Advisory Note on Autonomous Weapon Systems that Target Humans

July 2021

As set out in our recommendations, the Campaign to Stop Killer Robots[1] urges states to prohibit autonomous weapon systems that target humans as one part of a legal framework of prohibitions and regulations.[2] This note sets out the basis for such a prohibition including consideration of the moral, ethical, legal and practical dangers they would pose.

Autonomous weapon systems that target humans

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. We consider these systems unacceptable and recommend a prohibition against them.

Of course, objects such as vehicles may contain humans. Autonomous weapon systems that target such objects must still be subject to meaningful human control and be used in accordance with existing legal rules including international humanitarian law and human rights law. However, such autonomous weapon systems would not **necessarily** be subject to a prohibition because the target profiles that activate the weapon are not specifically designed to target people.[3]

The Campaign to Stop Killer Robots recommends that autonomous weapon systems that target human beings be prohibited, **regardless of whether they operate under meaningful human control.**

Moral, ethical, legal and practical risks

Autonomous weapon systems that target humans, whether civilian or combatant, would dehumanise people, harming them on the basis of a processing of sensor inputs - converting people into data, sensed and sorted by a machine. In killing or wounding people based on such abstractions, these systems would offend against human dignity. To allow sensors and software to determine who lives and who dies should be considered morally unacceptable.

In addition to this fundamental moral objection, using sensor-based systems to automatically target people 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.

---

[1] The Campaign to Stop Killer Robots is a coalition of 180 NGOs across 68 countries

[2] The overall structure for regulating autonomous weapon systems should include (a) a general obligation to maintain meaningful human control (b) prohibitions on (i) systems that cannot be meaningfully controlled and (ii) systems that target humans and (c) positive obligations to ensure meaningful human control over the broad range of autonomous weapon systems. For further details, see our Recommendations on the Normative and Operational Framework

[3] Such autonomous weapon systems should still be subject to a prohibition if they cannot be meaningfully controlled.

(a) Targeting a certain person or certain groups of people

If it is claimed that systems can distinguish between people to whom force should be applied and those who must be protected, then acute concerns arise regarding the legal requirements for distinction and for the protection of certain classes of persons. Under International Humanitarian Law (IHL), people are targetable only on a contextual, case-by-case basis. Claiming that machines can make such determinations undermines the law. Additional serious problems relating to prejudice and bias would arise if systems used target profiles to identify 'targetable people' based on certain characteristics such as race, gender or age and historically marginalized groups may face higher rates of error.[4]

Furthermore, systems that are particularly complex or opaque would make it difficult or perhaps impossible to offer a meaningful explanation of why certain people were targeted in certain circumstances.

(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 moral, legal and practical problems would be most straightforwardly addressed through a prohibition on autonomous weapon systems that target humans - **amongst other legal prohibitions and regulations.** There is strong public opinion against allowing autonomous weapon systems to target people. This is reflected in an appeal from thousands of AI and tech experts, an international interfaith statement and in a global survey - where 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 prohibition of sensor-based weapon systems that target people, as opposed to other objects, has precedent in the prohibition of anti-personnel landmines (and even the CCW itself recognises that anti-personnel mines should be subject to specific, more stringent rules than other mines.) Significantly, a prohibition on autonomous weapons that target people would not entail changes in practice for most militaries, as sensor-based systems targeting people without human decision-making are currently not in extensive use.

Thus, as a fundamental moral position, and on the basis of a precautionary orientation to protect existing law and avoid societal harms, states should call for a prohibition on autonomous weapons systems that would target people, as one component of a legal structure of prohibitions and regulations.

[4] See 'Autonomous weapons systems: an analysis from human rights, humanitarian and ethical artificial intelligence perspectives', Wanda Muñoz, SEHLAC