Tsvetelina van Benthem – Written Evidence (AIW0033)

I am grateful for the opportunity to submit evidence to the Artificial Intelligence in Weapon Systems Select Committee.

I am a lecturer in international law for the Oxford Diplomatic Studies Programme and a researcher at the Oxford Institute for Ethics, Law and Armed Conflict (ELAC) and Merton College. My areas of specialisation are the law of armed conflict, international criminal law, international human rights law, and the regulation of emerging technologies under international law. As part of my research on risk-taking and unintended engagements in armed conflict, I have been involved in the work of the Group of Governmental Experts on Lethal Autonomous Weapons Systems since 2018. In 2021, as a member of ELAC, I collaborated with the Centre for Data Ethics and Innovation on an academic workshop tasked with the review and appraisal of the then-draft Ethical Principles for AI in Defence.

At Oxford, I co-convene the Transitional Justice Research Group and am a member of the core team organising the Oxford Process on International Law Protections in Cyberspace.

Summary

The evidence focuses on the application of international law, and in particular international humanitarian law, to autonomous weapons systems (AWS). To address this question, the first two sections of the document review the convergence of state understandings on the characteristics of AWS and the particular concerns that such systems give rise to. The key messages of this submission are, as follows:

1. While there is no internationally agreed definition of AWS, most definitions advanced by states and other stakeholders emphasise a capacity of weapons systems to select and apply force to targets without human intervention. There is thus a convergence of views over the main characteristics of such systems.

2. AWS are tools designed to channel the intent of parties to conflict. They can offer a range of operational advantages. At the same time, they may exacerbate the risks of unintended engagements, including engagements where the actual target of an attack is not one intended by the party to conflict. For instance, although a party to conflict may only want an AWS to attack combatants and lawful military objectives, the system – either because of inadequate programming parameters, faulty sensors or an unforeseen reaction to environmental conditions – could engage civilians or civilian objects. Questions regarding the absence of intent vis-à-vis a particular outcome pose particular difficulties under existing international law.

3. International humanitarian law (IHL) applies to AWS. The more complex question is how precisely the law applies to these systems. If further clarified and specified, the wealth of positive and negative obligations under IHL can offer a robust and sufficient regime to regulate the development and use of AWS.

4. The inquiry of how IHL applies to AWS is complicated for two main reasons: first, there are ongoing disagreements over the content of particular obligations under IHL, and second, even where the content of obligations is not in dispute, the precise way in which they apply to AWS may be unclear due to the novel features of these technologies.

5. States must continue to clarify the rules of IHL. They must engage in this clarification exercise with more detail and granularity. High-level agreement can mask deep disagreements over the substantive scope of rules. For instance, agreement that the prohibition of attacking civilians is applicable and relevant to the use of AWS says little about whether this prohibition covers only the intentional targeting of civilians, or also reckless or negligent engagements.

6. The effectiveness of IHL depends on the clarity of its rules and the reasonable prospect of review and accountability in case of breach. If IHL obligations do not provide clear guidance to their addressees, the law could hardly exert meaningful constraining power. Equally, if states do not consider that there is a reasonable chance for the implementation of state responsibility in case of violations, they may be more likely to take unlawful risks that can ultimately harm civilians, their own troops or friendly forces.

7. Focusing on IHL should not detract from the examination of other relevant and applicable legal regimes, such as international human rights law and international criminal law. The interaction between these regimes should be further explored.

1. What do you understand by the term autonomous weapons system (AWS)? Should the UK adopt an operative definition of AWS?

1.1. While there is no internationally agreed definition of AWS, most definitions adopted by states and other stakeholders emphasise a capacity of systems to select and apply force to targets without human intervention.[1] How one defines this term of course has significant implications in terms of coverage. A broad definition of AWS would encompass systems currently in use, such as the Phalanx system[2] and the

[1] In March 2023, the Chair of the GGE on LAWS published a non-exhaustive compilation of definitions and characterisations submitted by states since 2017 - Non-exhaustive compilation of definitions and characterizations submitted by the Chairperson, CCW/GGE.1/2023/CRP.1, 10 March 2023. What can be observed from 2017 to 2023 is a standardisation of understandings across states. While important differences in terminology (and therefore of scope) remain, states are coalescing around the elements identified in this section. See also ICRC, Position on autonomous weapon systems, 12 May 2021.

HARPY 'fire-and-forget' weapon.[3] A narrow definition (for instance systems capable of setting of modifying their objectives or systems relying on particular forms of machine learning) would limit the weapon systems that fall within the confines of this term. It bears mentioning that there is a distinction to be drawn between AWS and artificial intelligence (AI). Depending on one's definition, a weapon system may qualify as autonomous even without relying on AI in its functions. And AI – itself a term without a universally agreed definition – refers to a set of technologies that enable the computation of a variety of advanced functions,[4] with applications extending well beyond AWS.

1.2. Definitional questions in relation to AWS have been debated at length at the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE on LAWS), a group operating within the framework of the Convention on Certain Conventional Weapons. Despite the lack of agreement on a specific definition of such systems, the GGE on LAWS is able to hold substantive discussions on the regulation of AWS. This is because there seems to be broad agreement on a number of premises and characteristics of AWS, which in turn allows the identification of legal, political and ethical challenges.

1.2.1. Concerning premises, there is broad agreement that (a) autonomy is a **relational concept**: one system, person or entity is autonomous in its relation to another system, person or entity; (b) autonomy **exists on a spectrum**: there are varying types and degrees of autonomy; and (c) autonomy can exist in relation to **a range of functions**: for instance, we may think of autonomy in the assessment of data to identify targets, or autonomy in decisions to apply force, among others.

1.2.2. Concerning characteristics, there is a degree of convergence over the following aspects of these systems: subsequent to activation by a human decision-maker, the weapon system has the capacity to (a) **identify, select and engage targets** (b) **with force** (c) **without further intervention by a human**. A key term here is 'intervention', and there is scope for interpretation whether *any* type of human input would qualify as 'intervention' in the operation of the system.[5] According to most states and commentators, of particular significance here is the relationship

---

[2] The Phalanx weapon system 'automatically detects, evaluates, tracks, engages and performs kill assessments against anti-ship missiles and high-speed aircraft threats' - <https://www.raytheonmissilesanddefense.com/what-we-do/naval-warfare/ship-self-defense-weapons/phalanx-close-in-weapon-system>.

[3] The HARPY loitering munition can search for radar emitters in a pre-defined loitering area – IAI, HARPY Autonomous Weapon for All Weather, at <https://www.iai.co.il/p/harpy>.

[4] The UK MoD understands AI 'as a family of general-purpose technologies, any of which may enable machines to perform tasks normally requiring human or biological intelligence, especially when the machines learn from data how to do those tasks.'

[5] For instance, Palestine has suggested that nominal human input after a system's activation does not amount to a human intervention. See State of Palestine's Proposal for the Normative and Operational Framework on Autonomous Weapons Systems, CCW/GGE.1/2023/WP.2/Rev.1, 3 March 2023.

between the input of the human decision-maker and the degree of independence delegated to the system, or in other words the type, degree and quality of human-machine interaction.[6]

1.3. It would be advisable for the UK to adopt a working definition of AWS. In a paper co-authored by the UK and submitted to the GGE on LAWS in March 2023, AWS are understood as weapon systems that, 'once activated, can identify, select, and engage targets with lethal force without further intervention by an operator.'[7] This understanding is largely in line with the characteristics and definitions advanced by other states, and provides a helpful baseline for advancing discussions.

1.4. What a working definition with such key yet broad characteristics allows is, first, the identification of features of AWS that may give rise to particular technological, legal, ethical or political challenges, and second, the establishment of an umbrella category that is not overly narrow. A definition that confines AWS to particular existing and foreseeable technologies would likely prove both arbitrary and of limited use given rapid technological developments in the area.

1.5. In sum, the adoption of a working definition of AWS that reflects a weapon system's capacity to identify, select and engage targets with force without human intervention subsequent to activation would align with current inter-governmental discussions and reflect the dynamics of distancing between direct human input and concrete battlefield outcomes.

2. What are the possible challenges, risks, benefits and ethical concerns of AWS? How would AWS change the makeup of defence forces and the nature of combat?

2.1. In the discussions on the legality and operational desirability of developing and employing AWS, states and commentators have identified a wide range of possible benefits and concerns. To start with the benefits, it is often asserted that AWS will bring increased precision to the battlefield, that, unlike humans, these systems will have the capacity to engage in assessments that are not influenced by anger, fear, or fatigue, and that autonomy can minimise target misidentification and reduce the risks to a state's own troops. Further, given the speed at which new weapons can travel, autonomous functions may provide the only route to meaningful defence. Despite the range of perceived advantages, serious concerns remain. Lack of predictability in concrete engagements,

[6] In a paper submitted by Argentina, Costa Rica, Guatemala, Kazakhstan, Nigeria, Panama, Philippines, Sierra Leone, State of Palestine and Uruguay, it is stated that 'a working characterization is a useful starting point and that such characterization should focus on the human element and human-machine interaction since these are essential to addressing the issue of attribution of responsibility.' - CCW/GGE.1/2022/WP.3 (2022).

[7] Draft articles on autonomous weapon systems – prohibitions and other regulatory measures on the basis of international humanitarian law ("IHL") submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States, CCW/GGE.1/2023/WP.4/Rev.1, 13 March 2023. As stated in the Draft Articles, this understanding of AWS is without prejudice to any other understandings of this or similar terms for other purposes.

unreliability and algorithms running on biased data are the main concerns that pervade the discussions on AWS. Thus, some states consider that autonomous capabilities lend themselves to 'layers of unpredictability and cascading impacts',[8] may herald new asymmetric methods of warfare that will increase the risk of miscalculation,[9] could imperil international peace and security through the possibility of unintended or uncontrollable levels of escalation[10] and be too opaque to be understood by human operators.[11]

2.2. It is clear that AWS can offer strategic and operational benefits to parties to conflict. It is equally clear that these systems raise a number of serious concerns. The concern that I will focus on in this section, and which is of particular relevance to the application of international law to AWS, is that of **unintended engagements**.

2.3. **AWS, as all weapons, are tools designed to channel the intent of parties to armed conflict**. There is little operational benefit to systems that are unreliable. The concern with AWS is not that these systems will be developed to purposefully target civilians or with the intention to strike military objectives and civilian objects without distinction. **The main risk is of unintended engagements where the intention of the party to conflict is not translated to the outcome produced by the AWS**. Such engagements can occur, for instance, if the AWS mistakenly identifies civilian persons or objects as lawful targets (*ie* combatants or military objectives) for reasons that may include inadequate programming parameters, faulty sensors or unforeseen reaction to environmental factors.

2.4. Factual uncertainty is pervasive in armed conflict and has been amplified by the migration of fighting to urban areas and asymmetric tactics. Contemporary armed conflicts are environments conducive to mistakes. As will be detailed in section 3, it is not infrequent for parties to conflict to engage protected persons and objects under a mistaken belief about their status as lawful military objectives – and this is regardless of the weapons used. In this sense, the development and use of AWS does not introduce a completely new vector of risk to the battlefield but may amplify existing concerns over the protection of civilians and civilian objects from the dangers of military operations.

2.5. Why could AWS amplify existing concerns over the protection of civilians and civilian objects?

> 2.5.1. **First, these systems may extend the distance between direct human input and a specific use of force against a target**. Distance in this context can be temporal (*ie* a wider temporal frame between activation of the system and target

[8] Pakistan, 'Proposal for an international legal instrument on Lethal Autonomous Weapons Systems (LAWS)' submitted to the GGE on LAWS, 8 March 2023, CCW/GGE.1/2023/WP.3/Rev.1, para. 11.
[9] id, para. 14.
[10] ibid.
[11] State of Palestine's Proposal for the Normative and Operational Framework on Autonomous Weapons Systems, CCW/GGE.1/2023/WP.2/Rev.1, 3 March 2023, para. 27.

engagement), geographical (*ie* human decision-makers may be completely removed from the geographical location of the battlefield), decisional (*ie* humans set the parameters of autonomous operation yet the actual target engagement does not require additional human input). Distance can entail risks. For instance, there may be more scope for malicious external interference with the operation of the system (through hacking, among others), impact of environmental factors (including weather conditions), change in battlefield conditions (such as civilians unexpectedly entering an area).

2.5.2. **Second, some have expressed concerns that AWS may be unreliable and/ or uncontrollable in ways that are difficult to predict**. In contrast, it is predictable that a bullet fired by a rifle may ricochet from a surface. Even though a bullet that ricochets is unlikely to channel the intent of the person who pulled the trigger, that person can reasonably predict the ways in which firing the rifle may cause unintended harm.

2.6. The concerns mentioned above mostly relate to the current and foreseeable state of technological development. It may be that AWS can be designed to parse the complexity and evolving conditions of battlefields, to operate within limited boundaries, to reliably channel human intent, and to lead to explainable and traceable outcomes. It may also be that AWS will lead to less mistakes in the identification of targets, and thus to a decrease in civilian harm. Even so, there always remains the risk of engagements that were not intended by the party to conflict deploying the AWS. This, in turn, leads to discrete legal questions. How does international law regulate unintended engagements? To what extent, if at all, can a party to conflict be held responsible for the mistaken targeting of a civilian or civilian object? What obligations does international law impose on parties to conflict in relation to the reduction and management of the risk of such engagements? **The clearer the obligations under international law and the more robust the legal protection in relation to unintended engagements, the lesser the risk of states deploying AWS that are not reliably tested, verified, and understood**.

3. Is existing International Humanitarian Law (IHL) sufficient to ensure any AWS act safely and appropriately? What oversight or accountability measures are necessary to ensure compliance with IHL? If IHL is insufficient, what other mechanisms should be introduced to regulate AWS?

3.1. IHL can provide a robust and sufficient legal framework for the safe and responsible development and use of AWS, subject to further clarification of the relevant obligations. Even though there are no obligations specifically regulating AWS under existing IHL, this legal

regime contains a wide range of obligations addressed to parties to conflict, both positive and negative in character, that bear on the development and use of AWS. Importantly, it is a basic premise of IHL that the right of parties to conflict to choose their means and methods of warfare is not unlimited. In addition to engaging in the clarification of existing obligations under IHL, states may choose to strengthen, detail and further develop this framework by concluding additional international agreements, for example in the form of a new protocol to the Convention on Certain Conventional Weapons.[12] Confidence-building[13] and capacity-building measures[14] can also complement the existing IHL framework.

3.2. **This section will focus on the need for clarification of existing obligations under IHL. Without such clarification, the rules binding parties to conflict under IHL may leave too much ambiguity and provide too little concrete guidance to decision-makers**. **Ambiguity in the law is itself a danger to the protection of civilians**. The addressees of obligations under IHL are parties to armed conflict (states and non-state groups), not weapons systems. In this sense, the key questions that arise are on the way in which these addressees can comply with *their* obligations *in the use of AWS*.

3.3. **While states agree on the applicability of IHL to AWS, the extent to which they agree on the elements of particular rules and on the way in which these rules apply to AWS is unclear.** IHL is a legal regime developed to accommodate the changing character of armed conflict,[15] and states have affirmed within the auspices of the GGE on LAWS that 'international humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems'.[16] To say that IHL *applies* is a first necessary step in the analysis. What is more difficult to determine is *how* precisely IHL applies to these emerging technologies.[17]

---

[12] Many states have called for the conclusion of a new AWS-specific treaty instrument. See, for instance, Communiqué of the Latin American and the Caribbean Conference of Social and Humanitarian Impact of Autonomous Weapons, February 23 – 24, 2023; Pakistan, 'Proposal for an international legal instrument on Lethal Autonomous Weapons Systems (LAWS)' submitted to the GGE on LAWS, 8 March 2023, CCW/GGE.1/2023/WP.3/Rev.1; State of Palestine's Proposal for the Normative and Operational Framework on Autonomous Weapons Systems, CCW/GGE.1/2023/WP.2/Rev.1, 3 March 2023.

[13] Confidence-building measures are measures designed to increase trust and understanding between states.

[14] Capacity-building measures are measures designed to develop and strengthen the skills, abilities, processes and resources available to states.

[15] International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion of 8 July 1996, para. 86.

[16] Guiding Principle (a), Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 25 September 2019, CCW/GGE.1/2019/3, Annex IV.

[17] A similar conversation is taking place in relation to the regulation of information and communications technologies (ICTs) under international law. See, for instance, the 2021 Report of the Open-Ended Working Group on ICTs; the 2021 Report of the Group of Governmental Experts on Advancing responsible State behaviour in cyberspace in the context of international security; The Oxford Process on International Law Protections in Cyberspace: A Compendium (October 2022), available at: <https://www.elac.ox.ac.uk/wp-content/uploads/2022/10/Oxford-Process-Compendium-Digital.pdf>, p. 13.

3.4. **Agreement on the applicability and relevance of certain rules and standards does not necessarily imply agreement on their substantive coverage**.

3.4.1. Thus, **states may**, for instance, **agree that parties to conflict must comply with obligations under IHL in their development and use of AWS, but disagree on the scope of these obligations**. States may agree on the importance of the principle of distinction and the prohibition of attacking civilians as one of the specific obligations that flow from it. However, they may disagree on whether this prohibition covers attacks against civilians conducted with intent in relation to their civilian status, or also attacks that are reckless or negligent in relation to that protected status. In this way, agreement on the principles and obligations without clarification of their precise elements may in fact mask deep disagreements over the substantive scope of existing rules.

3.4.2. **States may also use the same terms, but with different meaning**. For instance, both Palestine and the Russian Federation emphasise the importance of ensuring 'human control' in order to comply with IHL obligations. However, while Palestine understands human control to mean control over the operation of the weapon system across its lifecycle and up to the concrete application of force,[18] the Russian Federation argues that 'effective human control over system can be achieved beyond direct control' and 'specific forms and methods of such control should be left to the discretion of the state.'[19] To some, human decision-making at the stage of developing the AWS and setting its parameters would suffice. To others, meaningful scope for human decision-making must be ensured in the lead to and during the initiation of force towards a particular target.

3.4.3. And finally, **standards of 'adequacy', 'reasonableness' and 'feasibility' are often used to describe the care owed in using AWS. These standards, however, mean little without concrete guidance on the factors that determine what is *reasonable*, which precautions are *feasible* in particular battlefield contexts, what it means to *adequately* predict, understand and explain the effects of weapons**.[20]

3.5. **The question of *how* IHL applies to AWS is complicated for two main reasons**. **First**, decades-long debates on the scope of existing

---

[18] State of Palestine's Proposal for the Normative and Operational Framework on Autonomous Weapons Systems, CCW/GGE.1/2023/WP.2/Rev.1, 3 March 2023, para. 26.

[19] Russian Federation, 'Concept of Activities of the Armed Forces of the Russian Federation in the Development and Use of Weapons Systems with Artificial Intelligence Technologies', Unofficial English version, 7 March 2023, CCW/GGE.1/2023/WP.5.

[20] Pakistan proposal (note 8), para. 27(a) – 'Prohibitions on development, deployment and use in all circumstances of an autonomous weapon system that: • has effects which cannot be adequately predicted, understood and explained.'

obligations under IHL persist today. The debates around AWS thus expose **a lack of clarity in the content of legal obligations that exists generally, not only in relation to the use of autonomy**. **Second, even if states agree on the existence of obligations under IHL and their elements, they may disagree on the way in which these elements apply in the particular context of AWS.** I will address these two aspects in turn.

3.5.1. Lack of clarity over the elements of IHL obligations

Under both treaty and customary international law, parties to conflict must distinguish between civilians and civilian objects, on the one hand, and combatants and military objectives, on the other, refrain from launching disproportionate and other indiscriminate attacks, take precautions in attack and against the effects of attacks, and undertake legal reviews of new weapons. Even if the existence of these obligations is not in doubt, their elements leave scope for interpretation. Three IHL rules of special relevance to the use of AWS can be used to demonstrate this.

- ***The prohibition of attacking civilians***

  According to art. 51(2) of Additional Protocol I and customary international law, 'the civilian population as such, as well as individual civilians, shall not be the object of attack.' Attacks against civilian objects are similarly prohibited under art. 52(1) of Additional Protocol I and custom.[21]

  What does it mean to make the civilian population or individual civilians *the object of attack*? The text itself does not specify whether it is prohibited to direct attacks against civilians purposefully, knowingly, recklessly, negligently, or in all cases regardless of fault (strict liability). Some states and commentators see this prohibition as covering intentional conduct only, others consider that it extends to recklessness or negligence. For the treaty text, the answer to this question requires an analysis through the rules on treaty interpretation.[22] For the customary rule, the content of the prohibition can be discerned by examining evidence of state practice accepted as law.[23] I have written elsewhere that this prohibition covers not only the intentional targeting of civilians, but also, at the very least, direction of attacks against civilians with recklessness regarding their civilian status.[24]

---

[21] For reasons of scope, this section will focus on the prohibition of attacking civilians.

[22] Vienna Convention on the Law of Treaties, arts. 31 – 33.

[23] Statute of the International Court of Justice, art. 38(1)(b). See also International Law Commission, Draft conclusions on identification of customary international law (2018).

[24] van Benthem, 'Exploring Changing Battlefields: Autonomous Weapons, Unintended Engagements and the Law of Armed Conflict' (2022), available at: <https://www.ccdcoe.org/uploads/2022/06/CyCon_2022_book.pdf>, p. 189.

What the subjective element embedded in this prohibition is will, of course, determine the scope of conduct that would fall foul of the rule. Given that mistakes and malfunctions are regular occurrences in armed conflict, this discussion has a particular significance for civilian protection. From the bombing of the Chinese embassy in Belgrade[25] through the attack on the Médecins sans Frontières hospital in Kunduz[26] to the 2021 Kabul strike that mistakenly killed ten civilians, including seven children,[27] it is clear that faulty intelligence, institutional biases and deficient decision-making procedures can lead to target misidentifications causing enormous harm to civilians.

Consider the March 2022 attack by Russian Armed Forces against the Donetsk Regional Drama Theatre in Mariupol, Ukraine. The theatre was used by civilians as an air raid shelter, and satellite imagery shows that the word ДЕТИ ('children' in Russian) was spelled out in large letters outside the building as an attempt to indicate to attackers that the theatre is used by civilians. Even if Russian forces did not target the theatre with the intention of attacking civilians, they were at the very least reckless in launching this attack. If the obligation to refrain from making civilians and civilian objects the object of attack covers intentional conduct only, a mistake in relation to the status of the chosen target would place such a strike outside the rule's regulatory reach. If, on the other hand, the standard is one of recklessness or negligence, the key questions will be whether there was awareness of a risk that the target is in fact civilian or there should have been such awareness.

Because of the risk of unintended engagements outlined in section 2, the clarification of the subjective element in the prohibition of attacking civilians has particular relevance in the context of AWS. There is no doubt that the intentional targeting of civilians with AWS is prohibited. Yet this is not the main concern. The main concern is whether the unintended application of force to civilians and civilian objects can be considered a breach of the relevant prohibitions under IHL, and thus entail the responsibility of the state or non-state group party to conflict deploying the AWS.[28]

---

[25] Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia, paras. 80 – 85.

[26] Médecins sans Frontières, Kunduz Hospital Attack in Depth, available at: <https://www.msf.org/kunduz-hospital-attack-depth>.

[27] US Department of Defense, 'DoD: August 29 Strike in Kabul 'Tragic Mistake,' Kills 10 Civilians', 17 September 2021, available at: <https://www.defense.gov/News/News-Stories/Article/Article/2780257/dod-august-29-strike-in-kabul-tragic-mistake-kills-10-civilians/>.

[28] Mistaken strikes can also be assessed against the obligations to take precautions in attack, and in particular the

- ***The presumption of civilian status***

  'In case of doubt whether a person is a civilian, that person shall be considered to be a civilian.'[29] The consequence of this presumption is that, if there is doubt regarding status, a party must refrain from launching an attack. How to understand the standard of 'doubt' is, however, contested. It is unclear whether a party to conflict must be 'absolutely certain', 'near certain', 'reasonably certain', or 'act reasonably' in relation to the status of persons or objects. This question of legal standard of doubt is separate from the technical feasibility of developing AWS that can observe the presumption such that the party to conflict deploying them is in compliance with its obligations under IHL. One must first know what standard ought to be 'encoded' in the system.

- ***The prohibition of indiscriminate attacks***

  Under IHL, it is prohibited to launch indiscriminate attacks. One type of indiscriminate attack is an attack which employs 'a method or means of combat which cannot be directed at a specific military objective.'[30] The phrase 'means of combat' generally refers to the weapons or weapons systems being used by a party.[31] Of particular interest under this rule are weapons that are inaccurate or unreliable. It is unclear, however, how inaccurate or unreliable a weapon must be, in a particular context of deployment, for an attack to be considered indiscriminate under this rule. While there seems to be a tolerance level of errors in the use of weapons, where this tolerance level is set is not clearly specified under the relevant treaties, in international jurisprudence or in the practice of states. To be able to meaningfully operationalise protection under this rule, the question of acceptable margins of error, and the relevant subjective attitude vis-à-vis such margins of errors, must be further explored.[32]

3.5.2. Difficulties over the application of existing IHL obligations to the use of AWS

In addition to substantive disagreements over the elements and standards of relevant rules of IHL, the specifics of AWS may pose

---

obligation to verify targets.

[29] Additional Protocol I, art. 50(1). In relation to objects, the presumption states: 'In case of doubt whether an object which is normally dedicated to civilian purposes, such as a place of worship, a house or other dwelling or a school, is being used to make an effective contribution to military action, it shall be presumed not to be so used.' – art. 52(3).

[30] Additional Protocol I, art. 51(4)(b).

[31] ICRC, Commentary to Additional Protocol I, 1987, para. 1957.

[32] See Laura Bruun, Marta Bo and Netta Goussac, 'Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What does IHL Permit, Prohibit and Require?', SIPRI Report 2023. The report details both the prohibition on the use of weapons that are indiscriminate by nature and the prohibition on the indiscriminate use of weapons that are not otherwise prohibited.

particular difficulties to the law's application even where the elements and standards of obligations are clear. For instance, in the conduct of attacks, parties to conflict are required to take feasible precautions.[33] What the standard of 'feasibility' would mean in relation to the verification of targets or choice of means and methods in the use of AWS needs further unpacking.[34] While the analysis is context-dependent, there is a need to at least identify factors relevant to the operation of particular AWS systems that can guide decision-makers in their choice of adequate and effective precautionary measures.

3.6. IHL is a legal field containing wide-ranging protections capable of adapting to the use of emerging technologies in military decision-making. This field contains a wealth of negative obligations, that is, obligations to refrain from particular conduct (such as the prohibition of directing attacks against civilians and of indiscriminate attacks), and positive obligations, that is, obligations to take certain measures (such as the obligations to take precautions in attack or against the effects of attacks, and the obligation to conduct reviews of new weapons). This system of obligations can offer a viable route for ensuring the safe and responsible development and use of AWS. At the same time, remaining disagreements over the elements of relevant IHL rules and their application to the use of AWS may perpetuate an area of legal uncertainty that would hamper meaningful constraints over AWS development and use. Thus, **clarification of IHL should be a key priority**. States must set out their positions on the content of specific rules of IHL with more granularity and be upfront and transparent about the most contentious elements of key IHL rules.

3.7. That further detail on the elements of specific IHL rules is necessary to ensure the safe and responsible development and use of AWS is now well-understood. In the GGE on LAWS discussions, states and other stakeholders have repeatedly called for further specification of IHL protections. With each meeting, the discussions are increasing in legal sophistication, and this can be discerned both in session statements and working papers submitted to the GGE. That said, there is a need for more focused discussions on legal regulation, and in particular on the following:

3.7.1. **Identifying elements of specific IHL obligations**. In particular, subjective elements and elements related to acceptable margins of error must be discussed. They must be discussed with particular attention to the implications of their coverage for unintended harms.

3.7.2. **Strengthening procedural obligations of parties to conflict**. Procedural obligations, such as obligations to conduct legal

[33] Additional Protocol I, art. 57.

[34] van Benthem, 'The Regulation of Militarised Artificial Intelligence: Protecting Civilians through Legal Reviews of New Weapons and Precautions' in Bielicki (ed.), *Regulating Artificial Intelligence in Industry* (Routledge 2022).

reviews of new weapons, can serve as information-generators on the characteristics of weapons systems, modalities of human-machine interaction, and the relationship between such characteristics and interaction, on the one hand, and obligations under IHL, on the other.[35]

3.7.3. **Clarifying the relationship between IHL and other applicable rules of international law**. IHL is only one area that can be brought to bear on the development and use of AWS. A topic that remains underexplored in the context of AWS is the relationship between obligations under IHL and those under international human rights law (IHRL). Obligations related to the procedures leading to a use of force and investigations arise under the right to life and may be of particular relevance to AWS.[36] The potential unpredictability of AWS may imperil not only civilians, but also a state's own forces or allies who are present in the theatre of deployment, and thus cause foreseeable dangers to their life and limb. International criminal law (ICL), which regulates the criminal responsibility of individuals for genocide, crimes against humanity, war crimes and the crime of aggression, is another legal regime that needs further exploration in the context of AWS. While the law of war crimes builds on IHL, there are important differences in the elements of rules arising under these two regimes.[37]

3.8. States are bound by the IHL treaties they are parties to, and by customary IHL. Under the law of state responsibility, a breach of an international obligation that is attributable to a state constitutes an internationally wrongful act that entails that state's responsibility, and thus a duty to provide reparation.[38] The prospect of state responsibility may have a deterrent effect over the rapid development and deployment of AWS, and ensure a gradual, phased and responsible approach that mitigates possible risks to civilians. That said, access to international fora for the invocation and review of state responsibility is not always open. For instance, the jurisdiction of the International Court of Justice in contentious cases is subject to the consent of states.[39] Proceedings against a state may be initiated before human rights courts, subject to that state being a party to the relevant treaty instrument.[40] The human

---

[35] van Benthem, 'Exploring Changing Battlefields: Autonomous Weapons, Unintended Engagements and the Law of Armed Conflict' (2022), available at: <https://www.ccdcoe.org/uploads/2022/06/CyCon_2022_book.pdf>, p. 189.

[36] Particular interpretative controversies have arisen over the extraterritorial application of human rights treaties. Because of differing standards, the analysis of such application must be conducted for each specific human rights treaty. On a recent examination of the extraterritorial application of the European Convention on Human Rights, see European Court of Human Rights, *Ukraine and The Netherlands v Russia*, Applications nos. 8019/16, 43800/14 and 28525/20, Grand Chamber Decision, paras. 555 et seq.

[37] The default mental element under the Rome Statute of the International Criminal Court is intent and knowledge (art. 30), which may be higher than the elements present in obligations under IHL binding *states* and non-state groups parties to conflict.

[38] International Law Commission, Articles on the Responsibility of States for Internationally Wrongful Acts (2001), arts. 1, 2, 31.

[39] Statute of the International Court of Justice, art. 36.

rights court would be able to review the state's conduct against obligations arising under the human rights enshrined in its constitutive treaty, and these rights may, to the extent interpretation allows, accommodate the relevant standards under IHL.[41] In a non-judicial setting, the conduct of states may be subject to scrutiny within the institutional frameworks of international organisations – for instance through reporting obligations and regular review.[42] If states decide to move forward with the adoption of additional legal regulation of AWS, it would be advisable to consider, in addition to further substantive regulation, robust review mechanisms and platforms for continuous inter-governmental dialogue.

3.9. Measures taken at the international level are important, and so are those taken by states domestically. Compliance with IHL is crucial for a state's reputation not only as it is projected externally to other states and stakeholders, but also internally to its own citizens. This was acknowledged in the UK AI in Defence policy statement: a failure to recognise the risks and concerns about the impact of autonomy on humans may 'risk losing public consent'. Strong and robust domestic strategies on the development and use of AWS, mainstreaming of international law in its application to AWS in legislation, military instructions and training, and rigorous and continuous testing of AWS are all key to ensuring a responsible and safe approach to the introduction of autonomy in weapon systems.

3.10. To conclude, **IHL contains a range of positive and negative obligations that constrain the development and use of AWS. How effective IHL will be in ensuring meaningful regulation of AWS will depend, first, on the further clarification and specification of these obligations, and second, on the prospect of reviewing state behaviour and operationalising responsibility for violations of the law.**

4. What are your views on the Government's AI Defence Strategy and the policy statement 'Ambitious, safe, responsible: our approach to the delivery of AI-enabled capability in Defence'? Are these sufficient in guiding the development and application of AWS? How does UK policy compare to that of other countries?

4.1. This section will briefly address the 2022 policy statement. Focusing on the safe and responsible introduction of AI in the defence sector, as the UK does, is the right approach to take. The strong emphasis on legal regulation, the acknowledgment of concerns over algorithmic bias and

---

[40] For instance, applications related to alleged violations of the European Convention on Human Rights can only be brought to the European Court against states that are parties to the European Convention.

[41] See, for instance, Human Rights Committee, General comment No. 36 (2018) on article 6 of the International Covenant on Civil and Political Rights, on the right to life, CCPR/C/GC/36, 30 October 2018, para. 64.

[42] Consider the process of the Universal Periodic Review established within the confines of the Human Rights Council - <https://www.ohchr.org/en/hr-bodies/upr/basic-facts>.

unpredictability and the affirmation that risk ownership over the use of AI must be clearly defined provide a sound basis for achieving the goals of safety and responsibility. Overall, the five principles cover some of the main concerns over the use of AWS – loss or distancing of the human element, concerns over understanding, potential biases and unreliability, and risk of erosion of chains of responsibility. It is particularly welcome that the policy statement looks at autonomy from two responsibility lenses: what it means *to be responsible* in the introduction of autonomy in defence and how to ensure that decision-makers are *held responsible* for harmful outcomes.

4.2. In addition to the UK AI Defence Strategy and policy statement, the UK has been and continues to be active in the discussions at the GGE on LAWS.[43] Through statements and working papers, the UK has advanced the inter-governmental discussions, including on the regulation of AWS under IHL. In light of the analysis in section 3, further work is needed on clarifying the content of IHL and other relevant legal regimes and specifying their application to AWS. The UK has already suggested that a viable route forward may be the elaboration of a manual on the application of international law to AWS. An initiative of this kind can indeed bring more granularity to the discussions. Its success will rest on the buy-in from states and other stakeholders, the transparency of the process, and the rigour of the legal analysis. A continued emphasis on the clarification of international law is recommended.

5. Are existing legal provisions and regulations which seek to regulate AI and weapons systems sufficient to govern the use of AWS? If not, what reforms are needed nationally and internationally; and what are the barriers to making those reforms?

5.1. These questions are partly addressed in section 3. In this section, I will briefly review the most prominent approaches to AWS regulation advanced by states, together with some of the barriers facing these approaches.

5.2. To begin with, states are converging over a two-tiered approach to AWS: certain AWS that cannot comply with international law are prohibited, and the use of AWS that are not prohibited is regulated.[44] This

---

[43] See, for instance, UK Written Contributions on Possible Consensus Recommendations in Relation to the Clarification, Consideration and Development of Aspects of the Normative and Operational Framework on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (2021) and UK Proposal for a GGE Document on the Application of International Humanitarian Law to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (2022).

[44] Joint Statement by 70 States on Lethal Autonomous Weapons Systems, First Committee, 77th United Nations General Assembly Thematic Debate – Conventional Weapons, 21 October 2022; France, Intervention to GGE on LAWS, 6 March 2023, available at : <https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/20230306_DSMT_GENEVE_GGE_SALA_Intervention_g%C3%A9n%C3%A9rale.pdf>.

two-tiered structure is already part and parcel of existing IHL. Weapons of a nature to cause superfluous injury or unnecessary suffering,[45] as well as inherently indiscriminate weapons, are prohibited *as such*. And even if a weapon system is not prohibited in and of itself, its use is regulated under international law. Although the two-tiered approach does not add to existing IHL, it does provide a structured way of thinking about the nature and use of AWS. Under this approach, we circle back to the difficult questions around the clarification and specification of obligations under IHL – what is the threshold for considering a weapon inherently indiscriminate? What does it mean to direct attacks against civilians and civilian objects? Which factors determine the feasibility of precautions in a given context?

5.3. In recent years, a number of states have called for the adoption of a new legally binding instrument on AWS. In February 2023, 33 states from Latin America and the Caribbean adopted the Belén Communiqué, through which they commit to 'collaborate to promote the urgent negotiation of an international legally binding instrument, with prohibitions and regulations with regard to autonomy in weapons systems, in order to ensure compliance with International Law, including International Humanitarian Law, and ethical perspectives, as well as the prevention of the social and humanitarian impact that autonomy in weapons systems entail'.[46] While a new treaty may establish substantive and procedural obligations that are tailored to AWS and more granular than existing rules under IHL, it is unlikely that a new treaty would attract universal endorsement. In particular, states that have already invested substantially in the research and development of autonomous military capabilities, including Israel, Russia, the UK and the US seem to prefer the clarification of existing IHL over the negotiation of a new treaty.

5.4. Clarifying existing international law and reflecting on the benefits of a new treaty instrument can meaningfully happen in parallel. While clarifying the substantive coverage of IHL seems to be a prerequisite for discussions on what is needed *in addition* to existing rules, the lack of AWS-specific rules in the law and the relatively slow pace of the GGE discussions compared to the rate of technological development explain the interest in adopting detailed, concrete and tailored new rules. That said, states and other stakeholders must exercise caution when promoting the need for new rules. Using a language of insufficiency to describe existing IHL may undermine the protections under this regime and suggest an overly limited scope of existing obligations.[47] This can be

[45] Additional Protocol I, art. 35(2).

[46] Communiqué of the Latin American and the Caribbean Conference of Social and Humanitarian Impact of Autonomous Weapons, February 23 – 24, 2023, op. para. 1.

[47] A good example of a state advancing further regulation without suggesting the insufficiency of existing IHL comes from the Netherlands. In their 2022 national position on AWS, it is affirmed that 'there are various options for achieving further regulation for both fully autonomous and partially autonomous weapon systems, such as a new protocol to the CCW. This is not so much about developing new legal rules, but primarily about further specifying existing legal rules.' - Letter of 17 June 2022 to the House of Representatives from the

particularly counter-productive given that many states may decide to not become parties to the new AWS-specific treaty. They will remain bound by the obligations under treaties in force to which they are parties and customary international law.

5.5. And finally, institutional culture may pose a barrier to the responsible use of AWS. The promise of autonomy for operational advantage may incentivise the rapid introduction of autonomous functions, and a mentality of 'take risks – learn in the process'. The price of learning, however, may be very high. Taking risks could well be measured in the death and injury of civilians to whom those risks has been transferred. States that take their international obligations seriously will also take safety, risk mitigation and ownership seriously.

Thank you for the opportunity to submit this written evidence.

Tsvetelina van Benthem
May 2023

Minister of Foreign Affairs, Wopke Hoekstra, and the Minister of Defence, Kajsa Ollongren, presenting the government's response to AIV advisory report no. 119, CAVV Advisory Report 38, 'Autonomous Weapon Systems: The Importance of Regulation and Investment.