Fully Autonomous Weapons:
Questions and Answers

Human Rights Watch and
Harvard Law School's International Human Rights Clinic
October 2013

Advances in artificial intelligence (AI) and other technologies will soon make possible the development of fully autonomous weapons, which would revolutionize the way wars are fought. These weapons, unlike the current generation of armed drones, would be able to select and engage targets without human intervention. Military officials in the United States and other technologically advanced countries generally say that they prefer to see humans retain some level of supervision over decisions to use lethal force, and the US Defense Department has issued a policy directive embracing that principle for the time being.

But the temptation will grow to acquire fully autonomous weapons, also known as "lethal autonomous robotics" or "killer robots." If one nation acquires these weapons, others may feel they have to follow suit to avoid falling behind in a robotic arms race. Furthermore, the potential deployment and use of such weapons raises serious concerns about protecting civilians during armed conflict. Because of these concerns, fully autonomous weapons should be prohibited before it is too late to change course. Nations should agree that any decision to use lethal force against a human being should be made by a human being.

In our November 2012 report, *Losing Humanity: The Case against Killer Robots*, Human Rights Watch and Harvard Law School's International Human Rights Clinic (IHRC) discussed the move toward full autonomy in weapons systems and analyzed the risks the technology could pose to civilians. We also called on countries to prohibit fully

autonomous weapons through an internationally legally binding instrument and to adopt national laws and policies on the subject. This Question and Answer document summarizes, clarifies, and expands on some of the issues discussed in *Losing Humanity*. It examines the legal problems posed by fully autonomous weapons and then elaborates on why banning these weapons is the best approach for dealing with this emerging means of war.

Why are fully autonomous weapons a pressing issue?

While fully autonomous weapons technology does not exist yet, developments in that direction make it a pressing issue. The 2012 US Defense Department directive on autonomy mandates keeping humans in the loop for any decision about the use of lethal force for up to 10 years. Other US military documents, however, have indicated a long-term interest in full autonomy.

For example, a 2011 US roadmap specifically for ground systems stated, "There is an ongoing push to increase UGV [unmanned ground vehicle] autonomy, with a current goal of 'supervised autonomy,' but with an ultimate goal of full autonomy." A US Air Force planning document from 2009 said, "[A]dvances in AI will enable systems to make combat decisions and act within legal and policy constraints without necessarily requiring human input."

The United States has also been developing precursors to fully autonomous systems. The US X-47B aircraft is being designed to take off, land, and refuel on its own power and would have weapons bays that could be used to make later models serve a combat function. While this particular aircraft may not necessarily be given the power to determine when to fire, it reflects the move toward increased autonomy. Other countries pursuing ever-greater autonomy for weapons include China, Israel, Russia, South Korea, and the United Kingdom. As these nations go down that road, more may choose to follow.

What are the potential benefits of fully autonomous weapons?

A range of potential benefits touted by proponents have motivated some countries to pursue increasingly autonomous technology. Proponents contend that fully autonomous weapons could decrease the need for soldiers on the battlefield and thereby save military lives. These weapons could have the ability to detect and attack targets with greater speed

and precision than weapons directed by human beings. Pain, hunger, exhaustion, the instinct for self-defense, and emotions such as fear and anger would not influence fully autonomous weapons' determinations about when to use lethal force. These characteristics could entice high-tech militaries to deploy fully autonomous weapons despite their humanitarian drawbacks.

If fully autonomous weapons could have some advantages, why should they be prohibited?

The potential advantages of fully autonomous weapons would be offset by the lack of human control over the weapons. As elaborated on below, fully autonomous weapons would face challenges in being able to comply with the complex and subjective rules of international humanitarian law, which require human understanding and judgment.

In addition, while fully autonomous weapons would not share the emotional weaknesses of human soldiers, they would at the same time be bereft of other emotions, most notably compassion. Compassion can deter combatants from killing civilians, even in conflicts in which there is little regard for international humanitarian law and commanders order troops to target civilians.

Saving soldiers' lives is a laudable goal, but the concerns raised about the use of fully autonomous weapons suggest that with the use of these weapons, the danger for civilians would increase, running the risk of shifting the burden of conflict onto civilians.

Finally, it would appear that many of the potential benefits of fully autonomous weapons— such as the ability to process large amounts of data swiftly, to avoid actions driven by fear, and to reduce military casualties—could arguably be attained through the use of semi-autonomous weapons that are remotely controlled by humans. Humans bring judgment and compassion into decisions about the use of lethal force.

Could fully autonomous weapons comply with the requirements of international humanitarian law to protect civilians in armed conflict?

There are serious concerns about whether fully autonomous weapons could comply fully with important principles of international humanitarian law, a shortcoming that threatens

legal protections for civilians. Distinguishing between combatants, who may be targeted, and civilians, who may not be, is a core requirement of international humanitarian law.

There is no certainty that fully autonomous weapons would have the capacity to make such distinctions reliably. That is especially the case in the increasingly common scenarios of contemporary warfare, in which combatants do not identify themselves by uniforms or insignia. When there are no visible clues in a close combat situation, assessing an individual's intentions is key to determining a potential target's status and level of threat.

An important way to recognize intentions is to relate to the person's emotional state, which another person can do. But it would be difficult, and perhaps impossible, to program a robot with the innately human qualities crucial to assessing an individual's intentions.

It would also be difficult to program fully autonomous weapons to carry out the proportionality test, for prohibiting attacks in which expected civilian harm outweighs anticipated military advantage. Robots could not be programmed to determine that balance in every situation because there are an infinite number of possible situations. In addition, international humanitarian law depends on human judgment to make subjective decisions about the proportionality of attacks. There are serious doubts about whether fully autonomous weapons could exercise comparable judgment to assess proportionality in complex and evolving situations.

Are there other concerns under international humanitarian law?

The proposed use of fully autonomous weapons also potentially relates to an element of international law known as the Martens Clause, which sets limits on the conduct of war even when treaty law does not apply.

This provision of international humanitarian law, as articulated in Additional Protocol I to the Geneva Conventions, states:

> In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the

principles of international law derived from established custom, from the principles of humanity and from dictates of public conscience.[1]

States should take evolving public perspectives into account when determining whether fully autonomous weapons meet the dictates of public conscience.

For many people the prospect of fully autonomous weapons is disturbing. In discussions with government and military officials, scientists, and the general public, for example, Human Rights Watch has encountered tremendous discomfort with the idea of allowing military robots to determine on their own if and when to use lethal force against a human being.

A June 2013 national representative survey of 1,000 Americans found that, of those with a view, two-thirds came out against fully autonomous weapons: 68 percent opposed the move toward these weapons (48 percent strongly), while 32 percent favored their development. Interestingly, active duty military personnel were among the strongest objectors—73 percent expressed opposition to fully autonomous weapons.[2] These kinds of reactions to fully autonomous weapons raise serious concerns under the Martens Clause.

Is accountability an issue for fully autonomous weapons?

Accountability for violations of international humanitarian law is important for two reasons. First, the understanding that people will be held responsible for their actions can deter them from committing war crimes or being negligent. Second, accountability for unlawful

---

[1] 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, art. 1(2).

[2] Charli Carpenter, "US Public Opinion on Autonomous Weapons," June 19, 2013, http://www.whiteoliphaunt.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf (accessed June 21, 2013). Many who responded "not sure" preferred a precautionary approach "in the absence of information." Charli Carpenter, "How Do Americans Feel about Fully Autonomous Weapons?" *Duck of Minerva*, June 19, 2013, http://www.whiteoliphaunt.com/duckofminerva/2013/06/how-do-americans-feel-about-fully-autonomous-weapons.html (accessed June 21, 2013). These figures are based on a nationally representative online poll of 1,000 Americans conducted by Yougov.com. Respondents were an invited group of internet users (YouGov Panel) matched and weighted on gender, age, race, income, region, education, party identification, voter registration, ideology, political interest and military status. The margin of error for results is +/- 3.6%. A discussion of the sampling methods, limitations and accuracy can be found at: http://yougov.co.uk/publicopinion/methodology/.

acts dignifies victims by giving them recognition that they were wronged and satisfaction that someone was punished for inflicting the harm they experienced.

Holding a human responsible for the actions of a robot that is acting autonomously could prove difficult. The resulting accountability gap would undermine this valuable tool for protecting civilians. The lack of accountability for fully autonomous weapons could also drive some militaries to develop and acquire these weapons instead of others for which they could more easily be held accountable.

How would a new legal instrument for fully autonomous weapons supplement existing international humanitarian law?

Fully autonomous weapons would be a new category of weapons that could pose serious risks to civilians. International humanitarian law needs to be clarified and strengthened to address that issue.

While international humanitarian law already sets limits on problematic weapons and their use, responsible governments have found it necessary to supplement that legal framework for several weapons that significantly threaten civilians, including antipersonnel mines, biological weapons, chemical weapons, and cluster munitions. Fully autonomous weapons would have the potential to raise a comparable level of humanitarian concern.

An international prohibition would eliminate questions about the legality of these controversial weapons by standardizing rules across countries and overriding calls for case-by-case determinations. A stand-alone instrument could also address aspects of proliferation such as production and transfer, which traditional international humanitarian law does not.

Supplementing existing international humanitarian law with a new convention prohibiting fully autonomous weapons would also facilitate weapons reviews. Article 36 of Additional Protocol I to the Geneva Conventions requires a state party "[i]n the study, development, acquisition or adoption of a new weapon, means or method of war...to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."

Whether countries that are not party to Additional Protocol I consider a review process a customary international law obligation, many of them conduct one. Without a treaty banning these weapons, reviews might turn on case-by-case determinations. A strong convention, however, would help guide weapons reviews and standardize findings. It would also make the illegality of fully autonomous weapons clear even for countries that do not conduct reviews of new or modified weapons.

Why pursue a ban rather than regulation of fully autonomous weapons?

A prohibition would maximize protection for civilians in conflict because it would be more comprehensive than regulation. A ban would also be more effective as it would be clearer and thus simpler to enforce. Regulations, by contrast, would allow for the existence and potential misuse of fully autonomous weapons, be more difficult to enforce, and lead to proliferation.

If fully autonomous weapons came into existence under a regulatory regime, they would be vulnerable to misuse. Even if use of fully autonomous weapons were restricted to certain locations or to specific purposes, for example, countries that usually respect international humanitarian law might develop and deploy such weapons and be tempted to use them in inappropriate ways.

Enforcement of regulations can also be challenging and leave room for error, increasing the potential for harm to civilians. Instead of clearly understanding that any use of fully autonomous weapons is unlawful, countries, international organizations, and nongovernmental organizations would have to monitor the use of the weapons and determine in every case whether use of the weapon complied with the regulations. There would probably be debates about enforcement and the scope of the regulations—for example, what constituted a populated area, in which the use of certain weapons would be banned.

Finally, the existence of fully autonomous weapons would leave open the door to their acquisition by repressive regimes or non-state armed groups that might not care about the regulations and could turn off programming designed to regulate a robot's behavior. In addition, fully autonomous weapons could be perfect tools of repression for autocrats seeking to strengthen or retain power. Even the most hardened troops can eventually turn

on their leader if ordered to fire on their own people. An abusive leader who resorted to fully autonomous weapons would not have to fear that armed forces would resist being deployed against certain targets.

Why should countries institute a *pre-emptive* ban?

The many concerns about fully autonomous weapons already suggest the weapons would be unacceptable legally and morally. A pre-emptive ban would eliminate the need to deal with the foreseeable problems later on.

Furthermore, it is difficult to stop technology once large-scale investments are made. If these weapons were deployed and the concerns about their potential harm to civilians proved justified, it would be very hard to put the genie back in the bottle. Countries would be greatly tempted to use technology already developed and to incorporate it into military arsenals. Many countries would be reluctant to give up the technology, especially if their competitors were deploying it.

There is precedent for a pre-emptive ban on a class of weapons. In 1995, countries agreed to a ban on blinding lasers before the weapons had started to be deployed out of concerns for the humanitarian harm the weapons would cause.[3]

A first step toward an international, pre-emptive ban could be national moratoriums on fully autonomous weapons. Christof Heyns, the United Nations special rapporteur on extrajudicial killings, called for national moratoriums in his May 2013 report to the Human Rights Council.[4] These interim bans would ensure that problematic weapons do not come into being and are not deployed while countries negotiate an international treaty. Countries that would not join a treaty but that want to prohibit fully autonomous weapons at the domestic level could also adopt moratoriums.

---

[3] Convention on Conventional Weapons, Protocol IV on Blinding Laser Weapons, adopted October 13, 1995, entered into force July 30, 1998. See also International Committee of the Red Cross, "Ban on Blinding Laser weapons Now in Force," news release, 98/31, July 30, 1998, http://www.icrc.org/eng/resources/documents/misc/57jpa8.htm (accessed April 12, 2013).

[4] Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns, on Lethal Autonomous Robots, Human Rights Council, 23rd Session, A/HRC/23/47, April 9, 2013, pp. 21-22.

What weapons would the ban encompass?

In general the ban would apply to any fully autonomous weapon—a weapon that could select and fire on targets without human intervention. At this point it is too early to determine which specific weapons would be included in or excluded from the scope of the definition of fully autonomous weapons. Countries usually agree on definitions at the end of a treaty negotiation process, as they did during the Oslo Process, which produced the Convention on Cluster Munitions in 2008.

Treaty drafters would have to consider what constitutes human intervention as they crafted a definition. One particular issue to address will be the categorization of certain automatic weapons defense systems, which are designed to sense an incoming munition, such as a missile or rocket, and to respond automatically to neutralize the threat. These systems may be better classified as "automatic" than "autonomous."

Automatic systems follow specific pre-programmed commands with little room for variation in a "structured environment," while autonomous weapons have more freedom to determine their own actions in an "open and unstructured" environment.[5] In addition, these automatic weapons often pose less of a threat to civilians because they are intended for defensive use against materiel targets and sometimes allow a human override.

Some experts, including the UN special rapporteur on extrajudicial killing, have said, however, that to exclude a weapon, the human ability to override its use would have to be meaningful.[6]

Would a ban entail an absolute prohibition on all development of autonomous robotic technology?

No. Research and development activities should be banned if they are directed at technology that can only be used for fully autonomous weapons or that is explicitly intended for use in such weapons. A prohibition on the development of fully autonomous

---

[5] Noel Sharkey, "Automating Warfare: Lessons learned from the Drones," *Journal of Law, Information & Science* (2011), p. EAP 2; Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Surrey, UK: Ashgate Publishing Limited, 2009), pp. 43-44.

[6] "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns," UN Doc. A/HRC/23/47, April 9. 2013, p. 8.

weapons would not ban development of all fully autonomous robotics technology because the technology can have positive, non-military applications.

The prohibition would also not encompass development of semi-autonomous weapons because those that are not fully autonomous raise different kinds of concerns. Even if the prohibition is a narrow one, as a matter of principle, countries should not be permitted to contract specifically for the development of fully autonomous weapons systems.