Lethal Autonomous Weapon Systems

By L. Righetti, Q.-C. Pham, R. Madhavan, and R. Chatila

The topic of lethal autonomous weapon systems (LAWS) has recently caught public attention due to extensive news coverage and apocalyptic declarations from famous scientists and technologists. Elon Musk said, "I think we should be very careful about artificial intelligence (AI). If I had to guess at what our biggest existential threat is, it's probably that" [27]. Stephen Hawking was quoted by the BBC as saying "the development of full AI could spell the end of the human race" [28].

As scientists, it is somewhat difficult and often frustrating to follow this narrative, especially when robotics is associated with science-fiction characters that are far from being a tangible reality. Yet, weapon systems with increasing autonomy are being developed due to fast improvements in machine learning, robotics, and automation in general [9]. These developments raise important and complex security, legal, ethical, societal, and technological issues that are being extensively discussed by scholars, nongovernmental organizations (NGOs), militaries, governments, and the international community. Unfortunately, the robotics community has stayed out of the debate, for the most part, despite being the main provider of autonomous technologies. In this column, we review the main issues raised by the increase of autonomy in weapon systems and the state of the international discussion. We argue that the robotics community has a

The IEEE Robotics and Automation Research and Practice Ethics Committee (RARPEC) is intended as a platform to exchange ideas and discuss the impacts and practice of robotics and automation (RA) technologies in research, development, and deployment that appear to pose ethical questions for humanity. With increased awareness and controversies surrounding RA, RARPEC is publishing a series of opinion pieces that will focus on separating hype from reality by providing an objective and balanced treatment of technological, ethical, legal, and societal perspectives. First in the series, this piece focuses on the topic of lethal autonomous weapon systems. Please send your feedback and suggestions to the chair of the committee, Raj Madhavan, at raj.madhavan@ieee.org.

fundamental role to play in these discussions, for its own sake, to provide the often-missing technical expertise necessary to frame the debate and promote technological development in line with the IEEE Robotics and Automation Society (RAS) objective of advancing technology to benefit humanity.

The First International Multilateral Discussion on Autonomous Weapon Systems

The increasing autonomy in weapon systems became an international issue in April 2013 with the annual report to the United Nations (UN) Human Rights Council of Christof Heyns, the Special Rapporteur on extrajudicial, summary, or arbitrary executions, which discussed the development of weapons "that, once activated, can select and engage targets without further human intervention." He argued that these weapons raised concerns on the basis of their compliance with international humanitarian law (IHL) and on the ethical considerations that "their deployment may be unacceptable because no adequate system of legal accountability can be devised,

and because robots should not have the power of life and death over human beings" [1]. In November 2012, the U.S. Department of Defense had released a directive on autonomy in weapon systems to "establish policy and assign responsibilities for the development and use of autonomous and semi-autonomous functions in weapon systems," [19] making the United States the first country to have an official policy on LAWS. At about the same time, the Campaign to Stop Killer Robots, formed by ten NGOs and led by the Human Rights Watch, demanded "a comprehensive, preemptive prohibition on the development, production, and use of fully autonomous weapons" [13], [16].

The topic of lethal autonomous weapon systems has recently caught public attention due to extensive news coverage and apocalyptic declarations from famous scientists and technologists.

In 2013, in what can be considered the first multilateral international discussion on autonomous systems, the high contracting parties to the Convention on Certain Conventional Weapons (CCW) decided by consensus to begin yearly informal discussions on LAWS. (The CCW has a total of 125 states that are high contracting parties and five signatories. "The purpose of the Convention is to ban or restrict the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately" [20].) In December 2016, the contracting parties decided to "establish a Group of Governmental Experts… with a mandate to assess questions related to emerging technologies in the area of LAWS" [21], which underlines the importance of the issue for the international community. However, there is not yet a consensus on whether new international treaties are necessary to regulate these developments.

A few states, notably states with an advanced weapons industry, such as the Russian Federation, the United States, and Israel, argue that there is no need to negotiate a specific treaty and the current international law is sufficient. A growing number of states (22 as of December 2017, including Brazil, Iraq, and Pakistan) support a preemptive ban on LAWS. Most states agree that a certain degree of meaningful human control is necessary at all times when deploying increasingly autonomous weapons and that new regulations might be necessary [4], [5], [26]. (The term *meaningful human control* was first introduced by NGO Article 36 [15], which originally talked about "meaningful human control over individual attacks." The term has since been used more broadly to help frame the discussion on LAWS. See [5] for an in-depth discussion.)

A Complex Multifaceted Problem

Autonomy is becoming pervasive in noncritical components of weapon systems, such as transport, navigation, or surveillance, and has already had an impact on the use of military force by nations. Partial autonomy in the navigation and surveillance capabilities of drones, e.g., has been decisive in the rapid and extensive deployment of the controversial U.S. drone program [10], [11]. However, it is the autonomy in the critical functions [2]–[4], i.e., acquisition of a target and decision to kill, that is at the center of the debate on LAWS. Such systems have not been deployed yet, but they are not technologically out of reach.

While several systems that autonomously select and attack targets already exist, such as the Terminal High Altitude Area Defense or the Iron Dome, they are not necessarily an issue because they are designed to defensively respond to incoming missiles rather than offensively target humans [9]. One can also argue that automated systems targeting humans already exist, the simplest example being antipersonnel mines. The differences are that they are nondiscriminative, already banned by the Ottawa Treaty, have a limited scope of action, and, most importantly, there is no algorithmic process of target acquisition and decision to kill. While robotics rarely makes a distinction between automated and autonomous, the distinction plays an important role for policymakers to differentiate types of weapons. While there is no consensus on the definition, autonomous includes the notion of an algorithmic decision-making process difficult to predict, while automated is related to a response deterministically triggered by well-defined events.

A major legal concern of weapon systems able to autonomously target and decide to kill people is whether they can be compatible with IHL and, in particular, with its core principles of distinction, proportionality, and precaution. (See [29] for a proper overview of IHL.) The principle of distinction is the necessity to discriminate between combatants and civilians at all times. The principle of proportionality restricts the amount and kind of force that can be used in a conflict such that it does not create excessive injuries and death of civilians and damage to civilian objects with respect to the direct anticipated military advantage [17]. The principle of precaution states that the parties involved in a conflict must take all feasible precautions to protect the civilian population and objects under their control against the effect of attacks.

The first issue raised by LAWS is whether a machine can comply with these principles (e.g., can a machine distinguish civilians from combatants?). It is particularly difficult because the interpretation of IHL is heavily dependent on the context (how can an algorithm balance expected civilian casualties with direct military advantage?). The second issue is related to predictability: is it possible to guarantee the compliance of a machine with IHL in the battlefield at all times? How will the machine behave in extreme adversarial conditions when something unexpected happens? Will the machine be capable of accurate situation assessment at all times? The last issue is related to accountability: will it be possible to trace exactly what went wrong in case of failure? How can one attribute criminal responsibility in the case of IHL violation?

All these issues are tightly connected to the technology, what it is capable of doing, how it can fail, and the associated certification procedures. This becomes particularly problematic when considering advanced planning and machine-learning algorithms. It is seldom possible to predict every possible outcome, as these algorithms are not necessarily deterministic, and their behavior depends on the training

Most states agree that a certain degree of meaningful human control is necessary at all times when deploying increasingly autonomous weapons and that new regulations might be necessary.

data and the environment. One can easily see how any meaningful legal review of conformity with IHL will be difficult, if not impossible, in the near future [22]. Article 36 of Additional Protocol I of the 1949 Geneva Conventions imposes an obligation on states to conduct a review for all new weapons to determine if their use would be, in all or some circumstances, prohibited by IHL.

The strict adherence to the principle of proportionality and distinction is not sufficient to make the use of a weapon acceptable; ethical considerations are not strictly covered by the law. The Martens Clause [23], also known as the *dictate of public conscience*, states that the fact that there is no law prohibiting a weapon does not mean that its use is permitted. There are ethical aspects that need to be taken into account. In other words, the usage of a weapon is also associated with its societal and ethical acceptance.

It was argued that the dictate of public conscience played an important role in the prohibition of chemical weapons, blinding lasers, and antipersonnel landmines [6]. But ethical considerations can vary across societies, cultures, and people. For example, based on ethical considerations, there have been (very few) proponents of LAWS, arguing that they are a moral obligation if they allow the reduction of harm. Robots would not kill out of anger, which makes them more ethical [14].

A major ethical concern, however, is raised in all cases: can we delegate the decision to kill to an algorithm? We argue here that even if a robot could lawfully decide to kill, allowing this would undermine our very notion of humanity and could be considered an affront to human dignity [1], [2], [6]. Moreover, when allowing an algorithm to make a life-and-death decision, there is, in principle, no reason to limit the abilities of algorithms in other contexts (law enforcement, health care, and so forth). Setting limits on what machines can and cannot do, based on ethical considerations, might have consequences to both

civilian and military applications independent of which one sets these norms first.

In addition to these legal and ethical issues, many fear that the development of LAWS will lead to an arms race between nations developing advanced weapons, with the associated risks for international stability and security. Proliferation is another concern, as LAWS can be rather cheap to develop with the increasing availability of off-the-shelf software and hardware components that can have a dual use. Without adequate legislation and norms to control the development of LAWS, one can easily imagine usage of such weapons by nonstate actors and terrorist groups. The deployment of weapon systems with increased autonomy, with very fast decision loops and potentially low or no human oversight, together with cybersecurity threats also pose serious concerns to the operational risks created by these technologies [7], [8], [16], [24].

The Role of Scientists in the Debate

While scientists and technology leaders have already been quite active in the debate, e.g., with multiple open letters with national and international resonance, these are individual initiatives that come mainly from the AI community. The 2015 open letter from the Future of Life Institute calling for a ban on LAWS beyond meaningful human control was signed by more than 3,000 robotics and AI researchers. In 2017, an open letter signed by more than 100 founders of robotics and AI companies urged the UN to address the issues raised by LAWS. Each time, the letters were publicly announced at the International Joint Conference on AI. In November 2017, hundreds of AI scientists in Canada, Belgium, and Australia sent open letters to their respective governments urging them to take action on the issue or sharing their concerns about the development of increasingly autonomous weapons.

The robotics community as a whole has largely remained silent despite the

development of associated underlying technologies and, more importantly, the active role it could play in this debate. There is an important need for independent expertise for all the parties involved in the discussions (and not only for actors with already high technological capabilities) to better frame the debate and design meaningful policies. In addition to its technical expertise, the robotics community can actively develop and enforce ethical norms for the use of autonomous technologies, which should include LAWS. While the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems is already addressing some of these aspects [18], [25], there is still a need for more important effort to be made to establish better practices, e.g., to define the appropriate level of human control in autonomous systems.

> **There is an important need for independent expertise for all the parties involved in the discussions to better frame the debate and design meaningful policies.**

In this column, we argue that it is not possible to advocate for the status quo and that it is our responsibility as a community to provide technical expertise and participate in the design of national and international regulations. More importantly, it is our responsibility to establish ethical norms, standards, and practices for weapon systems with increased autonomy in a manner that is compatible with the mission of RAS, which is "to foster the development and facilitate the exchange of scientific and technological knowledge in RA that benefits members, the profession, and humanity" [30].

Acknowledgments

Sweden, for their valuable comments on the initial version of this column.

References

[1] C. Heyns. (2013). Annual report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns. United Nations Human Rights Council. [Online.] Available: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G13/127/76/PDF/G1312776.pdf?OpenElement

[2] "Autonomous weapon systems: Technical, military, legal and humanitarian aspect," Expert Meeting Report, International Committee of the Red Cross, 2014.

[3] "Autonomous weapon systems: Implications of increasing autonomy in the critical function of weapons," Expert Meeting Report, International Committee of the Red Cross, 2016.

[4] Framing discussions on the weaponization of increasingly autonomous technologies (part I) (2014). United Nations Institute for Disarmament Research. [Online]. Available: http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-phase-iii

[5] The weaponization of increasingly autonomous technologies (part II): Considering how meaningful human control might move the discussion forward. United Nations Institute for Disarmament Research. (2014). [Online]. Available: http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-phase-iii

[6] The weaponization of increasingly autonomous technologies (part III): Considering ethics and social values. United Nations Institute for Disarmament Research. (2014). [Online]. Available: http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-phase-iii

[7] The weaponization of increasingly autonomous technologies (part VI): Concerns, characteristics and definitional approaches. United Nations Institute for Disarmament Research. (2017). [Online]. Available: http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-phase-iii

[8] The weaponization of increasingly autonomous technologies (part VII): Autonomous weapon systems and cyber operations. United Nations Institute for Disarmament Research. (2017). [Online]. Available: http://www.unidir.org/programmes/emerging-security-issues/the-weaponization-of-increasingly-autonomous-technologies-phase-iii

[9] "Mapping the development of autonomy in weapon systems," Stockholm International Peace Research Institute, 2017.

[10] "Living under drones: Death, injury, and trauma to civilians from US drone practices in Pakistan," Stanford Law School and NYU School of Law, 2012.

[11] "Will I be next? U.S. drone strikes in Pakistan," Amnesty International, 2013.

[12] "Autonomous weapon systems under international law," Geneva Academy of International Humanitarian Law and Human Rights, 2014.

[13] "Losing humanity: The case against killer robots," Human Rights Watch, 2013.

[14] R. Arkin. Governing Lethal Behavior in Autonomous Robots. Boca Raton, FL: Chapman and Hall/CRC Press, 2009.

[15] Article 36. Killer robots: U.K. government policy on fully autonomous weapons, Article 36. (2013, Apr.) [Online]. Available: http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf, 2013.

[16] Campaign to stop killer robots. [Online]. Available: http://www.stopkillerrobots.org/

[17] Customary international humanitarian law. [Online]. Available: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_cha_chapter4_rule14

[18] Reframing autonomous weapons systems. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. [Online]. Available: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

[19] "US Department of Defense Directive Number 3000.09: Autonomy in weapon systems," U.S. Department of Defense, Nov. 21, 2012.

[20] The United Nations Office at Geneva. The convention on certain conventional weapons. [Online]. Available: https://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument

[21] The United Nations Office at Geneva. Background on lethal autonomous weapons systems in the CCW. [Online]. Available: https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument

[22] V. Boulanin and M. Verbruggen, "Article 36 reviews: Dealing with the challenges posed by emerging technologies," Stockholm International Peace Research Institute, 2017.

[23] R. Ticehurst, "The Martens Clause and the laws of armed conflict," Int. Rev. Red Cross, no. 317, Apr. 1997.

[24] P. Scharre. Autonomous weapons and operational risk. (2016, Feb.) [Online]. Available: https://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515

[25] "Ethically aligned design, a vision for prioritizing human well-being with artificial intelligence and autonomous systems," IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 2016.

[26] V. Boulanin, "Mapping the debate on LAWS at the CCW: Taking stock and moving forward," European Union Non-Proliferation Paper No. 49, Stockholm International Peace Research Institute.

[27] S. Gibbs. (2014, 27 Oct.) Elon Musk: Artificial intelligence is our greatest existential threat. The Guardian. [Online]. Available: https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat

[28] R. Clellan-Jones. (2014, 2 Dec.) Stephen Hawking warns artificial intelligence could end mankind. BBC News. [Online]. Available: http://www.bbc.com/news/technology-30290540

[29] International Committee of the Red Cross. IHL database: Customary IHL. [Online]. Available: https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul

[30] IEEE Robotics and Automation Society. [Online]. Available: http://www.ieee-ras.org/about-ras