



# The Need for and Elements of a New Treaty on Fully Autonomous Weapons

June 2020\*

The rapid evolution of autonomous technology threatens to strip humans of their traditional role in the use of force. Fully autonomous weapons, in particular, would select and engage targets without meaningful human control. Due in large part to their lack of human control, these systems, also known as lethal autonomous weapons systems or “killer robots,” raise a host of legal and ethical concerns.

States parties to the Convention on Conventional Weapons (CCW) have held eight in-depth meetings on lethal autonomous weapons systems since 2014. They have examined the extensive challenges raised by the systems and recognized the importance of retaining human control over the use of force. Progress toward an appropriate multilateral solution, however, has been slow. If states do not shift soon from abstract talk to treaty negotiations, the development of technology will outpace international diplomacy.

Approaching the topic from a legal perspective, this chapter argues that fully autonomous weapons cross the threshold of acceptability and should be banned by a new international treaty. The chapter first examines the concerns raised by fully autonomous weapons, particularly under international humanitarian law. It then explains why a legally binding instrument best addresses those concerns. Finally, it proposes key elements of a new treaty to maintain meaningful human control over the use of force and prohibit weapons systems that operate without it.

---

\* This chapter is part of a forthcoming publication by Fundação Alexandre de Gusmão, based on a presentation at the Rio Seminar on Autonomous Weapons Systems, February 20, 2020. The author, Bonnie Docherty, is a lecturer on law and the associate director of armed conflict and civilian protection at Harvard Law School’s International Human Rights Clinic. She is also a senior researcher in the Arms Division of Human Rights Watch, which has coordinated the Campaign to Stop Killer Robots since its inception in 2012. Docherty has participated in every Convention on Conventional Weapons meeting about lethal autonomous weapons systems and has published extensively on the topic. See Human Rights Watch and the Harvard Law School International Human Rights Clinic (IHRC), “Reviewing the Record: Reports on Killer Robots,” [http://hrp.law.harvard.edu/wp-content/uploads/2018/08/Killer\\_Robots\\_Handout.pdf](http://hrp.law.harvard.edu/wp-content/uploads/2018/08/Killer_Robots_Handout.pdf) (accessed May 22, 2020).

## The Problems Posed by Fully Autonomous Weapons

Fully autonomous weapons would present significant hurdles to compliance with international humanitarian law's fundamental rules of distinction and proportionality.<sup>1</sup> In today's armed conflicts, combatants often seek to blend in with the civilian population. They hide in civilian areas and wear civilian clothes. As a result, the ability to distinguish combatants from civilians or those *hors de combat* often requires gauging an individual's intentions based on subtle behavioral cues, such as body language, gestures, and tone of voice. Humans, who can relate to other people, can better interpret those cues than inanimate machines.<sup>2</sup>

Fully autonomous weapons would find it even more difficult to weigh the proportionality of an attack. The proportionality test requires determining whether expected civilian harm outweighs anticipated military advantage on a case-by-case basis in a rapidly changing environment. Evaluating the proportionality of an attack involves more than a quantitative calculation. Commanders apply human judgment, informed by legal and moral norms and personal experience, to the specific situation. Whether the human judgment necessary to assess proportionality could ever be replicated in a machine is doubtful. Furthermore, robots could not be programmed in advance to deal with the infinite number of unexpected situations they might encounter on the battlefield.<sup>3</sup>

The use of fully autonomous weapons also risks creating a serious accountability gap.<sup>4</sup> International humanitarian law requires that individuals be held legally responsible for war crimes and grave breaches of the Geneva Conventions. Military commanders or operators could be found guilty if they deployed a fully autonomous weapon with the intent to commit a crime. It would, however, be legally challenging and arguably unfair to hold an operator responsible for the unforeseeable actions of an autonomous robot.

---

<sup>1</sup> Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), adopted June 8, 1977, 1125 U.N.T.S. 3, entered into force December 7, 1978, arts. 48 and 51(4-5).

<sup>2</sup> Human Rights Watch and IHRC, *Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban* (December 2016), [https://www.hrw.org/sites/default/files/report\\_pdf/arms1216\\_web.pdf](https://www.hrw.org/sites/default/files/report_pdf/arms1216_web.pdf) (accessed May 21, 2020), p. 5.

<sup>3</sup> *Ibid.*, pp. 5-8.

<sup>4</sup> See generally Human Rights Watch and IHRC, *Mind the Gap: The Lack of Accountability for Killer Robots* (April 2015), [https://www.hrw.org/sites/default/files/reports/arms0415\\_ForUpload\\_o.pdf](https://www.hrw.org/sites/default/files/reports/arms0415_ForUpload_o.pdf) (accessed May 20, 2020).

Finally, fully autonomous weapons contravene the Martens Clause, a provision that appears in numerous international humanitarian law treaties.<sup>5</sup> The clause states that if there is no specific law on a topic, civilians are still protected by the principles of humanity and dictates of public conscience.<sup>6</sup> Fully autonomous weapons would undermine the principles of humanity because of their inability to show compassion or respect human dignity.<sup>7</sup> Widespread opposition to fully autonomous weapons among faith leaders, scientists, tech workers, civil society organizations, the public, and more indicate that this emerging technology also runs counter to the dictates of public conscience.<sup>8</sup>

Fully autonomous weapons pose numerous other threats that go far beyond concerns over compliance with international humanitarian law. For many, delegating life-and-death decisions to machines would cross a moral red line.<sup>9</sup> The use of fully autonomous weapons, including in law enforcement operations, would undermine the rights to life, remedy, and dignity.<sup>10</sup> Development and production of these machines could trigger an

---

<sup>5</sup> See generally Human Rights Watch and IHRC, *Heed the Call: A Moral and Legal Imperative to Ban Killer Robots* (August 2018), [https://www.hrw.org/sites/default/files/report\\_pdf/arms0818\\_web.pdf](https://www.hrw.org/sites/default/files/report_pdf/arms0818_web.pdf) (accessed May 20, 2020).

<sup>6</sup> See, for example, Convention (II) with Respect to the Laws and Customs of War on Land and its Annex: Regulations concerning the Laws and Customs of War on Land, The Hague, adopted July 29, 1899, entered into force September 4, 1900, pmbL., para. 8; Protocol I, art. 1(2).

<sup>7</sup> Human Rights Watch and IHRC, *Heed the Call*, pp. 19-27.

<sup>8</sup> See, for example, PAX, “Religious Leaders Call for a Ban on Killer Robots,” November 12, 2014, <https://www.paxforpeace.nl/stay-informed/news/religious-leaders-call-for-a-ban-on-killer-robots>; “Autonomous Weapons: An Open Letter from AI & Robotics Researchers,” opened for signature July 28, 2015, <https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1> (signed, as of May 2020, by 4,502 AI and robotics researchers and 26,215 others); Scott Shane and Daisuke Wakabayashi, “The Business of War: Google Employees Protest Work for the Pentagon,” *New York Times*, April 4, 2018, <https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html?partner=IFTTT>; Campaign to Stop Killer Robots, “Learn: The Threat of Fully Autonomous Weapons,” <https://www.stopkillerrobots.org/learn/>; Ipsos, “Six in Ten (61%) Respondents across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems,” January 21, 2019, <https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons> (all accessed May 21, 2020). See also Human Rights Watch and IHRC, *Heed the Call*, pp. 28-43.

<sup>9</sup> UN Human Rights Council, Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns, “Lethal Autonomous Robotics,” [http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47\\_en.pdf](http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf) (accessed May 21, 2020), p. 17 (writing, “Machines lack morality and mortality, and should as a result not have life and death powers over humans”).

<sup>10</sup> See generally Human Rights Watch and IHRC, *Shaking the Foundations: The Human Rights Implications of Killer Robots* (May 2014), [https://www.hrw.org/sites/default/files/reports/arms0514\\_ForUpload\\_0.pdf](https://www.hrw.org/sites/default/files/reports/arms0514_ForUpload_0.pdf) (accessed May 20, 2020). See also Heyns, “Lethal Autonomous Robotics,” pp. 6 (on the right to life: “the introduction of such powerful yet controversial new weapons systems has the potential to pose new threats to the right to life”), 15 (on the right to remedy: “If the nature of a weapon renders responsibility for its consequences impossible, its use should be considered unethical and unlawful as an abhorrent weapon”), and 20 (on dignity: “there is widespread concern that allowing [fully autonomous weapons] to kill people may denigrate the value of life itself”).

arms race, and the systems could proliferate to irresponsible states and non-state armed groups.<sup>11</sup> Even if new technology could address some of the international humanitarian law problems discussed above, it would not resolve many of these other concerns.

## The Need for a Legally Binding Instrument

The unacceptable risks posed by fully autonomous weapons necessitate creation of a new legally binding instrument. It could take the form of a stand-alone treaty or a protocol to the Convention on Conventional Weapons. Existing international law, including international humanitarian law, is insufficient in this context because its fundamental rules were designed to be implemented by humans not machines. At the time states negotiated the additional protocols to the Geneva Conventions, they could not have envisioned full autonomy in technology. Therefore, while CCW states parties have agreed that international humanitarian law applies to this new technology, there are debates about how it does.<sup>12</sup>

A new treaty would clarify and strengthen existing international humanitarian law. It would establish clear international rules to address the specific problem of weapons systems that operate outside of meaningful human control. In so doing, the instrument would fill the legal gap highlighted by the Martens Clause, help eliminate disputes about interpretation, promote consistency of interpretation and implementation, and facilitate compliance and enforcement.<sup>13</sup>

The treaty could also go beyond the scope of current international humanitarian law. While the relevant provisions of international humanitarian law focus on the use of weapons, a new treaty could address development, production, and use. In addition, it could apply to

---

<sup>11</sup> “Autonomous Weapons: An Open Letter from AI & Robotics Researchers”; Human Rights Watch and IHRC, *Making the Case*, pp. 29-30.

<sup>12</sup> The applicability of international humanitarian law to lethal autonomous weapons systems is the first of 11 guiding principles adopted by CCW states parties. “Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems,” CCW/GGE.1/2018/3, October 23, 2018, [https://www.unog.ch/80256EDD006B8954/\(httpAssets\)/20092911F6495FA7C125830E003F9A5B/\\$file/CCW\\_GGE.1\\_2018\\_3\\_final.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/20092911F6495FA7C125830E003F9A5B/$file/CCW_GGE.1_2018_3_final.pdf) (accessed May 21, 2020), para. 26(a).

<sup>13</sup> Campaign to Stop Killer Robots, “Key Elements of a Treaty on Fully Autonomous Weapons: Frequently Asked Questions,” February 2020, <https://www.stopkillerrobots.org/wp-content/uploads/2020/03/FAQ-Treaty-Elements.pdf> (accessed May 21, 2020), p. 2.

the use of fully autonomous weapons in both law enforcement operations as well as situations of armed conflict.<sup>14</sup>

A legally binding instrument is preferable to the “normative and operational framework” that the CCW states parties agreed to develop in 2020 and 2021.<sup>15</sup> The phrase “normative and operational framework” is intentionally vague, and thus has created uncertainty about what states should be working toward. While the term could encompass a legally binding CCW protocol, it could also refer to political commitments or voluntary best practices, which would be not be enough to preempt what has been called the “third revolution in warfare.”<sup>16</sup> Whether adopted under the auspices of CCW or in another forum, a legally binding instrument would bind states parties to clear obligations. Past experience shows that the stigma it would create could also influence states not party and non-state armed groups.

## The Elements of a New Treaty

CCW states parties have discussed the problems of fully autonomous weapons and the adequacy of international humanitarian law since 2014. It is now time to move forward and determine the specifics of an effective response. This chapter will lay out key elements of a proposed treaty, which were drafted by the International Human Rights Clinic at Harvard Law School and adopted by the Campaign to Stop Killer Robots in 2019.<sup>17</sup>

The proposal outlined below does not constitute specific treaty language. States will determine the details of content and language over the course of formal negotiations. Instead, the proposal highlights elements that a final treaty should contain in order to effectively address concerns that many states, international organizations, and civil society have identified. The elements include the treaty’s scope, the underlying concept of meaningful human control, and core obligations.

---

<sup>14</sup> Ibid.

<sup>15</sup> CCW Meeting of High Contracting Parties, “Final Report,” CCW/MSP/2019/9, December 13, 2019, <https://undocs.org/CCW/MSP/2019/9> (accessed May 21, 2020), para. 31.

<sup>16</sup> “Autonomous Weapons: An Open Letter from AI & Robotics Researchers.”

<sup>17</sup> Campaign to Stop Killer Robots, “Key Elements of a Treaty on Fully Autonomous Weapons,” November 2019, <https://www.stopkillerrobots.org/wp-content/uploads/2020/04/Key-Elements-of-a-Treaty-on-Fully-Autonomous-WeaponsvAccessible.pdf> (accessed May 21, 2020). See also Campaign to Stop Killer Robots, “Key Elements of a Treaty on Fully Autonomous Weapons: Frequently Asked Questions.”

## Scope

The proposal for a new treaty recommends a broad scope of application. The treaty should apply to any weapon system that selects and engages targets based on sensor processing, rather than human input.<sup>18</sup> The breadth of scope aims to ensure that all systems in that category—whether current or future—are assessed, and that problematic systems do not escape regulation. The prohibitions and restrictions, which are detailed below, however, are future looking and focus on fully autonomous weapons.

## Meaningful Human Control

The concept of meaningful human control is crucial to the new treaty because the moral, legal, and accountability problems associated with fully autonomous weapons are largely attributable to the lack of such control.<sup>19</sup> Recognizing these risks, most states have embraced the principle that humans must play a role in the use of force.<sup>20</sup> While they have used different terminology, many states and experts prefer the term “meaningful human control.” “Control” is stronger than alternatives such as “intervention” and “judgment” and is broad enough to encompass both of them; it is also a familiar concept in international law.<sup>21</sup> “Meaningful” ensures that control rises to a significant level.<sup>22</sup>

---

<sup>18</sup> Article 36, “Autonomy in Weapons System: Mapping a Structure for Regulation through Specific Policy Questions,” November 2019, <http://www.article36.org/wp-content/uploads/2019/11/regulation-structure.pdf> (accessed May 21, 2020), p. 1.

<sup>19</sup> Human Rights Watch and IHRC, “Killer Robots and the Concept of Meaningful Human Control,” April 2016, [https://www.hrw.org/sites/default/files/supporting\\_resources/robots\\_meaningful\\_human\\_control\\_final.pdf](https://www.hrw.org/sites/default/files/supporting_resources/robots_meaningful_human_control_final.pdf) (accessed May 21, 2020), pp. 2-6.

<sup>20</sup> Ray Acheson, “It’s Time to Exercise Human Control over the CCW,” *Reaching Critical Will’s CCW Report*, vol. 7, no. 2, March 27, 2019, <https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2019/gge/reports/CCWR7.2.pdf> (accessed May 21, 2020), p. 2 (reporting that “[o]nce discussions got under way, it became clear that the majority of governments still agree human control is necessary over critical functions of weapon systems”).

<sup>21</sup> Campaign to Stop Killer Robots, “Key Elements of a Treaty on Fully Autonomous Weapons: Frequently Asked Questions,” p. 5.

<sup>22</sup> According to Article 36, “The term ‘meaningful’ can be argued to be preferable because it is broad, it is general rather than context specific (e.g. appropriate) [and] derives from an overarching principle rather being outcome driven (e.g. effective, sufficient).” Article 36, “Key Elements of Meaningful Human Control,” April 2016, <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> (accessed May 21, 2020), p. 2.

States, international organizations, nongovernmental organizations, and independent experts have identified numerous components of meaningful human control.<sup>23</sup> This chapter distills those components into three categories:

- Decision-making components give humans the information and ability to make decisions about whether the use of force complies with law and ethics. For example, a human operator should have: an understanding of the operational environment; an understanding of how the system functions, such as what it might identify as target; and sufficient time for deliberation.
- Technological components are embedded features of a weapon system that enhance meaningful human control. Technological components include, for example, predictability and reliability, the ability of the system to relay information to a human operator, and the ability of a human to intervene after activation of the system.
- Operational components limit when and where a weapon system can operate and what it can target. Factors that could be constrained include the time between a human’s legal assessment and a system’s application of force, the duration of a system’s operation, and the nature and size of the geographic area of operation.<sup>24</sup>

None of these components are independently sufficient, but they each increase the meaningfulness of control, and they often work in tandem. The above list may not be exhaustive; further analysis of existing and emerging technologies may reveal others. Regardless, a new legally binding instrument should incorporate such components as prerequisites for meaningful human control.

---

<sup>23</sup> See, for example, Allison Pytlak and Katrin Geyer, “News in Brief,” *Reaching Critical Will’s CCW Report*, vol. 7, no. 2, March 27, 2019, pp. 10-12 (summarizing states’ views on control from one CCW session); International Committee of the Red Cross, “Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control,” August 2019, <https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control>; International Committee for Robot Arms Control, “What Makes Human Control over Weapons Systems ‘Meaningful?’” August 2019, [https://www.icrac.net/wp-content/uploads/2019/08/Amoroso-Tamburrini\\_Human-Control\\_ICRAC-WP4.pdf](https://www.icrac.net/wp-content/uploads/2019/08/Amoroso-Tamburrini_Human-Control_ICRAC-WP4.pdf); International Panel on the Regulation of Autonomous Weapons (iPRAW), *Focus on Human Control* (August 2019), [https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09\\_iPRAW\\_HumanControl.pdf](https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-09_iPRAW_HumanControl.pdf) (all accessed May 21, 2020).

<sup>24</sup> Campaign to Stop Killer Robots, “Key Elements of a Treaty on Fully Autonomous Weapons,” pp. 3-4.

## Core Obligations

The heart of a legally binding instrument on fully autonomous weapons should consist of a general obligation combined with prohibitions and positive obligations.<sup>25</sup>

### *General Obligation*

The treaty should include a general obligation for states to maintain meaningful human control over the use of force. This obligation establishes a principle to guide interpretation of the rest of the treaty. Its generality is designed to avoid loopholes that could arise in the other, more specific obligations. The focus on conduct (“use of force”) rather than specific technology future proofs the treaty’s obligations because it is impossible to envision all technology that could prove problematic. The reference to use of force also allows for application to both situations of armed conflict and law enforcement operations.

### *Prohibitions*

The second category of obligations is prohibitions on weapons systems that select and engage targets and by their nature—rather than by the manner of their use—pose fundamental moral or legal problems. The new treaty should prohibit the development, production, and use of systems that are inherently unacceptable. The clarity of such prohibitions facilitates monitoring, compliance, and enforcement. Their absolute nature increases stigma, which can in turn influence states not party and non-state actors.

The proposed treaty contains two subcategories of prohibitions. First, the prohibitions cover systems that always select and engage targets without meaningful human control. Such systems might operate, for example, through machine learning and thus be too complex for humans to understand and control. Second, the prohibitions could extend to other systems that select and engage targets and are by their nature problematic: specifically, systems that use certain types of data—such as weight, heat, or sound—to represent people, regardless of whether they are combatants. Killing or injuring humans based on such data would undermine human dignity and dehumanize violence. In

---

<sup>25</sup> These obligations are drawn from the Campaign to Stop Killer Robots, “Key Elements of a Treaty on Fully Autonomous Weapons.”



addition, whether by design or due to algorithmic bias, they may rely on discriminatory indicators to choose targets.<sup>26</sup>

### *Positive Obligations*

The third category of obligations encompasses positive obligations to ensure meaningful human control is maintained over all other systems that select and engage targets. These systems may not be prohibited under the treaty as inherently problematic, but they might have the potential to be used without meaningful human control. The positive obligations apply to all systems that select and engage targets based on sensor processing, and they establish requirements to ensure that human control over these systems is meaningful. The components of meaningful human control discussed above can help determine the criteria necessary to ensure systems are used only with such control.

## Other Elements

The elements outlined above are not the only elements of a new legally binding instrument. While beyond the scope of this chapter, other important elements include:

- A preamble, which would articulate the treaty's purpose;
- Reporting requirements to promote transparency and facilitate monitoring;
- Verification and cooperative compliance measures to enforce the treaty's provisions;
- A framework for regular meetings of states parties to review the status and operation of the treaty, identify implementation gaps, and set goals for the future;
- Requirements to adopt national implementation measures; and
- The threshold for entry into force.<sup>27</sup>

## Conclusion

After six years of CCW discussions, states should actively consider the elements of a new treaty and pursue negotiations to realize them. In theory, negotiations could lead to a new

---

<sup>26</sup> For further discussion of the second subcategory of prohibitions, see Article 36, "Targeting People: Key Issues in the Regulation of Autonomous Weapons Systems," November 2019, <http://www.article36.org/wp-content/uploads/2019/11/targeting-people.pdf> (accessed May 21, 2020).

<sup>27</sup> Campaign to Stop Killer Robots, "Key Elements of a Treaty on Fully Autonomous Weapons," p. 9.

CCW protocol, but certain states have taken advantage of the CCW's consensus rules to block progress. Therefore, it is time to consider an alternative forum. States could start an independent process of the kind used to create the Mine Ban Treaty or the Convention on Cluster Munitions, or they could adopt a treaty under the auspices of the UN General Assembly as was done for the Arms Trade Treaty and the Treaty on the Prohibition on Nuclear Weapons. Ultimately, states should pursue the most efficient path to the most effective treaty that preempts the dangers posed by fully autonomous weapons.