ can be retained.⁶

‘AUTONOMOUS WEAPONS SYSTEMS’ AND ‘SENSOR-BASED TARGETING SYSTEMS’

Different stakeholders use different language when discussing autonomy in weapons systems. Despite differences of language, at the root of the issue are concerns relating to removing the ‘human element’ or ‘human control’ from the processes by which targets are identified and force is applied. Whilst the discussion has primarily been focused on systems by which force is applied automatically in response to a machine detecting certain patterns of sensor information (a target profile), some of the problems that arise in that process would also arise if, instead, the system were recommending (or proposing) that a human operator apply force to that same machine-detected object.

Thus, when deciding on what kind of systems should fall under the scope of future treaty, particular attention should be paid to the risks that emerge from autonomy in specific functions *within* the process of identifying a target object through to the physical application of force.

It seems reasonable that two types of systems are taken into consideration:

x autonomous weapons systems - understood as systems that, when in use, apply force automatically, at a time and place that is determined by matching sensor inputs from the environment against encoded profiles of intended target-types, without human assessment of those sensors inputs;

and

x sensor-based targeting systems - understood as systems designed to support the targeting process by detecting and proposing potential targets to human operators, where such systems operate by matching sensor inputs from the environment against encoded profiles of intended target-types, without human assessment of those sensors inputs.

When we compare these two kinds of systems, we can say that in case of autonomous weapons systems the process of...

$$\{sensing\} \rightarrow \{match\ sensor\ inputs=profile\} \rightarrow \{force\}$$

...is effectuated without a person being involved after the weapons system is activated.

On the other hand, in case of sensor-based targeting systems, the human is being given the option to “pull the trigger” or “push the fire button” in response to the outputs generated by the system:

$$\{sensing\} \rightarrow \{match\ sensor\ inputs=profile\} \rightarrow \underline{human\ judgement} \rightarrow \{force\}$$

Thus, whilst the system is proposing targets based on the automatic matching of sensor inputs against target profiles, there is a subsequent possibility for human judgement to be exerted. As such, it can be considered as a tool supporting the process of targeting rather than a weapon system *per se* (though specific types may, physically, be fully integrated within a weapons system).

As described here, both kinds of system operate in similar way - in fact they follow the same process up to a certain point. We have described the sensing and the analysis phases in the same terms. Only the last stage i.e. the stage of applying ‘force’, or engaging the target, is different. In the case of a sensor-based targeting system, a human pulls the trigger - in an autonomous weapons system, the trigger is pulled automatically.

Structurally, it is important to recognise that this opportunity for the exertion of human judgement means that additional information can be brought to bear on

that decision - information that may be generated from different sources, and subsequent to the point when the targeting system was activated. On the other hand, whilst additional information *may* be brought to bear and an additional process of human judgement *may allow* further human moral reflection, this does not have to happen in a meaningful way. There is the possibility that such systems produce a form of confirmation bias, where people come to assume the validity of the sensor-based targeting system’s proposals.

From this perspective it is easy to recognise that if a human user of a sensor-based targeting system simply pulled the trigger automatically (‘mindlessly’) in response to every target proposed by that system, then the effect would be the same as if it were an autonomous weapons system (i.e. no wider, or more recent information would be informing such decisions, no additional moral reflection would be occurring).

Therefore, although there is a significant structural distinction between the two processes resulting from the additional opportunity for human judgement, there are clearly certain issues that are at stake in both. Considering the abovementioned similarities between the two kinds of systems and the similar challenges that their use may pose, it seems justified to suggest that not only sensor-based weapons systems require controls and rules to be put in place to ensure their legitimate use but the same goes for the sensor-based targeting systems.

To illustrate it more specifically, below we present some general rules that might be taken into consideration when elaborating on the legal framework to address the challenges posed by ‘autonomy in weapons’ - and try to situate them in relation to both sensor-based weapons systems and sensor-based targeting systems.

TOWARDS AN INTERNATIONAL TREATY

Taking into consideration the direction in which the international discussion on ‘autonomy in weapons’ is moving, the question on how to frame the subject matter of future legally binding instrument that shall regulate the weapons featuring autonomy in their critical functions, becomes vital and practical. Thus, we find it useful to put forward some ideas that might feed the discussions on the subject matter and help ensuring that all relevant types of systems that raise concerns in the context of their autonomous functions undergo the necessary scrutiny under the new rules that shall be developed.

As a structural suggestion it would be possible to develop an approach such that:

- x ‘sensor based targeting systems’, are systems that propose targets by matching sensor inputs from the environment, and possibly other information, against target profiles.
- x ‘autonomous weapons systems’, are systems where force is applied automatically on the basis of a ‘sensor-based targeting system.’

Table: Comparison of rule applications

This table suggests a mapping of rules that Article 36 has proposed elsewhere against the different types of system under consideration in this paper. This mapping is simply provided as a 'food for thought' exercise, to illustrate that one approach to regulation might be to build on the interconnection of sensor-based targeting systems and autonomous weapons

		Sensor-based targeting systems	Autonomous weapons systems (sensor-based weapons systems)
		Systems that support a targeting process by detecting and 'proposing' potential targets based on machine analysis of sensors inputs. Such systems may operate by automatically matching sensors inputs from the environment, in combination with other data, against pre-encoded target profiles.	Systems that automatically apply force on the basis of a sensor-based targeting system, without human assessment of the individual proposals made by that system.
Policy building block	Rule/requirement		
Not killing people on the basis of sensors	Systems that target people should be prohibited because they undermine human dignity.	RULE WOULD NEED AUGMENTATION - such systems could be used to detect people, but it is a human that should be legally obliged to make a morally engaged and informed judgment as to whether a detected person should be attacked, and can be attacked in accordance with existing legal rules.	RULE APPLIES
Prohibiting systems that cannot be effectively controlled	It is not acceptable to use systems where the location and duration of their functioning cannot be appropriately limited.	RULE APPLIES - detecting people as <i>possible</i> targets would not be permissible outside of certain conflict circumstances, therefore some capacity for constraint must be required.	RULE APPLIES
Prohibiting systems that cannot be effectively controlled	It is not acceptable to use systems where the target profiles cannot be appropriately understood (systems functioning in a way that cannot be explained, systems that during their operation can change the profiles that they identify and propose as targets, or systems that generate their own target profiles should be prohibited)	RULE APPLIES - even if force is not being applied automatically it would be problematic if a system were proposing targets without the basis for such proposals being understood. The machine proposal would have a weight in the balance of considerations that it should not have.	RULE APPLIES
Prohibiting systems that cannot be effectively controlled	The users of a system must be able to provide a meaningful explanation of the external conditions and characteristics that will be identified and proposed as a target.	RULE APPLIES - as above, though these formulations point more directly to concerns regarding subsequent accountability.	RULE APPLIES
Ensuring meaningful human control over the systems that are not prohibited	Rules on the use of these systems should require users to appropriately control the location, duration and target specification, as well as other aspects of design and use.	RULES APPLY - in part these follow from the points above. <u>Autonomous weapons systems may require a more restrictive interpretation of certain rules.</u>	RULES APPLY
Ensuring meaningful human control over the systems that are not prohibited	System users need to have sufficient information about the context in which force would / may occur (e.g. about the presence of civilians and civilian objects in the vicinity/deployment area), to enable them to make a sufficient legal judgement on each attack.	RULE APPLIES - here it is an obligation to ensure sufficient information in relation to a specific proposed target, in a specific location, <u>subsequent</u> to the system's functioning.	RULE APPLIES - here it is an obligation to preemptively have sufficient information, about an area and time period of system functioning, in order to make a decision <u>prior</u> to system functioning.
Ensuring meaningful human control over the systems that are not prohibited	The users of a system must understand the nature and extent of force that a system will exert in any application of force.	RULE DOES NOT APPLY - this system does not apply force.	RULE APPLIES
Ensuring meaningful human control over the systems that are not prohibited	The users of a system must limit the number of applications of force that a system can undertake within the context of an attack, such that they can make appropriate judgements about the use of that system in that attack.	RULE DOES NOT APPLY - this system does not apply force.	RULE APPLIES

ENDNOTES

- 1 The Convention on prohibitions or restrictions on the use of Certain Conventional Weapons which may be deemed to be excessively injurious or to have indiscriminate effects adopted on 10 October 1980.
- 2 States contributions on possible consensus recommendations in relation to the clarification, consideration and development of aspects of the normative and operational framework are available here: https://meetings.unoda.org/section/ccw-gge-2021_documents_14090_documents_14570/
- 3 See e.g. the contributions submitted by Brazil, Chile and Mexico: <https://documents.unoda.org/wp-content/uploads/2021/06/Brazil-Chile-Mexico.pdf> and the contributions submitted by the Republic of Costa Rica, the Republic of Panama, the Republic of Peru, the Republic of the Philippines, the Republic of Sierra Leone and the Eastern Republic of Uruguay: <https://documents.unoda.org/wp-content/uploads/2021/06/Costa-Rica-Panama-Peru-the-Philippines-Sierra-Leone-and-Uruguay.pdf>
- 4 ICRC Position on Autonomous Weapon Systems and Background Paper, 12 May 2021, p.2, <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>
- 5 See "Regulating Autonomy in Weapons", Article 36, 2020, <https://article36.org/wp-content/uploads/2020/10/Regulating-autonomy-leaflet.pdf>; Elizabeth Minor, "Autonomous weapons: building towards international regulation from contributions to the CCW", Article 36, 2020, <https://article36.org/wp-content/uploads/2020/10/CCW-commentaries-analysis.pdf>.
- 6 See "Regulating Autonomy in Weapons", Article 36, 2020, <https://article36.org/wp-content/uploads/2020/10/Regulating-autonomy-leaflet.pdf>.

Acknowledgements:

Lead author on this paper was

Anna Turek
Attorney-at-law, member of the Human Rights
Section at Warsaw Bar Association
annaturek.3@gmail.com

Opinions expressed herein reflect only her
personal views.