The Hague
Centre for
Strategic
Studies

HCSS Security

Governing autonomous weapon systems

Expanding the solution space, from scoping to applying

Esther Chavannes, Klaudia Klonowska, Tim Sweijs

Executive Summary

Current discussions around autonomous weapon systems (AWS) have generated concerns whether existing regulatory regimes are still fit for purpose in light of the challenges posed by new military technologies. Major military powers are actively exploring the utility and use of autonomous systems in war in order to gain a military competitive edge. Thus far, the notion of 'meaningful human control' (MHC) has been central to international discussions about AWS. These discussions have stalled, both because of different interpretations of MHC and due to the unwillingness of key participants to curtail their ability to potentially leverage AWS in the future.

This paper therefore seeks to reframe the discussion. It first surveys alternative suggestions for AWS oversight and control, and then moves beyond current conceptual scoping to offer actionable insights and recommendations for policymakers working on AWS governance at the crossroads of security and technology in the ministries of Foreign Affairs, Defense, and Economic Affairs. The relevance for these three groups of policymakers is arms control and international cooperation on related technology; the development of these technologies within legal and ethical boundaries; and the development of relevant industry standards respectively.

The initial scoping exercise on the literature on regulation of AWS generated a series of solutions (see Figure 1 below) which fall into three categories:

- First, framework-based solutions which focus on accountability throughout the life cycle of AWS (colored blue in Figure 1);
- Second, law-based proposals which emphasize accountability not just for states, but also for private actors (yellow);
- Third, technological solutions which aim to improve the explainability and verifiability of systems in a way that can be understood and operationalized by everyone – from policymakers to end users (green).

Most of the proposed solutions in these three categories are complementary in nature. They were aligned along the AWS life cycle in order to identify the focus of existing solutions and assess potential gaps in the solution space.

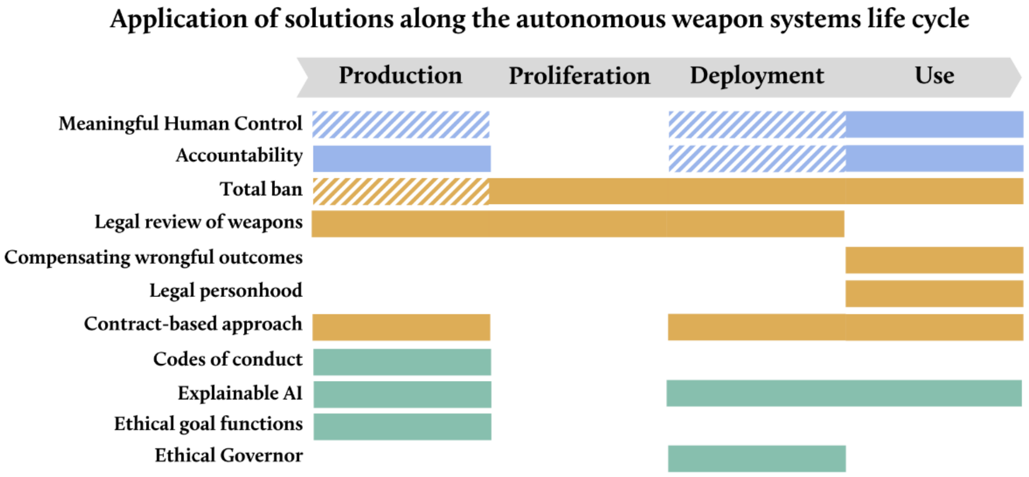


Figure 1. Application of solutions along the autonomous weapon system life cycle

At a workshop organized in late 2019, approximately twenty legal, military, political, and technology experts convened to discuss these proposed solutions. The authors of this paper identify five salient insights based on this workshop:

- First, the potential uses of AWS are not governed in a legal vacuum. Rather than creating new regulation there is a need to clarify and specify the application of existing binding legal frameworks to AWS and the technologies of which they are composed;
- Second, to the extent that AWS are covered by existing regulation, it is important to improve and ensure states' compliance with these existing regulations;
- Third, non-legal approaches, including technology-based solutions, can complement existing or improved regulation in order to properly govern and regulate private actors involved in the development of emerging technologies associated with AWS.
- Fourth, smaller states, such as the Netherlands, industry, and epistemic communities can work together in coalitions of the willing in order to ensure that high-quality domestic standards for relevant technologies are promoted and used as starting point for the development and adoption of standards for the production and use of technologies worldwide.
- Fifth, the proper translation of higher-level legal and ethical norms and rules into the regulation of lower-level technology standards and protocols that can be promoted domestically and internationally, requires enhanced technological literacy on the part of policymakers.

Based on the analysis of assembled solutions and the expert session insights, the authors of this paper propose the following recommendations for policymakers. At the domestic level:

1. *Ensure and safeguard legal compliance* – review compliance with, and publish and regularly update guidelines to, the applicability of international humanitarian law (IHL) to AWS and related technology, with particular attention to the early stages of the life cycle, in order to set domestic standards;
2. *Set standards and protocols* – define the purpose and goals of acquired and manufactured military technology and set up guidelines for context-specific Codes of Conduct for private actors. Identify desired and undesired outcomes, as well as (technological) solutions to improve the verifiability and explainability of these outcomes;
3. *Tailor contractor work to military needs* – continuously test and re-evaluate the pre-determined settings and capabilities of a system, in order to ensure that manufacturing is tailored to the use environment;
4. *Improve technological literacy of policymakers* – ensure adequate and recurrent cooperation with researchers and military personnel who are spearheading development, innovation, and application. Where necessary, consider bringing in new employees with technical expertise in order to increase understanding of the technological conditions for policies;
5. *Improve engagement of and with military personnel* – cooperate closely with private contractors and manufacturers of AWS as well as policymakers, in order to ensure interaction between legal frameworks, technological standards, military effectiveness, and rules of engagement.

At the international level:

6. *Promote legal compliance* – clarify application of existing international regimes to AWS and related technologies, and promote widespread compliance among states, especially concerning the Article 36 legal review of weapons. Initiatives such as the UN GGE's 11 Guiding Principles can provide a starting point in how this can be done;
7. *Lead by example* – lead the implementation of domestically developed doctrines and standards into international standards for the production and use of AWS. This furthers ethical considerations in technology development as well as standards that align with existing international laws and norms;
8. *Improve transparency* – share with other states and with national publics the insights gained from setting standards and protocols and applying regulations

around AWS and related technology, in order to improve international harmonization and understanding;

9. *Strengthen cooperation on standard-setting* – establish a 'coalition of the willing' among states to share practices, promote standards for new technologies relevant to the military field, and share the burden of expenses related to the validation and verification of the technologies that go into AWS.
10. *Stimulate epistemic communities* – reinforce and stimulate an international epistemic community to facilitate the exchange of technical and legal knowledge and inform international policymaking with expertise, and build bridges between research and industry to translate legal and ethical rules and norms into technology standards and protocols.

Table of Contents

Introduction

International discussions on the regulation of autonomous weapon systems (AWS) have been stalling. China states it supports a ban on AWS use, but refuses to curb research and development in this area;[1] Russia's military chief of staff has described the possibility of, and interest in, creating a robotized unit;[2] and the US encourages innovations in the area of AWS development, citing their potential "to reduce the risk of civilian casualties and damage to civilian objects."[3] Meanwhile, civil society as well as many smaller and middle powers watch these developments with apprehension as they continue their efforts to create binding international agreements on the moving target that is AWS regulation. At the same time, actors have vastly differing views on the definition of 'meaningful human control' (MHC), which is currently the main frame of reference for regulating AWS. We might therefore rethink whether MHC is the most effective and/or the sole frame through which to approach the regulation of AWS taking into account different incentive structures of the principal actors. In reframing this discussion, this paper canvasses alternative solutions to oversight and control of AWS and offers a set of practical, actionable insights on new ways forward in AWS governance.

This paper is structured as follows. First, it highlights the most important ethical and legal, political-economic, and military-strategic incentives of different stakeholders – including states, tech companies, and civil society – to (not) regulate AWS and related technologies (section 1). Second, it surveys challenges seen in current regimes (section 2), analyzes solutions that have been proposed in the literature on AWS regulation (section 3), and aligns them along the AWS life cycle, which comprises production, proliferation, deployment, and use (section 4). This provides a comprehensive overview of the current state of debate and helps to assess potential gaps in the current approach to AWS governance. This paper then draws key insights from an expert session on AWS regulation (section 5) and presents these takeaways as recommendations for policymakers working on AWS governance at the crossroads of security and emerging technology, in the ministries of Foreign Affairs, Defense, and Economic Affairs of the Netherlands, as well as those in similar lines of work in the ministries of like-minded governments (section 6).

[1] Kania, "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems."

[2] Scharre, *Army of None: Autonomous Weapons and the Future of War*.

[3] "Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapon Systems. Submitted by the United States of America."

1. Why regulate AWS: different actors, different incentives

A variety of stakeholders are involved in debates regarding the regulation of AWS. Depending on certain motivations or incentives, these actors range from advocates for stricter regulations to opponents of further limitations on the production, proliferation, deployment, and use of AWS. This section describes three categories of considerations stakeholders may have: ethical and legal, political-economic, and military-strategic.

1.1 Ethical-legal

One of the main reasons to aim for stricter regulation of AWS lies in discomfort with the idea of delegating decisions to kill to an algorithm. Actors subscribing to this ethical viewpoint are predominantly civil society, as well as a good number of states.[4] There is additionally a substantive group of private actors and researchers that subscribe to and propagate these ethical standards. Notably, these actors have formed a coalition to push for further regulations of lethal autonomous weapon systems (LAWS) in The Campaign to Stop Killer Robots. Additionally, a number of states have expressed an interest in fully banning LAWS including Brazil, Colombia, Mexico, Pakistan, and Venezuela.[5]

A greater group of states, however, does not consider a ban to be the best or only option to address ethical concerns linked to AWS, but instead they emphasize the role of legal frameworks. In this view, states recognize the role of AWS in improving the accuracy of targeting and reducing collateral damage on the one hand, and the need to regulate far-reaching reliance on algorithms to avoid bias or other issues in military decision-making on the other. This approach focuses on the regulation of quality and diversity of data while leaving open the possibilities for the development of legally sound military technologies. Among others, the United States is an advocate of this approach, highlighting the application of principles of international humanitarian law (IHL) to AWS.[6]

1.2 Political-economic

The abovementioned movement to ban AWS has led to scrutiny of states and companies involved in their production and use. Technology companies have been

[4] Netherlands Advisory Council on International Affairs, "Autonomous Weapon Systems: The Need for Meaningful Human Control."

[5] "Country Views on Killer Robots."

[6] "Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapon Systems. Submitted by the United States of America."

seen to withdraw from AWS-related projects in order to preserve their reputation and economic prospects.[7] However, commitments of tech companies to not produce AWS as a whole are not necessarily transferable to related technologies that such systems are composed of. Companies have significant economic incentives to continue the production and development of AI that may have civilian applications, as well as potential use for the military.

Similarly, states can be shamed by the public, civil society, and smaller states for their investments in or development of AWS. Smaller states who are unable to compete against advanced military forces are pushing for stricter regulations of both the development and the use of AWS. Regardless of these pressures, bigger states with investments in highly sophisticated technologies remain reluctant to impose strict regulations, both for political and economic reasons. Among the political reasons is a trend of societal intolerance for civilian casualties,[8] resulting in an exacerbated "need for advanced systems of defense and precision-guided weapons."[9] Among the economic reasons is the fact that states benefit economically from innovation, development, and production of military technologies in the private sector.[10] States with advanced technological capabilities profit from the growth of domestic technological companies in many fields related to artificial intelligence (AI). Therefore, while potentially supportive of the restrictions on the use of AWS, these technologically advanced countries may be reluctant to regulate their research and development.

1.3 Military-strategic

Apart from ethical-legal and political-economic considerations, AWS have come to be seen as a threat to international peace and security. AWS are expected to lower the threshold to wage war and increase the speed of decision-making in a way that some fear will lead to rapid conflict escalation.[11] Meanwhile, relatively sophisticated military technology is becoming increasingly accessible to non-state actors.[12] Stakeholders calling for stricter regulations in order to prevent the escalation of conflict include a number of smaller and middle-sized countries with less developed military forces, as well as, to a larger extent, international organizations and civil society, such as the UN

[7] Beenes et al., "Don't Be Evil? A Survey of the Tech Sector's Stance on Lethal Autonomous Weapons," 4.

[8] Scharre, *Army of None: Autonomous Weapons and the Future of War*; McLean, "Drones Are Cheap, Soldiers Are Not."

[9] Chavannes and Arkhipov-Goyal, "Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context," 8.

[10] Bloom et al., "Policies to Promote Innovation"

[11] Beenes et al., "Don't Be Evil? A Survey of the Tech Sector's Stance on Lethal Autonomous Weapons," 9.

[12] Beenes et al., 9.

Secretary General António Guterres, Human Rights Watch, and the Netherlands-based organization PAX for Peace.[13]

At the same time, some states such as the United States, France, the United Kingdom, and China have already invested in AWS-related technologies and are thus far less willing to regulate their production.[14] They are also both aware of and wary of the state of military-technological developments in adversarial countries.[15] Instead of increasing regulation, therefore, some of these states propose to develop and use AWS according to 11 Guiding Principles. This Guiding Principles initiative came forth after discussions in the Convention on Certain Conventional Weapons' (CCW) Group of Governmental Experts (GGE), and it was pushed forward and opened for endorsement by France and Germany in an 'Alliance of Multilateralism' event in September of 2019.[16] Regulating AWS using a principle-based approach is motivated by these countries' wish to respond to the ethical concerns while ensuring progress of innovation and development of strategically advantageous military technology.[17] Proposed solutions for the regulation of AWS will need to take into account this constellation of different incentive structures.

[13] "Autonomous Weapons That Kill Must Be Banned, Insists UN Chief"; Kayser, "Killer Robots"; Docherty, "Losing Humanity."

[14] Wareham, "As Killer Robots Loom, A Push to Keep Humans in Control of Use of Force."

[15] Jones et al., "Managing the New Threat Landscape: Adapting the Tools of International Peace and Security."

[16] "11 Principles on Lethal Autonomous Weapons Systems (LAWS)."

[17] Boulanin and Verbruggen, "Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies."

2. Problems with existing regimes

This section presents the state of the current regimes for AWS oversight and/or control as identified in the literature. Arms control started as a strategic measure after the Second World War, but has in the past few decades also been interpreted from a more humanitarian point of view.[18]

Currently, the most prominent issues appearing in the literature surrounding AWS that spur the call for increased or improved regulation of some form, lie in (1) concerns over insufficiencies in current regulations to ensure compliance with international humanitarian law (IHL); (2) difficulties in establishing lawful preemption and use of force by AWS; (3) a lack of clearly established accountability frameworks; and (4) an inadequate or unclear positioning of AWS within existing arms exports regimes.

2.1 Autonomous decision-making and compliance with humanitarian law

At the moment, there are no bans or regulations under international law that outlaw AWS explicitly. At the same time, however, using AWS in conflict means that their development, proliferation, deployment and use are bound by IHL.[19]

2.1.1 General applicability of IHL principles

It is at present time unclear whether and how AWS and related technologies can comply with IHL principles.[20] The concern is that the more open-ended legal provisions of IHL leave room for context-specific interpretation which cannot be easily translated or incorporated into a weapon system's functionality.[21] For example, the principle of distinction is at risk, as the assessment of the civilian/combatant status requires a complex analysis of an actor's actions, motives and the intensity of involvement in hostilities – especially in densely populated urban areas. On the contrary, there are certain principles of IHL, such as the principle of precaution in attacks, which might be interpreted to improve with the use of algorithms when their

[18] Okano-Heijmans and Klijn, "Input Paper: Managing RAS: The Need for New Norms and Arms Control."

[19] IHL comprises international rules protecting people and property in conflict by setting limits on how conflicting parties may choose their methods and means of warfare. The key principles of IHL are proportionality, military necessity, distinction, and humanity (also referred to as the Martens Clause). See for example Bouvier, "International Humanitarian Law and the Law of Armed Conflict."

[20] Petman, *Autonomous Weapons Systems and International Humanitarian Law*.

[21] Schuman, "Situational Awareness and Adhere to the Principle of Distinction as a Necessary Condition or Lawful Autonomy." For the legal complexities of the civilian/combatant status, see also Melzer, "Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law."

decisions are more accurate than those of the military personnel in order to prevent excessive loss of lives.[22]

2.1.2 Principle of humanity (Martens Clause)

The lack of a clear definition of the Martens Clause (principle of humanity) hinders its application to AWS.[23] The Martens Clause is one of the oldest principles of IHL, and it requires state conduct in warfare to adhere to norms as "dictated by public conscience."[24] Accordingly, states should stop the development of AWS if they consider the killing of humans by an autonomous machine contrary to the generally accepted norms.[25] Alternatively, they could innovate in certain areas of these technologies if they see it as a way to improve adherence to societal norms.[26] The Martens Clause is phrased in an open-ended manner, which means that depending on an actor's interests, interpretations of the principle of humanity in the context of AWS can be anywhere from very prohibitive to nearly redundant.

2.1.3 Legal review of new weapons (Article 36)

The legal review mechanism already applies to AWS, yet it is uncertain whether and how it can effectively prevent violations of IHL. Under Article 36 of the Additional Protocol I to the Geneva Conventions, states are under an obligation to conduct a review of the legality of new weapons – which would include AWS.[27] Since neither the Geneva Conventions nor the International Committee of the Red Cross (ICRC) that helped develop this mechanism provide guidelines to the application of the legal review, there is much uncertainty about how to deal with the complexity of emerging military technologies and whether the review should include related non-weaponized technologies assisting in the targeting process. For example, the use of machine learning or other AI-based techniques can lead to not easily predictable outputs that

[22] By the principle of precaution in attack, states are obliged to choose the means and methods of warfare that are expected to cause least damage to civilian objects and harm to the civilian population. Therefore, if military technologies become more accurate and reliable in certain areas than human decision-makers, their use might be required in armed conflict to prevent excessive loss of civilian lives. See Queguiner, "Precautions under the Law Governing the Conduct of Hostilities," 798–99; Lawland, "Reviewing the Legality of New Weapons, Means and Method of Warfare"; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I).

[23] Sparrow, "Ethics as a Source of Law"; "Expert Meeting on Lethal Autonomous Weapons Systems."

[24] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I). These principles originate from 1899 and 1907 Hague Conventions.

[25] Docherty, "Losing Humanity," 26.

[26] "Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapon Systems. Submitted by the United States of America."

[27] Article 36 - New weapons - Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I).

pose a challenge to trial-stage assessments of expected consequences. Furthermore, the effective control of AWS through Article 36 is hindered by a lack of a widespread or transparent implementation.[28]

2.2 (Semi-)autonomous use of force and international peace and security

The use of autonomy in weapon systems arguably lowers the threshold to wage war thus posing a threat to the prohibition of the use of force.[29] The speed with which AI-based military technology makes decisions poses a challenge to 'strategic stability' and international peace and security.[30] The interaction between (semi-)autonomous systems and their predictive nature makes it less predictable whether and when the system reacts with force, possibly leading to unintended escalation.[31] Especially in cases of self-defense mechanisms, there is a possibility that automated systems miscalculate or misidentify threats, which could lead to premature or otherwise unlawful use of force.[32] A concerning example is the use of (semi-)autonomous systems for nuclear weapons, which threatens the nuclear deterrence regime. This challenge indicates that deployment and use of AWS must be strictly controlled in order to prevent states from unwillingly breaching their obligations to sustain international peace and security.[33]

2.3 Wrongful acts and the attribution of accountability

2.3.1 Difficulties in establishing state accountability

The use of AWS blurs the responsibilities of the numerous actors involved in the development, deployment and (semi-)autonomous use of force, which hinders the determination of accountability. As established in the International Law Commission's (ILC) principles for wrongful acts of states, the only conduct attributable to states at the international level is that of their government organs or of

[28] Kayser and Denk, "Keeping Control."

[29] "LAWS: Ten Problems for Global Security"; Amoroso, "Jus in Bello and Jus Ad Bellum Arguments against Autonomy in Weapon Systems: A Re-Appraisal." The Austrian and Chinese delegations at the 2014 CCW Meeting of Experts argued that the use of autonomous weapons removes the restrains of the conduct of war related to the risks of the deaths of soldiers, thus making the use of force a more likely. The body of international law that regulates and attempts to restraint the use of force is referred to as 'jus ad bellum'.

[30] Strategic stability occurs when adversaries lack significant incentive that would lead them to engage in provocative behavior. See also "LAWS: Ten Problems for Global Security."

[31] "LAWS: Ten Problems for Global Security"; Shaw, *Predator Empire*; Johnson, "Artificial Intelligence & Future Warfare." CYBERCOM, for example, can "deter and respond to preempt cyber threats in all phases of conflict," see Soesanto, "The Evolution of US Defense Strategy in Cyberspace (1988 – 2019)"; Geist and Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?," 8; Haner and Garcia, "The Artificial Intelligence Arms Race"; Bromley and Maletta, "The Challenge of Software and Technology Transfers to Non-Proliferation Efforts."

[32] It is possible that certain technical weaknesses will be exploited – intentionally or even accidentally – leading the system to react with force at a time where it was not warranted. Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States."

[33] "Charter of the United Nations."

people acting under these organs' "direction, instigation or control" as "agents of the State."[34] In the complicated life cycle of autonomous military systems, it becomes nebulous who counts, on the one hand, as agents taking orders within the official structure, and on the other hand, as agents of the state making individual decisions that result in wrongful acts outside of an "effective command and control structure."[35] In other words, there is a lack of clarity on who is responsible for wrongful acts resulting from the use of AWS and whether they are sufficiently linked to the acts of the state.

2.3.2 Criminal law and the notion of intent

The use of fully or semi-autonomous weapons inherently lacks the intention to cause harm since the decisions are made by machines and not humans, which challenges the application of criminal liability for war crimes. For international criminal law (ICL), the evidence of intent is crucial to the attribution of wrongful act.[36] However, as AWS have no consciousness, no criminal intent can be established, and therefore ICL cannot be applied directly to AWS. It is unclear to what extent the responsibility for wrongful acts should fall on the commanders, the human operator, the software developer, or the designer of the military technology. The number of actors involved in the life cycle of AWS significantly challenges the attribution of accountability. Thus, it must be established how criminal intent is attributed in the case of AWS. This determination is likely to be an outcome of the judicial procedures, for example, judgements of the International Criminal Court.

2.4 Non-proliferation efforts and the applicability of export control regimes

The complex nature of AWS – the combination of hardware and software – poses a challenge to the adequate regulation of exports and proliferation. In turn, exports control regimes struggle with the adequate regulation of all the different elements that make up AWS, some of them of a dual-use nature. For example, some smaller civilian-use drones may double as a weapon when embedded with additional software, such as a facial recognition, improved upon to serve for the purposes of target identification. Non-proliferation efforts are additionally complicated by the rapid advancement in digital sharing and storage possibilities. Cloud computing, for

[34] International Law Commission, "Responsibility of States for Internationally Wrongful Acts," 38, para 2.

[35] Chavannes and Arkhipov-Goyal, "Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context," 53; "Killer Robots and the Concept of Meaningful Human Control."

[36] Crootof, "War Torts: Accountability for Autonomous Weapons"; "Willfully" here means that someone must have acted either intentionally or recklessly, see Sandoz, Swinarski, and Zimmermann, "Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949"; Chavannes and Arkhipov-Goyal, "Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context," 54.

example, raises complicated questions of "whether, how and when controls on transfers of software and technical data should be applied."[37] An example of an existing regime to limit the exports of dual-use technologies is the Wassenaar Arrangement.[38] The regime establishes a list of export controlled items and promotes transparency on the basis of voluntary information sharing, in order to control the global exchange of dual-use goods and technologies.[39] Further efforts are necessary to continuously adjust the list to adequately address the technologies used in AWS.

2.5 Conclusion

The discussion regarding AWS shows that the regulation of AWS at the very least pose a challenge to existing regimes. Some of the existing regulations could be reinterpreted and adjusted to challenges posed by the emerging use of autonomous systems in military forces. Where AWS are deployed in armed conflicts, fundamental questions relating to IHL must be answered: first, how to apply IHL principles to military technology, and second, how to adequately review the legality of military-use technology. It must further be established who should be accountable for wrongful acts and whether the responsibility lies with state actors or other actors involved in the entire life cycle of the hard- and/or software. In these regards, efforts to regulate the production, proliferation, deployment, and use of AWS and related technologies require critical analysis and creative thinking. The following section surveys an assortment of solutions presented by governments, civil society, and academics.

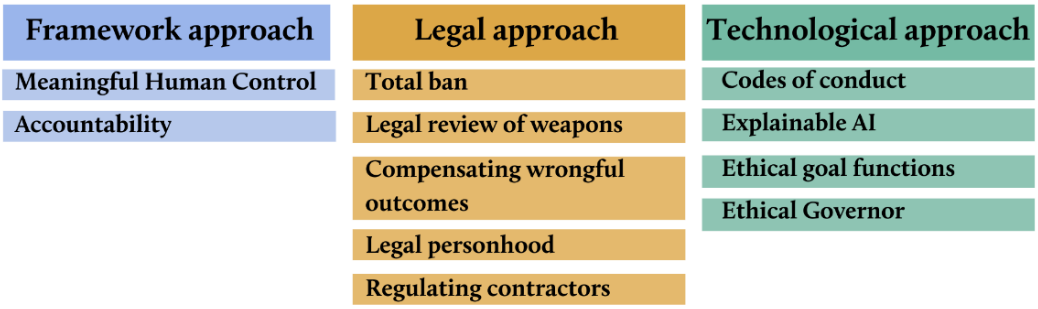
[37] Bromley and Maletta, "The Challenge of Software and Technology Transfers to Non-Proliferation Efforts."

[38] "Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies. Public Documents Volume I: Founding Documents."

[39] Kimball, "The Wassenaar Arrangement at a Glance."

3. Exploring the AWS solution space

The previous section addressed the most pressing problems in existing forms of regulation of AWS, this section will discuss a number of solutions proposed thus far. These have been classified into (1) framework-, (2) law-, and (3) technology-based solutions (see Figure 1). Each solution is explained briefly, the advantages and drawbacks are highlighted, and the actors and capabilities required for implementation are mentioned. Notably, the framework approach provides overarching concepts that can be formalized more concretely through the legal and/or technical solutions. These solutions address the regulation of AWS at different stages of the life cycle. In section 4, this approach will feature explicitly as the solutions below will be clustered along the AWS life cycle of production, proliferation, deployment, and use.



Framework approach	Legal approach	Technological approach
Meaningful Human Control	Total ban	Codes of conduct
Accountability	Legal review of weapons	Explainable AI
	Compensating wrongful outcomes	Ethical goal functions
	Legal personhood	Ethical Governor
	Regulating contractors	

Figure 2. Solution portfolio

3.1 Framework-based solutions

3.1.1 Implementing 'meaningful human control' over AWS

'Meaningful human control' (MHC) is a proposed solution aimed at addressing several regime gaps at once, and one which has thus far seen the widest support from the international community.[40] International discussion is still ongoing concerning whether MHC refers to a wider process of control over weapon systems, or just to control taking place at a certain stage, e.g., during the targeting phase.[41] The wider life cycle approach is gaining traction, as the term 'meaningful' comes to include the design and development phases of autonomous systems.[42] This approach

[40] The term has also been referred to as the human intervention, appropriate level of human judgement by the US, and the intelligent partnership by the UK. See Ekelhof, "Complications of a Common Language: Why It Is so Hard to Talk about Autonomous Weapons"; Moyes, "Key Elements of Meaningful Human Control."

[41] Kayser and Denk, "Keeping Control."

[42] Ekelhof, "Complications of a Common Language: Why It Is so Hard to Talk about Autonomous Weapons"; Boothby, "Conflict Law: The Influence of New Weapons Technology, Human Rights, and Emerging Actors."

is supported by technological solutions (section 3.3) and the institutionalization of accountability (section 3.1.2). It appears most likely that MHC will take the form of 'if–then' rules, for example indicating necessary levels of human–machine interaction ordered by decreasing levels of human control and increasing levels of machine control.[43] One option to formalize MHC is through the creation of an additional protocol to the Convention on Certain Conventional Weapons (CCW).[44] This approach was put on the agenda predominantly by smaller and middle powers and by civil society organizations, and it takes a top-down, state-based approach to regulating AWS.

Not necessarily tied to the manner of implementation, international organizations – including International Committee for Robot Arms Control (ICRAC), Article 36, ICRC and NATO – have identified key elements of importance to the clarification of the concept of MHC. In Table 1, we have listed three elements that are most commonly mentioned: (1) the adjustment of system functions, (2) the understanding of the system, and (3) the improvement of the situational awareness. Importantly, all three elements point towards the interaction between the human and the machine, to ensure that human intervention is meaningful.

Key elements	ICRAC[45]	Article 36 (org)[46]	ICRC[47]	NATO[48]
(1) System functions	Capability for "the rapid suspension or abortion of the attack"	Predictable, reliable, and transparent technology	Predictable system that allows for human intervention	Output presented in a comprehensive and meaningful manner
(2) System understanding	Ability to identify and "react to any changes or unanticipated situations" that alter the analysis,	Understanding "what type of objects a weapons system will identify as target profiles"	Sufficient information and understanding of the weapon	Access to and understanding of the sources of information

[43] Tamburrini and Amoroso, "What Makes Human Control over Weapons Systems 'Meaningful'?"; Sharkey, "Towards a Principle for the Human Supervisory Control of Robot Weapons," 11.

[44] Tamburrini and Amoroso, "What Makes Human Control over Weapons Systems 'Meaningful'?"

[45] Sharkey, "Guidelines for the Human Control of Weapons Systems," 4.

[46] Moyes, "Key Elements of Meaningful Human Control.," 4.

[47] "Expert Meeting on Lethal Autonomous Weapons Systems."

[48] Roorda, "NATO's Targeting Process: Ensuring Human Control over and Lawful Use of 'autonomous' Weapons."

	objective, or legitimacy of targets	and how kinetic force would be applied	system	
(3) Situational awareness	“Full contextual and situational awareness of the target area at the time of a specific attack”	Clear military objective and definition of (potential) unintended consequences	Understanding of the operating environment and its interaction with the system	

Table 1. Key elements to defining 'meaningful human control'

A few issues persist with the concept of MHC. Firstly, due to cognitive limitations humans are prone to automatic reasoning that neglects inconsistencies and are biased to confirm, thus in practice leading to human oversight instead of control.[49] Secondly, the increasing speed of military operations calls into question whether there is in practice sufficient time for a human to review the algorithmic recommendations in terms of necessity, objective, and proportionality – again pointing to the potential importance of a life-cycle-based approach to MHC.[50]

3.1.2 Establishing AWS accountability frameworks

A holistic, life-cycle-based approach gives room for consideration of all elements that go into creating, using and regulating weapon systems, as well as for evaluation feedback loops when needed. Throughout procurement processes, governments should ensure that the concept of morally responsible engineering is applied in the design phase and that weapons are extensively tested under realistic conditions. It is also important to ensure that training programs for military personnel – in particular commanders – devote attention to ethical issues relating to the deployment of autonomous weapons. At the same time, companies working on the development of relevant technologies should have high-level ethical standards translated into practical requirements for programmers and others involved in all stages of development of weapon systems.

[49] Sharkey, “Towards a Principle for the Human Supervisory Control of Robot Weapons,” 8–9; Schwarz, “The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems”; Sharkey, “Guidelines for the Human Control of Weapons Systems.”

[50] Moyes, “Key Elements of Meaningful Human Control.”; Sharkey, “Guidelines for the Human Control of Weapons Systems”; Schwarz, “The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems.”

An example of an approach that could serve as a model to establish an accountability framework is a suggestion made by Australia's delegation to the CCW GGE meeting in March of 2019 as the country's "system of control."[51] Australia is of the opinion that the term MHC as used in most of these GGE meetings "does not adequately cover the plethora of practical controls or systems utili[z]ed by states and their militaries."[52] The system Australia proposed, as seen in Figure 2, addresses responsibilities in eight phases, from the policy and legal framework down to the final use of force. This approach also allows for specific controls and the actors responsible for them to be tailored to different AWS and their unique operating environments. This wider system of control resembling a whole-of-decision-chain approach gives room for consideration of all elements that go into creating, using and regulating these weapon systems, as well as for evaluation feedback loops when needed, allowing for clarity on *who* may be responsible for *which* tasks or requirements, and at *what* stages of the cycle.[53]

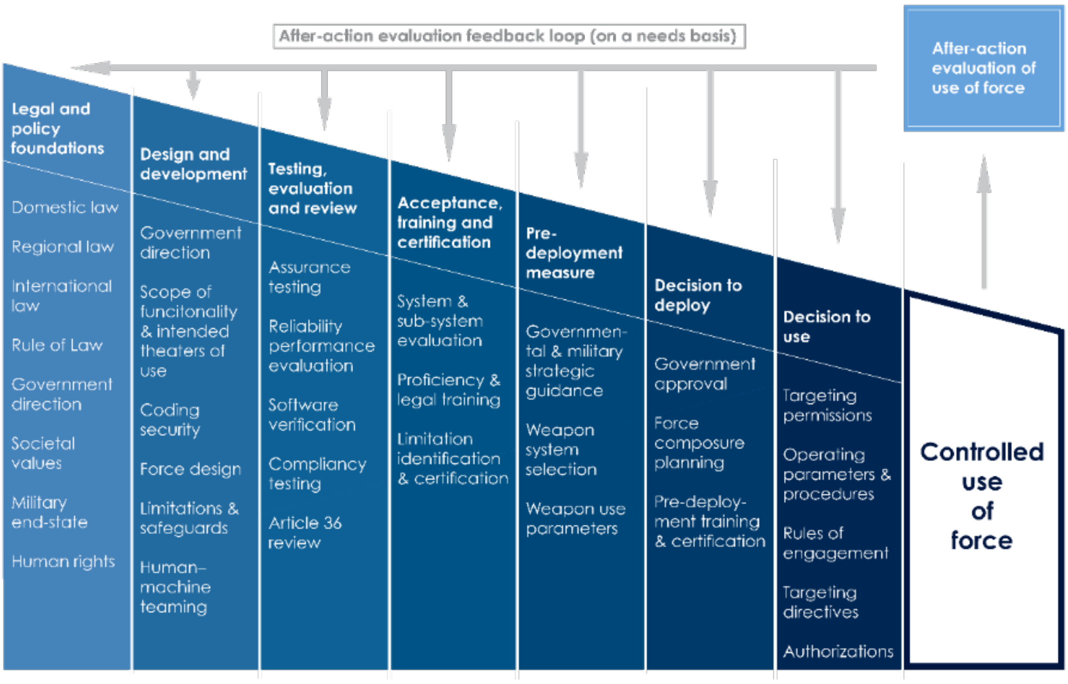


Figure 2. Australia's system of control[54]

[51] "Australia's System of Control and Applications for Autonomous Weapon Systems."

[52] "Australia's System of Control and Applications for Autonomous Weapon Systems," 5.

[53] "Australia's System of Control and Applications for Autonomous Weapon Systems"; Verdiesen, "Agency Perception and Moral Values Related to Autonomous Weapons"; Marra and Mcneil, "Understanding 'The Loop': Regulating the Next Generation of War Machines."

[54] "Australia's System of Control and Applications for Autonomous Weapon Systems."

3.2 Legal solutions

3.2.1 Implementing a total ban on AWS

A **total ban on AWS is one of the solutions proposed to effectively prevent the development and/or use of AWS.** Human Rights Watch, as part of the "Stop Killer Robots" campaign, that a prohibition of 'lethal AWS' or LAWS is necessary to ensure that "firing decisions are made by humans," who, they claim, can make more accurate and humane judgements in warfare.[55] The implementation of a ban could solve the issues with the Martens Clause (see section 2.1.2), the lack of accountability (section 2.3), or the use of force (section 2.2) in relation to AWS. The notion of a ban, however, has recently been featured less prominently on the agenda of the GGE meetings. As highlighted by Paul Scharre, "someday machines might outperform humans in warfare,"[56] which is why some say a ban on AWS could prevent the possibility of more accurate and lawful targeting decisions.[57] The implementation of a full ban on AWS has received support of the international civil society, but would require a wider-spread agreement of states to a new treaty.

3.2.2 Improving the legal review of weapons (Article 36)

Article 36 is an existing provision calling upon states to implement a legal review of weapons – a mechanism that aims to address the legality of AWS at the early stages of design and development. The advantage of the Article 36 mechanism is that weapons are reviewed with consideration of the principles of IHL as well as any other relevant international obligations, thus providing a holistic approach to the implementation of a lawful conduct in war.[58] However, Article 36 requires institutionalization and formalization of the legal review in domestic procedures, such as in military manuals. Nevertheless, the ICRC highlights that the review of legality in design and development of AWS may adequately address faulty or biased decisions.[59] Thus, it would be necessary to call upon more states to institutionalize the

[55] Docherty, "Losing Humanity," 46.

[56] Scharre, "Human Judgement: Lethal Decision-Making in War."

[57] "Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centered Approach."

[58] Article 36 - New weapons - Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 36.

[59] The strength and accuracy of AI-based technologies depends largely on the quality of available and usable data for algorithm training. In practice in other fields outside of the military realm, time and again it turns out after the roll-out of a new system that certain biases in data used to train the algorithms in question shine through in the outcomes and the decisions made. Ultimately, "the behavior of the AWS is determined by human-written software", thus regulating the design hopes to guarantee the compatibility of technology with the IHL principles, see McFarland, "Autonomous Weapons and Human Control"; Backstrom and Henderson, "New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Review"; Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States"; Deeks, "Predicting Enemies."

legal review in order to ensure a widespread practice.[60] International civil society may play an important role in this process. At the same time, states that already have conducted legal reviews should share practices regarding the review of AI-based systems.

3.2.3 Drawing on tort law to compensate wrongful outcomes of AWS

This approach responds to the need to attach responsibility to unlawful acts where there is no clear evidence of harmful intent, which may especially be the case with complex autonomous systems.[61] Professor Rebecca Crootof uses the term 'war torts' to refer to the use of civil liability strategies to regulate wrongdoings in warfare committed by AWS where the intent is not easily deductible.[62] This solution could form a complementary legal regime to fill in the gaps left by international criminal law and to "hold states accountable for the injurious wrongs that are the side effects of employing [AWS]."[63] A practical option for implementation could be a system by which fully autonomous lethal weapon systems fall under a 'strict liability regime', meaning that for an ultra-hazardous or dangerous activity defendants can be held liable even in absence of fault, while semi-autonomous and non-lethal AWS are addressed by negligence standards.[64]

The tort law approach responds to already committed wrongdoings and is therefore reactive in nature, but if enforced adequately may nevertheless affect states' considerations in the stage of planning attacks. Notably, this approach does not constrain innovation in the field of military technology. In order to implement this approach, it would be necessary to first establish exactly where a gap of adequate response to wrongdoing exists, then to translate the legal strategies of tort law into international law. This would require great cooperation between international actors, including states, the United Nations and civil society.

3.2.4 Giving autonomous systems legal 'personhood'

Giving autonomous systems some form of legal 'personhood' may aid accountability and remedy harmful outcomes, in a similar way to 'corporate

[60] Kayser and Denk, "Keeping Control."

[61] The simplest explanation of this concept of civil liability, also known as tort law, could be as follows: If you are backing out of a parking space and accidentally collide with another car that is driving by, you had no criminal intent to damage someone else's life or property, but you still caused legal injury. From this follows that you will need to pay damages according to your responsibility for the outcome.

[62] Crootof, "War Torts: Accountability for Autonomous Weapons."

[63] Crootof, 1348.

[64] Fuzaylova, "War Torts, Autonomous Weapons, and Liability."

personhood'. A key difference between corporations and AI-based systems, of course, is that unlike is the case for AWS, a corporation cannot 'do' anything without human agents acting on its behalf.[65] Putting this potential solution into practice, an example has come from the European Parliament, which (albeit mainly in the context of civil use) proposed assigning a certain form of legal or 'electronic' personhood to robots – with mixed responses, it should be said.[66] The aim is in essence to make it possible to establish "a causal link between the harmful behavior of the robot and the damage suffered by the injured party" that is sufficient to claim compensation.[67] States would need to enact new legislation, either at a regional or even at an international level and additionally companies and individuals would indirectly become involved if AWS causes a wrongful outcome that requires compensation.

3.2.5 Regulating private contractors

Implementing principles from international law into contracts can allow states to regulate private contractors involved in the development, deployment and use of AWS. This requires constant involvement from a wide variety of private actors of which regulation is otherwise inhibited for due to jurisdictional issues. An existing example is the Montreux Document, an agreement that specifies obligations for signatories regarding private military and security companies in war zones.[68] It has been suggested that it could be relevant also to improve regulation of AWS to a certain extent in a similar, contract-law-based manner.[69] One issue with this solution is that it may further incentivize denial or trivialization of potential war crimes, due to the impersonal and heavily legalistic nature of the language used in contract law.[70]

3.3 Technological solutions

3.3.1 Back-end technological solutions

3.3.1.1 Codes of conduct on an engineering level

Organization-level Codes of Conduct can ensure that high-level ethical considerations are applied in practice by the people designing and developing AWS-related technologies. The Institute of Electrical and Electronics Engineers (IEEE) has

[65] Scherer, "Op-Ed."

[66] Committee on Legal Affairs, "Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))"; Chavannes and Arkhipov-Goyal, "Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context."

[67] Bryson, Diamantis, and Grant, "Of, For, and By the People: The Legal Lacuna of Synthetic Persons."

[68] "The Montreux Document - On Pertinent International Legal Obligations and Good Practices for States Related to Operations of Private Military and Security Companies during Armed Conflict."

[69] Liu, "Contract Law as Cover;" Dickinson, "Contract as a Tool for Regulating Private Military Companies."

[70] Dickinson, Outsourcing War and Peace, 69–101; HCSS RAS paper p57.

presented a set of principles for 'ethically aligned design'.[71] When it comes to AWS, an important conclusion is that existing ethics codes or codes of conduct may fail to address (moral) responsibility for autonomous systems or may fail to clarify engineers' obligations in this regard. For example, many AI-based programs currently suffer from a lack of regard for the flaws in the data that is used to train algorithms, resulting in discriminatory or simply low-quality outcomes that challenge the legality of the AWS deployment. As defense policymakers establish clear requirements for outsourcing technology, contracting parties should review, revise, and/or extend their codes of ethics with respect to AWS in order to apply these requirements in practice to all those involved within their respective companies.[72]

3.3.1.2 Implementing 'explainable AI' methods

'Explainable AI' is a technological solution that quantitatively justifies algorithmic outputs to improve human operators' understanding of the autonomous systems.[73] For example, facial recognition software could inform an operator that an identified object "looks like a face" with 99.9% certainty.[74] This technical understanding of the system allows for the strengthening of meaningful oversight and/or control of the autonomous weapons (see also section 3.1.1 on MHC). This method, originating from the Silicon Valley, is already being implemented into military technologies by the Defense Advanced Research Projects Agency (DARPA).[75] This solution requires cutting-edge technological development. Additionally, an institutionalized top-down approach would be needed in order to formalize the requirements for military technologies.

3.3.1.3 Giving AWS ethical goal functions

Emphasizing the use of goal functions that are interpretable by a machine, outcomes can be produced within a clear framework of legal, ethical, and military guidelines. Researchers of AI, and artificial general intelligence (AGI) in particular, are working towards this by applying 'orthogonality' in military-use algorithms. Put simply, this refers to the principle that "more or less any level of intelligence could be

[71] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design."

[72] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 118.

[73] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, "Ethically Aligned Design."

[74] Kelion, "Google Tackles the Black Box Problem with Explainable AI."

[75] See Salmanowitz, "Explainable AI and the Legality of Autonomous Weapon Systems." DARPA is an agency of the United States Department of Defense responsible for the development and research of the military technologies.

combined with more or less any final goal."[76] The intelligent system must never be allowed to change this function to set its own goal functions, however. This way it is ensured that the ethical decisions on framework and final or intermediate goal setting are always in the hands of humans. This approach aims to combat the reality that a modern military operation is "a distributed and partially automated process, often difficult to oversee."[77] As with the previous solution, the developers of technology are at the heart of this and these are mostly private actors and companies. At the same time, knowledge institutions have a large part in researching this option and making the use of the orthogonality principle apparent for military technology contexts, making partnerships between policymakers, knowledge institutes and private actors crucial. With recent European emphasis on explainable and trustworthy AI, researchers may look to this option specifically as a form of 'augmented utilitarianism'.[78]

3.3.2 Implementing an 'Ethical Governor'

The 'Ethical Governor' is a technological solution that implements a two-step process of verification of the legality of actions by an independent algorithm, instead of changing the design of military technology.[79] Essentially, the 'Ethical Governor' algorithm checks, first, whether a decision meets the requirements of IHL and, second, if it does, whether it responds directly to the operational orders. If the answer is negative to either of the two questions, the system would prevent an AWS from locking on to a target. This solution has been argued to possibly become more effective than humans in the assessment of legality of algorithmic decisions, since it is faster and is capable to understand and evaluate the algorithmic processes.[80] It is also an easily scalable solution to a variety of other, non-weaponized autonomous systems used in military decision-making and targeting processes. It could also be an addition to human control of AWS. However, the same difficulty of translating the legal principles to an algorithmic code prevails (see more in section 1.1.1 on IHL principles).[81] Therefore, this solution requires additional research into the technological possibilities and limitations before the relevant actors can implement it domestically.

[76] Bostrom, "The Superintelligent Will."

[77] Elands et al., "Governing Ethical and Effective Behaviour of Intelligent Systems," 213.

[78] Aliman and Kester, "Requisite Variety in Ethical Utility Functions for AI Value Alignment."

[79] Docherty, "Losing Humanity"; Petman, *Autonomous Weapons Systems and International Humanitarian Law*.

[80] Robert Arkin, Patrick Ulam, and Alan Wagner, "Moral Decision-making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust and Deception", *Georgia Institute of Technology* (n.d.).

[81] Docherty, "Losing Humanity."

3.4 Summary and feasibility assessment of key findings from literature review

3.4.1 Framework-based solutions

Whether oversight and control of AWS is implemented in the form of MHC or not, there is a need to implement a wide, life-cycle-based approach in order to assure both control and accountability from the policy and development stages up until – and even after – the use of AWS. By and large, there is consensus especially among international civil society representatives that interactions between humans and machines must be accompanied by an understanding of system functions and outcomes. Facilitating understanding requires in-depth training and re-training of military personnel regarding systems' decision-making processes; institutionalized communication channels between policymakers, end users, developers (especially external parties), and legal advisors to ensure transparent translation of legal and ethical standards into technological parameters; and sharing of best practices regarding the aforementioned implementation and translation of IHL principles into AWS functions.

3.4.2 Law-based solutions

Overall, the conclusion from the various legal challenges and solutions put forward in this paper do not necessarily point towards strict legalism and the creation of new hard law as the main way forward with AWS. This is also because much of the technologies' possibilities are still at least partially unknown. This leads, first, to definitional unclarities that hamper effective and precise lawmaking, and second, to resistance to additional international legal agreements due to states' fears they would inhibit innovation. The main solutions identified in this section were a ban on AWS proliferation and use, and possibly also on development; improved legal reviews of new military technologies; the adaptation of tort and/or contract law mechanisms to fill certain gaps in accountability through current application of the legislation applicable AWS; and the option of giving autonomous systems some form of legal 'personhood'.

Out of these identified solutions, the most feasible option would appear to be a pathway through improvement of existing regimes. Most effective methods of improving both oversight and control are likely to come in a hybrid form: in the time that it takes to develop the necessary trust to come to agreements on regulating AWS, states can be made to comply with mechanisms for compensation of wrongful outcomes.

3.4.3 Technology-based solutions

Technology-based solutions for now are largely aimed at making explicit the consideration of ethical and IHL requirements or principles at the design and engineering stages. The intended result is to indirectly improve the systems' performance, by increasing understanding on the military user's end and optimizing human-machine teaming. The central importance of the engineering end of military technology needs closer consideration in order to ensure that a potential AWS is compliant with IHL. Standardizing a form of explainable AI, coupled with trainings of military personnel, is necessary in order to diminish the role of biases and to ensure that misunderstandings of the system do not lead to unwanted outcomes. Two considerations for technological solutions to complex ethical problems such as the deployment of AWS, are, first, that they cannot stand alone, and second, that they can only emerge in the context of a broader (societal, political, professional) culture that places emphasis on the principles behind these solutions.

4. Solutions and their place in the AWS life cycle

The solutions proposed in the previous section do not stand alone, but rather, most of them would add in a complementary fashion onto an existing patchwork of regimes that apply to AWS. Different (constellations of) solutions may be most effective, depending on the actors involved their motivation to curb the development or use of AWS – or lack thereof. In explaining the solution space, it was also indicated what types of actors would be involved in the solution in question, and in what way it would affect the life cycle of AWS. A way to now cluster these solutions comprehensively is to align them along the cycle of production, proliferation, deployment, and use phases of AWS, indicating visually in which of these four stages each certain solution aims to improve oversight and/or control. This also allows for identification of potential gaps.

The solutions proposed in the literature can be aligned along this life cycle. This assessment by our experts is based on the literature review, and is visualized in Figure 3 below. MHC and accountability are framework-based approaches, shown in blue. Legal solutions are shown in yellow. Technology solutions are shown in green. Solid-colored bars indicate that a solution certainly applies in a particular phase, whereas the half-colored/striped bars indicate that a solution potentially affects a particular phase.

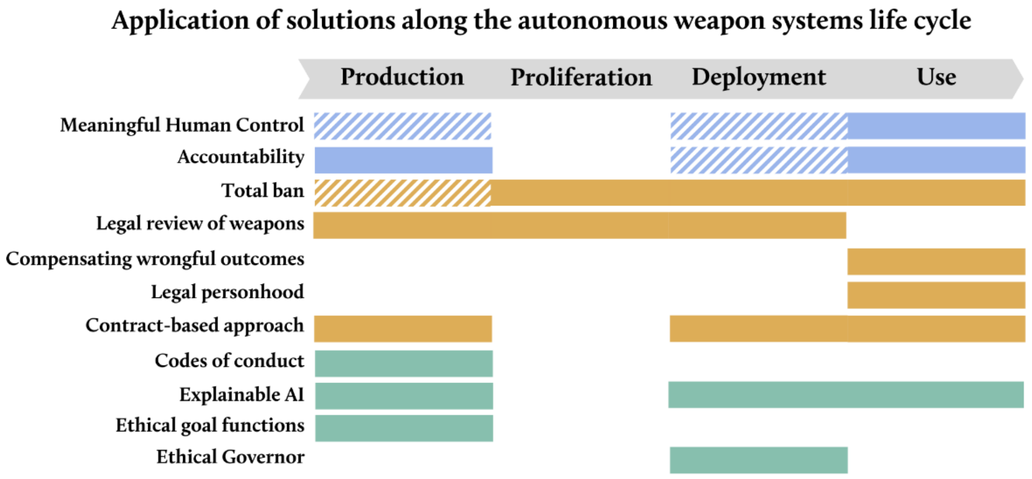


Figure 3. Application of solutions along the autonomous weapon system life cycle

Sections 2 and 3 laid out current thinking on potential ways to govern the development and use of AWS. They serve as food for thought for an open discussion on where new approaches are needed and possible, as well as what shape or form of these approaches or new regulation might be most effective when trying to address the opportunities and threats that arise from emerging technologies. Figure 3 already indicates, for example, an apparent gap in the solution space when it comes to regulating proliferation in particular. Building on this notion of food for thought, an expert session was organized in order to assess the literature's solutions and address any gaps in the understanding of this topic. The next section lay out the key takeaways from that expert session.

5. Key insights from the AWS expert session

It is time to move beyond broad, high level, conceptual discussions on to more operational and actionable insights. To this effect, from the literature analysis of current regimes surrounding AWS and the expert session, the authors of this paper have identified a number of salient insights. In the textbox below the set-up of the expert workshop is described. In the rest of this section, the most important insights have been laid out as five key takeaways relevant to policymakers, in particular those working on AWS governance at the crossroads of security and emerging technology in the Netherlands' ministries of Foreign Affairs, Defense, and Economic Affairs, and in general those working elsewhere in the international relations and defense sectors.

Expert session methodology

The expert session held at *The Hague* Centre for Strategic Studies (HCSS) on December 2nd, 2019 involved approximately twenty legal, military, political, and technological experts from ministries, armed forces, NGOs, industry, and academic and research institutions. Participants were divided into groups, each diverse in terms of academic background and experience. The expert session was based on a moderated but open discussion with two leading questions. First: are the solutions described in the original position paper (here sections 2 through 4) the only options or are there other solutions that we can envisage? Second: how do potential solutions align with the incentive structures of key actors, and which potential solutions would appear most feasible? A detailed outline of the questions used to guide the discussions is included in Annex 1. After an analysis by our in-house experts based on the conclusions from these discussions, sections 5 and 6 of this paper illustrate the main insights and subsequent recommendations.

5.1 Existing regulation suffices but requires tailored application

Existing regulations are considered to provide sufficient depth to regulate emerging military technologies. This sentence, of course, requires further specification. According to many participants, relevant legal frameworks that pertain to these technologies do exist. In other words, AWS do not exist in a legal vacuum. They are regulated by IHL and other international and domestic laws concerning, for instance, non-proliferation, international technological standards, or export controls. In addition, many international regulations already focus on the potentially unforeseen effects of new (technological) developments, rather than on specific technological applications. However, there is certainly a need for topic-specific clarification of how the existing laws apply or should be applied to many emerging technologies, including those relevant to the military domain. Therefore, rather than focusing on a supposed lack of regulation, further efforts are required to clarify how existing agreements apply specifically to AWS and the technologies of which they are composed.

5.2 Opportunities lie in improving compliance with existing regulations

Alongside the need for clarification of existing rules, countries should place an emphasis on promoting application of and compliance with these rules. The enforcement of international agreements continues to be a problem. Article 36's legal review of new weapons is the only binding mechanism of its kind, and while such legal reviews are relatively widely regarded as an international principle, Article 36 has been formalized only in a few states. On the one hand, this is due to the lack of clear guidelines to the implementation of international principles into AWS. Even states who have implemented the legal review of weapons, such as the Netherlands, struggle to adequately adjust the mechanism to meet the particularities of AWS. On the other hand, there is a lack of political will. As a result, non-compliant states proceed with the research and production of AWS while other countries find themselves at a political as well as a strategic disadvantage. Thus, ensuring wider-spread compliance with principles of new weapon system reviews should be one of the main objectives in the regulation of AWS. In this process, forming alliances or establishing 'coalitions of the willing' is instrumental to the promotion and enforcement of existing agreements and principles.

5.3 The role of private actors creates need for non-legal solutions

The dependence on private actors in the field of emerging technologies highlights the need for non-legal solutions that offer guidance beyond the governing of state behavior. For example, technical solutions, such as explainable AI and an Ethical Governor, may be deemed useful or even necessary to improve the verifiability and reliability of weapon systems. In addition to such approaches in the deployment and use phases of a system's life cycle, it is also important to implement other non-legal solutions in the first stages of the research and development of military technologies – especially when those stages tend to be outsourced. The objective is to ensure that legal knowledge and military needs are adequately translated into the parameters set in the technology that goes into an AWS. Specifically, the use of Codes of Conduct (CoCs), context-specific manuals listing technological requirements according to military objectives, are expected to provide greater clarity in the application of international standards. Furthermore, CoCs should state the responsibilities of the parties involved in the military (weapon) system's life cycle and promote transparency as well as cooperation regarding system development, improvements, and potential operational shortcomings.

5.4 National standards can be promoted to improve international standards

The Netherlands and likeminded countries could work together more on domestically setting in place standards and doctrines for the production of technologies relevant to AWS. These can then be translated into international principles for the production and use thereof. It is herein key to reinforce and stimulate an international epistemic community, in order to facilitate the exchange of technical and legal knowledge and inform international policymaking with expertise. Such a standard-setting approach could be taken, for instance, through ensuring sufficient participation on product standards through the UN's International Telecommunication Union (ITU). At the moment, international standards on technology and certification are not necessarily set by countries with the same approaches or ethical standards. Currently, China is leading the ITU's drafting of facial recognition standards. This is relevant, as these international standards may feed into the AWS life cycle – especially given defense sectors' reliance on private actors and internationally spread-out players for the development and production of their systems. Taking such an approach that addresses the early stages of production life cycles is a way through which to attain a universal application of existing international regimes and law and to further ethical considerations in technology development.

5.5 Technological complexity requires improved technical literacy

As regulating emerging technologies becomes increasingly complex, there is a widening knowledge gap between technical and technological experts and policymakers. This can in part be attributed to, for example, the outsourcing of research and production of military technologies to private actors. The understanding of sophisticated systems and their design is crucial, both to policymakers in order to cope with the complexity and to enhance the specificity of existing regulations, and to military personnel in order to interact with the systems in an informed manner and prevent a lack of accountability. A solid understanding of the what constitutes AWS and what may therefore pose a threat or present an opportunity, will make rules-based as well as 'softer' approaches to regulating these systems more relevant, applicable, and effective.

6. Recommendations

Building on the analysis of proposed solutions and the five key insights from the discussions during the expert session, this paper provides a number of concrete recommendations for policymakers working on AWS governance at the crossroads of security and emerging technology, in the ministries of Foreign Affairs, Defense, and Economic Affairs. The relevance for these three groups of policymakers is respectively arms control and international cooperation on related technology, the development of these technologies within legal and ethical boundaries, and the development of relevant industry standards. The insights and recommendations presented here aim to move discussions forward, from conceptual framing toward actionable steps.

At the domestic level:

1. *Ensure and safeguard legal compliance* – review compliance with, and publish and regularly update guidelines to, the applicability of international humanitarian law (IHL) to AWS and related technology, with particular attention to the early stages of the life cycle, in order to set domestic standards;
2. *Set standards and protocols* – define the purpose and goals of acquired and manufactured military technology and set up guidelines for context-specific Codes of Conduct for private actors. Identify desired and undesired outcomes, as well as (technological) solutions to improve the verifiability and explainability of these outcomes;
3. *Tailor contractor work to military needs* – continuously test and re-evaluate the pre-determined settings and capabilities of a system, in order to ensure that manufacturing is tailored to the use environment;
4. *Improve technological literacy of policymakers* – ensure adequate and recurrent cooperation with researchers and military personnel who are spearheading development, innovation, and application. Where necessary, consider bringing in new employees with technical expertise in order to increase understanding of the technological conditions for policies;
5. *Improve engagement of and with military personnel* – cooperate closely with private contractors and manufacturers of AWS as well as policymakers, in order to ensure interaction between legal frameworks, technological standards, and military rules of engagement.

At the international level:

6. *Promote legal compliance* – clarify application of existing international regimes to AWS and related technologies, and promote widespread compliance among states, especially concerning the Article 36 legal review of weapons. Initiatives

such as the UN GGE's 11 Guiding Principles can provide a starting point in how this can be done;

7. *Lead by example* – lead the implementation of domestically developed doctrines and standards into international standards for the production and use of AWS. This furthers ethical considerations in technology development as well as standards that align with existing international laws and norms;
8. *Improve transparency* – share with other states and with national publics the insights gained from setting standards and protocols and applying regulations around AWS and related technology, in order to improve international harmonization and understanding;
9. *Strengthen cooperation on standard-setting* – establish a 'coalition of the willing' among states to share practices, promote standards for new technologies relevant to the military field, and share the burden of expenses related to the validation and verification of the technologies that go into AWS;
10. *Stimulate epistemic communities* – reinforce and stimulate an international epistemic community to facilitate the exchange of technical and legal knowledge and inform international policymaking with expertise, and build bridges between research and industry to translate legal and ethical rules and norms in technology standards and protocols.

Annex 1. Questions guiding expert group discussions

Solutions: discuss solutions in literature and input paper, and envisage new solutions

1. What do we mean by "regulation"? What do we want to prevent/avoid?
2. How do we want to achieve this effect? What is the best approach? (Framework/legal/technological or a combination)
3. Are the solutions mentioned in the literature review satisfactory?
4. Are there any solutions that were not mentioned in the position paper that you have come across or want to bring to the table?

Assessment: political and technological feasibility

1. Think back to the solutions along the system life cycle – which steps of the cycle necessitate human control? Which ones require human oversight?
2. Which steps of the system life cycle are not covered by the solutions we discussed? Is it a problem? Can we think of other ways to regulate them during this stage?
3. Are these solutions technically feasible? If not, what is required to make them feasible?

Implementation: analysis of the implementation process

1. Who are the relevant actors? What are their motivations and concerns?
2. What is the feasibility of each solution?
3. What are the necessary elements to implement the proposed solutions?

Bibliography

France Diplomatie Ministry for Europe and Foreign Affairs. “11 Principles on Lethal Autonomous Weapons Systems (LAWS),” n.d. https://www.diplomatie.gouv.fr/en/french-foreign-policy/united-nations/alliance-for-multilateralism-63158/article/11-principles-on-lethal-autonomous-weapons-systems-laws.

Aliman, Nadisha-Marie, and Leon Kester. “Requisite Variety in Ethical Utility Functions for AI Value Alignment.” In *ArXiv:1907.00430 [Cs]*. Cornell University, 2019. http://arxiv.org/abs/1907.00430.

Amoroso, Danielle. “Jus in Bello and Jus Ad Bellum Arguments against Autonomy in Weapon Systems: A Re-Appraisal.” *Questions of International Law Journal*, October 31, 2017. http://www.qil-qdi.org/jus-bello-jus-ad-bellum-arguments-autonomy-weapons-systems-re-appraisal/#_ftnref108.

Article 36 - New weapons - Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (1977). https://ihl-databases.icrc.org/ihl/WebART/470-750045?OpenDocument.

International Committee of the Red Cross. “Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centered Approach,” June 6, 2019. https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach.

“Australia’s System of Control and Applications for Autonomous Weapon Systems.” Geneva, 2019. https://www.unog.ch/80256EDD006B8954/(httpAssets)/39A4B669B8AC2111C12583C1005F73CF/$file/CCW_GGE.1_2019_WP.2_final.pdf.

the United Nations News. “Autonomous Weapons That Kill Must Be Banned, Insists UN Chief,” March 25, 2019. https://news.un.org/en/story/2019/03/1035381.

Backstrom, Alan, and Ian Henderson. “New Capabilities in Warfare: An Overview of Contemporary Technological Developments and the Associated Legal and Engineering Issues in Article 36 Weapons Review.” *International Committee of the Red Cross* 94/886 (2012).

Beenes, Maaike, Frank Slijper, Alice Beck, and Daan Kayser. “Don’t Be Evil? A Survey of the Tech Sector’s Stance on Lethal Autonomous Weapons.” *PAX for Peace*, August 19, 2019.

Boothby, William. “Conflict Law: The Influence of New Weapons Technology, Human Rights, and Emerging Actors.” *Geneva Centre for Security Policy*, 2014.

Bostrom, Nick. “The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents.” *Minds and Machines* 22, no. 2 (May 2012): 71–85. https://doi.org/10.1007/s11023-012-9281-3.

Boulanin, Vincent, and Maaike Verbruggen. “Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies.” Solna, Sweden: SIPRI, December 2017. https://www.sipri.org/sites/default/files/2017-12/article_36_report_1712.pdf.

Bouvier, Antoine A. “International Humanitarian Law and the Law of Armed Conflict.” Edited by Harvey J. Langholtz. Peace Operations Training Institute, 2012. http://cdn.peaceopstraining.org/course_promos/international_humanitarian_law/international_humanitarian_law_english.pdf.

Bromley, Mark, and Giovanni Maletta. “The Challenge of Software and Technology Transfers to Non-Proliferation Efforts,” n.d.

Bryson, Joanna, Mihailis Diamantis, and Thomas Grant. “Of, For, and By the People: The Legal Lacuna of Synthetic Persons.” *Artificial Intelligence and Law* 25, no. 3 (September 30, 2017): 273–91.

“Charter of the United Nations.” United Nations, June 26, 1945. https://www.un.org/en/charter-united-nations/.

Chavannes, Esther, and Amit Arkhipov-Goyal. “Towards Responsible Autonomy: The Ethics of Robotic and Autonomous Systems in a Military Context.” The Hague: The Hague Centre for Strategic Studies, September 2019. https://www.hcss.nl/sites/default/files/files/reports/Towards%20Responsible%20Autonomy%20-%20The%20Ethics%20of%20RAS%20in%20a%20Military%20Context.pdf.

Committee on Legal Affairs. “Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).” European Parliament, May 31, 2016. http://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.

“Country Views on Killer Robots.” The Campaign to Stop Killer Robots, October 25, 2019. https://www.stopkillerrobots.org/wp-content/uploads/2019/10/KRC_CountryViews_25Oct2019rev.pdf.

Crootof, Rebecca. “War Torts: Accountability for Autonomous Weapons.” *University of Pennsylvania Law Review* 164 (May 2016): 56.

Deeks, Ashley. “Predicting Enemies.” *Virginia Law Review*, Virginia Public Law and Legal Theory Research Paper, 104 (March 1, 2018). https://papers.ssrn.com/abstract=3152385.

Deeks, Ashley, Noam Lubell, and Daragh Murray. “Machine Learning, Artificial Intelligence, and the Use of Force by States.” *Journal of National Security Law & Policy* 10, no. 1 (2019): 1–25.

Docherty, Bonnie. “Losing Humanity: The Case against Killer Robots.” Human Rights Watch, 2012.

Ekelhof, Merel. “Complications of a Common Language: Why It Is so Hard to Talk about Autonomous Weapons.” *Journal of Conflict and Security Law* 22, no. 2 (2017): 311–31. https://doi.org/10.1093/jcsl/krw029.

Elands, P.J.M., A.G. Huizing, L.J.H.M. Kester, S. Oggero, and M.M.M. Peeters. "Governing Ethical and Effective Behaviour of Intelligent Systems: A Novel Framework for Meaningful Human Control in a Military Context." *Militaire Spectator*, 2019.

International Committee of the Red Cross. "Expert Meeting on Lethal Autonomous Weapons Systems." Statement, November 15, 2017. https://www.icrc.org/en/document/expert-meeting-lethal-autonomous-weapons-systems.

Fuzaylova, Elizabeth. "War Torts, Autonomous Weapons, and Liability: Why a Limited Strict Liability Tort Regime Should Be Implemented." *Cardozo Law Review* 40, no. 3 (March 5, 2019). http://cardozolawreview.com/war-torts-autonomous-weapon-systems-and-liability/.

Geist, Edward, and Andrew J. Lohn. "How Might Artificial Intelligence Affect the Risk of Nuclear War?" Perspective: Security 2040. RAND Corporation, 2018. https://doi.org/10.7249/PE296.

Haner, Justin, and Denise Garcia. "The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development." *Global Policy* 10, no. 3 (September 2019): 331–37. https://doi.org/10.1111/1758-5899.12713.

"Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapon Systems. Submitted by the United States of America." Geneva, Switzerland, 2018. https://www.unog.ch/80256EDD006B8954/(httpAssets)/7C177AE5BC10B588C125825F004B06BE/$file/CCW_GGE.1_2018_WP.4.pdf.

International Law Commission. "Responsibility of States for Internationally Wrongful Acts." United Nations, 2001. http://legal.un.org/ilc/texts/instruments/english/draft_articles/9_6_2001.pdf.

Johnson, James. "Artificial Intelligence & Future Warfare: Implications for International Security." *Defense & Security Analysis* 35, no. 2 (April 3, 2019): 147–69. https://doi.org/10.1080/14751798.2019.1600800.

Jones, Bruce, Charles T Call, Daniel Toubolets, and Jason Fritz. "Managing the New Threat Landscape: Adapting the Tools of International Peace and Security." *Brookings Institute*, Foreign Policy at Brookings, September 2018, 23.

Kania, Elsa. "China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems." Lawfare, April 17, 2018. https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

Kayser, Daan. "Killer Robots." PAX for Peace, n.d. https://www.paxforpeace.nl/our-work/programmes/killer-robots.

Kayser, Daan, and Stepan Denk. "Keeping Control: European Positions on Lethal Autonomous Weapon Systems." PAX, October 2017. https://www.paxforpeace.nl/publications/all-publications/keeping-control.

Kelion, Leo. "Google Tackles the Black Box Problem with Explainable AI." *BBC News*, November 24, 2019.

"Killer Robots and the Concept of Meaningful Human Control." Human Rights Watch, 2016. https://www.hrw.org/sites/default/files/supporting_resources/robots_meaningful_human_control_final.pdf.

Kimball, Daryl. "The Wassenaar Arrangement at a Glance." Arms Control Association, December 2017. https://www.armscontrol.org/factsheets/wassenaar.

Konaev, Margarita. "With AI, We'll See Faster Fights but Longer Wars." *Texas National Security Review*, October 29, 2019. https://warontherocks.com/2019/10/with-ai-well-see-faster-fights-but-longer-wars/.

Lawland, Kathleen. "Reviewing the Legality of New Weapons, Means and Method of Warfare." 88/864: International Commission of the Red Cross, 2006.

"LAWS: Ten Problems for Global Security." *International Committee for Robot Arms Control*, April 2015. https://www.icrac.net/wp-content/uploads/2018/03/LAWS-10-Problems-for-Global-Security.pdf.

Marra, William C., and Sonia K. Mcneil. "Understanding 'The Loop': Regulating the Next Generation of War Machines." *Harvard Journal of Law & Public Policy* 36, no. 3 (May 2013): 1139–85.

McFarland, Tim. "Autonomous Weapons and Human Control." Humanitarian Law & Policy Blog, July 18, 2018. https://blogs.icrc.org/law-and-policy/2018/07/18/autonomous-weapons-and-human-control/.

McLean, Wayne. "Drones Are Cheap, Soldiers Are Not: A Cost-Benefit Analysis of War." The Conversation. Accessed June 28, 2019. http://theconversation.com/drones-are-cheap-soldiers-are-not-a-cost-benefit-analysis-of-war-27924.

Melzer, Nils. "Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law." International Commission of the Red Cross, 2009.

Moyes, Richard. "Key Elements of Meaningful Human Control." Geneva: Article 36, 2016. http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf.

Netherlands Advisory Council on International Affairs. "Autonomous Weapon Systems: The Need for Meaningful Human Control." Netherlands Advisory Council on International Affairs, October 2015. https://aiv-advies.nl/8gr#government-responses.

Okano-Heijmans, Maaike, and Hugo Klijn. "Input Paper: Managing RAS: The Need for New Norms and Arms Control." Clingendael Institute & HCSS, 2019.

Petman, Jarna. *Autonomous Weapons Systems and International Humanitarian Law: "Out of the Loop"?* Research Reports. Helsinki: Erik Castrén Institute of International Law and Human Rights, 2018. https://um.fi/documents/35732/48132/autonomous_weapon_systems_an_international_humanitarian_law__out_of_the.

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (1977).

Queguiner, Jean-Francois. "Precautions under the Law Governing the Conduct of Hostilities." International Commission of the Red Cross, 2006.

Roorda, Mark. "NATO's Targeting Process: Ensuring Human Control over and Lawful Use of 'autonomous' Weapons," n.d., 19.

Salmanowitz, Natalie. "Explainable AI and the Legality of Autonomous Weapon Systems." Lawfare - DayZero: Cybersecurity Law and Policy, November 21, 2018. https://www.lawfareblog.com/explainable-ai-and-legality-autonomous-weapon-systems.

Sandoz, Yves, Christophe Swinarski, and Bruno Zimmermann. "Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949." International Commission of the Red Cross, 1987. https://perma.cc/5XKM-QQYV.

Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. WW Norton & Co, 2018.

———. "Human Judgement: Lethal Decision-Making in War." Humanitarian Law & Policy Blog, n.d. https://blogs.icrc.org/law-and-policy/2018/04/11/human-judgment-lethal-decision-making-war/.

Scherer, Matt. "Op-Ed: If AI Systems Can Be 'Persons,' What Rights Should They Have?" Future of Life Institute, July 20, 2016. https://futureoflife.org/2016/07/20/op-ed-ai-systems-can-persons-rights/.

Schuman, Lucy. "Situational Awareness and Adhere to the Principle of Distinction as a Necessary Condition or Lawful Autonomy." In *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security*, edited by Robin Geiß. Berlin: German Federal Foreign Office, 2017.

Schwarz, Elke. "The (Im)Possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems." International Committee of the Red Cross. *Humanitarian Law & Policy* (blog), August 29, 2018. https://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/.

Sharkey, Noel. "Guidelines for the Human Control of Weapons Systems." ICRAC, April 2018. https://www.icrac.net/icrac-working-paper-3-ccw-gge-april-2018-guidelines-for-the-human-control-of-weapons-systems/.

———. "Towards a Principle for the Human Supervisory Control of Robot Weapons." *Politica & Società* 2 (August 2014). https://www.unog.ch/80256EDD006B8954/(httpAssets)/2002471923EBF52AC1257CCC0047C791/$file/Article_Sharkey_PrincipleforHumanSupervisory.pdf.

Shaw, Ian. *Predator Empire: Drone Warfare and Full Spectrum Dominance*. Minneapolis, MN: University of Minnesota Press, 2016.

Soesanto, Stefan. "Trend Analysis: The Evolution of US Defense Strategy in Cyberspace (1988 – 2019)." CSS Cyber Defense Project. Zürich: Center for Security Studies (CSS), ETH Zürich, August 2019. https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Cyber-Reports-2019-08-The-Evolution-of-US-defense-strategy-in-cyberspace.pdf.

Sparrow, Rob. "Ethics as a Source of Law: The Martens Clause and Autonomous Weapons." Humanitarian Law & Policy Blog, November 14, 2017. https://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/.

Tamburrini, Guglielmo, and Daniele Amoroso. "What Makes Human Control over Weapons Systems 'Meaningful'?" ICRAC, August 2019. https://www.icrac.net/wp-content/uploads/2019/08/Amoroso-Tamburrini_Human-Control_ICRAC-WP4.pdf.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. "Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems." IEEE, 2017. https://doi.org/10.1007/978-3-030-12524-0_2.

"The Montreux Document - On Pertinent International Legal Obligations and Good Practices for States Related to Operations of Private Military and Security Companies during Armed Conflict." *International Committee of the Red Cross*, 2009, 48.

Verdiesen, Ilse. "Agency Perception and Moral Values Related to Autonomous Weapons: An Empirical Study Using the Value-Sensitive Design Approach. Masters Thesis." TU Delft, 2017.

Wareham, Mary. "As Killer Robots Loom, A Push to Keep Humans in Control of Use of Force." The Human Rights Watch, January 2, 2020. https://www.hrw.org/news/2020/01/02/killer-robots-loom-push-keep-humans-control-use-force.

"Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies. Public Documents Volume I: Founding Documents." Wassenaar Arrangement Secretariat, February 2017. https://www.wassenaar.org/app/uploads/2019/consolidated/WA-DOC-17-PUB-001-Public-Docs-Vol-I-Founding-Documents.pdf.

Disclaimer

The research undertaken for this report has received financial support within the context of the PROGRESS framework agreement, Lot 4, commissioned by the Netherlands' ministries of Foreign Affairs and Defence 5. Responsibility for the contents and for the opinions expressed rests solely with the authors. Publication does not constitute nor should be construed as an endorsement by the Netherlands' ministries of Foreign Affairs and Defence.