Dr Elliot Winter – Written Evidence (AIW0001)

Introduction

I am a lecturer at Newcastle University Law School. I research and teach international armed conflict and global security law. My principal specialism is the regulation of new military technologies including autonomous weapons systems (AWS). I have published a range of articles in peer-reviewed journals on this topic. I have no external funding or interests to declare; I wish to participate in this call for evidence simply to contribute to the ongoing debate and to share the key findings of my research. I gave verbal evidence to the Select Committee on 23 February 2023 and was invited by the organisers to contribute further written evidence.

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

Broad definitions for AWS have already been in use for over a decade. The definitions used by actors at different ends of the spectrum of the debate are surprisingly similar. In 2012, the US Department of Defense defined an AWS as a 'weapon system that, once activated, can select and engage targets without further intervention by a human operator'.[1] In the same year, Human Rights Watch defined autonomous weapons as 'robots that are capable of selecting targets and delivering force without any human input or interaction'.[2] Likewise, the International Committee of the Red Cross has defined an AWS as any weapon system with autonomy in its critical functions that can select and attack targets without human intervention.[3] These definitions accurately capture the essence of an AWS. The UK could usefully synthesise these definitions into a formulation suitable for its own doctrine simply to clarify its position. However, any substantive change seems unnecessary and would likely be unwise as it would put the UK out of step with established discourse.

There has been a reluctance to specify more precisely what form an AWS may take in terms of its hardware or software. The Group of Governmental Experts (GGE) was formed under the auspices of the Convention on Conventional Weapons to investigate AWS. It has asserted

[1] United States Department of Defense, 'Autonomy in Weapons Systems' (2012) Directive 3000.09, Glossary Part II <https://bit.ly/2UCP4fc>.
[2] Human Rights Watch, *Losing Humanity: The Case Against Killer Robots* (International Human Rights Clinic 2012) 2.
[3] International Committee of the Red Cross, *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons* (International Committee of the Red Cross 2016) 8.

that a definition based on technological attributes alone would be of limited utility, as technology develops so quickly that any definition agreed upon would soon be rendered redundant.[4] Instead, as the chairperson observed, the GGE favours focussing on the extent of the link between machines and human beings: the 'human-machine interface'.[5] A flexible and future-proofed approach such as this is indeed preferable. The UK already takes a 'technology agnostic' approach that focusses on the level of human control over a weapon rather than its technical characteristics.[6] Therefore, there is no need to change this aspect of the UK stance.

Question 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?

Intentionally blank.

Question 3: What safeguards (technological, legal, procedural or otherwise) would be needed to ensure safe, reliable and accountable AWS?

Intentionally blank.

Question 4: 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?

Existing IHL is sufficient to regulate AWS. The regime is comprised of several core concepts: (i) humanity, (ii) military necessity, (iii) distinction, (iv) proportionality and (v) precaution. Humanity and military necessity represent the foundational tension that underpins IHL and these considerations must be kept in balance at all times. As Dinstein put it, IHL is 'predicated on a subtle equilibrium between the two diametrically opposed stimulants of military necessity and humanitarian considerations'.[7] In this sense, humanity and military necessity are not

[4] United Nations Office for Disarmament Affairs, 'Background on LAWS in the CCW' (*United Nations*, 2023) <https://bit.ly/42d8S9F>.

[5] Group of Governmental Experts, 'Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (23 October 2018) UN Doc CCW/GGE.1/2018/3, Annex III (Chair's Summary) paras 2 and 5.

[6] United Kingdom, 'Statement to the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' Plenary Meeting of the Group of Governmental Experts (25–29 March 2019) para 3.

hard rules or principles in themselves and so do not permit or prohibit anything in isolation. Instead, they are simply foundational pillars that inform the development and interpretation of other rules and principles.[8] However, distinction, proportionality and precaution are hard rules and principles of IHL that directly regulate AWS.

Distinction requires that participants in a conflict 'distinguish between the civilian population and combatants … and … direct … operations only against military objectives'.[9] Proportionality prohibits attacks that would cause collateral damage to civilians or their property that is 'excessive in relation to the concrete and direct military advantage anticipated.'[10] Precaution is a reiteration of distinction and proportionality, augmented by more specific guidance on the practical steps to be taken in the planning and execution of an attack, combined with requirements to use the least destructive means and methods of warfare that are viable in the circumstances and, where possible, to issue warnings.[11] I have written on each of these issues in the past and explained in detail why AWS cannot comply with distinction,[12] proportionality[13] or precaution[14]. In essence, the problem is that each of these concepts are complex, context-sensitive, requirements that require a high level of judgement-making capacity for their implementation.[15] While machines are adept at limited tasks such as classifying a weapon from visual imagery[16] or winning board games such as chess or even the Chinese game 'go',[17] they do not possess the higher-level understanding and reasoning required to, for example, identify an injured or surrendering combatant or to make inherently impressionistic, non-formulaic, decisions about what level of collateral damage is tolerable for a given attack. For example, as recently as January 2023, marines in a training exercise were able to fool artificial intelligence designed for use in a sentry system simply by moving towards it in plain sight under a cardboard box.[18] The system had not

---

[7] Yoram Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (3rd edn, CUP 2016) 9.
[8] Elliot Winter, 'Pillars not Principles: The Status of Humanity and Military Necessity in the Law of Armed Conflict' (2020) 25 Journal of Conflict and Security Law 1.
[9] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (adopted 08 June 1977, entered into force 07 December 1978) 1125 UNTS 3, Article 48.
[10] ibid, Articles 51(5)(b) and 57(2)(a)(iii).
[11] Ibid, Article 57(2).
[12] Elliot Winter, 'The Compatibility of Autonomous Weapons with the Principle of Distinction in the Law of Armed Conflict' (2020) 69(4) International and Comparative Law Quarterly 845.
[13] Elliot Winter, 'Autonomous Weapons in Humanitarian Law: Understanding the Technology, Its Compliance with the Principle of Proportionality and the Role of Utilitarianism' (2018) 6(1) Groningen Journal of International Law 183.
[14] Elliot Winter, 'The Compatibility of the Use of Autonomous Weapons with the Principle of Precaution in the Law of Armed Conflict' (2020) 58(2) Military Law and the Law of War Review 240.
[15] Elliot Winter, 'The Compatibility of Autonomous Weapons with the Principles of International Humanitarian Law' (2022) 27(1) Journal of Conflict and Security Law 1.
[16] Martin Cronin, 'PatScan Platform Detects Hidden Weapons, Chemicals, and Bombs' (*TechRepublic*, 2020) <https://tek.io/2IdOFZB>.
[17] Elizabeth Gibney, 'Google AI Algorithm Masters Ancient Game of Go' (2016) 529 Nature 445.

been programmed to identify a moving box as a threat and did not possess any real understanding or contextual awareness to allow it to apprehend the potential danger. Therefore, the 'Defence Artificial Intelligence Strategy' (DAIS) is correct in its summary that 'machines are good at doing things right (e.g. quickly processing large data sets) [whereas] people are good at doing the right things (e.g. evaluating complex, incomplete ... information guided by values such as fairness'.[19]

Levels of artificial intelligence beyond what is currently available would be required for satisfactory judgement-making capacity; perhaps even 'artificial general intelligence' which is as intelligent as humans.[20] According to software experts, such technology is unlikely to be available for at least 20[21] to 40[22] years, if at all. For now, only humans can make those decisions. Hence the real question is how properly to integrate humans into the operation of autonomous weapons such that they can comply with IHL. The solution adopted will comply with IHL provided that, as noted above, it strikes a balance between the considerations of humanity and military necessity.

It is hard to see how other mechanisms beyond existing IHL could realistically be called upon to regulate AWS. There have long been calls for a new international treaty on AWS and many States, non-governmental organisations and private individuals have shown interest in this cause (with the 'Campaign to Stop Killer Robots' advocacy group being a vocal supporter).[23] However, the key global military powers have continued to invest substantially in this technology and made it clear that they see AWS as a critical capability.[24] This, combined with the current condition of international relations after the invasion of Ukraine, would seem to suggest that the likelihood of a new treaty is slim. The UK has proposed the creation of a 'manual' to help regulate AWS.[25] However, manuals in this context are essentially monographs (albeit highly respected ones) created by groups of experts assembled by small groups of like-minded States. Perhaps the best and most closely related example is the Tallinn Manual on cyber warfare.[26] These documents are non-binding. Moreover, traditionally, manuals seek only to capture and codify

---

[18] ExtremeTech, 'US Marines Defeat DARPA Robot by Hiding Under a Cardboard Box' (*ExtremeTech* 2023) <https://bit.ly/3YNeFzV>.
[19] Ministry of Defence, *Defence Artificial Intelligence Strategy* (Ministry of Defence 2022) para. 1.3.2.
[20] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014).
[21] Vincent Muller and Nick Bostrom, 'Future Progress in Artificial Intelligence: A Survey of Expert Opinion' in Vincent Muller (ed), *Fundamental Issues of Artificial Intelligence* (Springer 2016).
[22] Tony Walsh, *2062: The World that AI Made* (La Trobe University Press 2018).
[23] Campaign to Stop Killer Robots, 'Less Autonomy: More Humanity' (*Stop Killer Robots* 2023) <https://bit.ly/429DA3E>.
[24] Winter (n 15) 5-6.
[25] United Kingdom, 'Proposal for a GGE Document on the Application of International Humanitarian Law to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS)' (*Reaching Critical Will*, 2022) <https://bit.ly/3ThUngJ>.
[26] Michael Schmitt (ed), *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge University Press 2013).

the settled *lex lata* (existing law) rather than to develop *lex ferenda* (future law) as they are written principally for the reference of practitioners who require an accurate account of existing law rather than musings over what the law might become. Thus, they are not suited to effecting revolutionary change.

Question 5: 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?

The UK Government recognises that the deployment of (fully) autonomous weapons would be incompatible with IHL. Specifically, it states in the 'Ambitious, Safe, Responsible' policy document (ASR) that the use of (fully) autonomous weapons is 'not acceptable [and that] the United Kingdom does not possess fully autonomous weapon systems and has no intention of developing them'.[27] It also recognises that the only way to overcome the technological limitations at present is to integrate human decision making into the operation of AWS. The UK Government states in ASR that the solution to the problem is to ensure that 'context-appropriate human involvement' is maintained during the deployment of AWS.[28] Similarly, in DAIS, the UK Government refers to 'the development of effective Human-Machine Teaming, combining human cognition, inventiveness and responsibility with machine-speed analytical capabilities'.[29] These terms reflect the language in the broader literature that AWS must be subject to 'meaningful human control'.[30] In essence then, the critical debate today is not about *whether* to incorporate humans into the operation of AWS, but *how* to do so.

Neither ASR nor DAIS give much detail on the precise extent of human involvement envisaged in the operation of AWS. There is a spectrum here. At the more permissive end, a 'human *on* the loop' model could be adopted whereby humans simply supervise the operation of the AWS and can at any time override the system and countermand its determinations. At the more restrictive end, a 'human *in* the loop' model could be adopted whereby humans must make any 'critical decisions' (any decisions involving targeting). As with any spectrum, there are costs and benefits depending on where one situates oneself. For example, the closer one sits to the permissive 'on the loop' end, the faster AWS will be able to react as they are free from requiring human input on critical decisions.

---

[27] Ministry of Defence, *Ambitious, Safe, Responsible: Our Approach to the Delivery of AI-Enabled Capability in Defence* (Ministry of Defence 2022) 13.
[28] ibid 3.
[29] Ministry of Defence (n 19) Executive Summary.
[30] Daniele Amoroso and Guglielmo Tamburrini, 'Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues' (2020) 1 Current Robotics Reports 187.

However, this comes at the expense of diluted, arguably tokenistic, human involvement. Conversely, the closer one sits to the restrictive 'in the loop' end, the more sound the critical decisions are likely to be as humans will have made them. However, this comes at the expense of a loss of speed, the loss of the ability to function in denied ('jammed') environments and an increased demand for human resources. From the perspective of IHL, the only place that one can justifiably sit on this spectrum is right in the middle given that, as noted above, IHL is predicated on maintaining an equilibrium between humanity and military necessity.

How the UK Government plans to strike this balance is the key issue that requires elucidation and scrutiny. ASR observes that the 'appropriate degree ... of "human control" need[s] to be considered carefully on a case-by-case basis'.[31] This may well be a fair point, and achieving the humanity and military necessity balance might well look different in different contexts. However, it poses a regulatory problem by leaving a lot of ambiguity over precisely what standard will be applied in any given situation. For example, ASR observes that, in some situations, the appropriate level of 'context-appropriate human involvement' may be none as it would pose an 'inappropriate constraint'. It highlights the situation where, for example, 'to defend a maritime platform against hypersonic weapons we may need defensive systems which can detect incoming threats and open fire faster than a human could react'.[32] That is an acceptable position in IHL as destroying an incoming hypersonic missile is unlikely to have any humanitarian cost, while its destruction achieves a clear military necessity. However, there may be a temptation to exclude human involvement in additional scenarios in order to exploit maximum military advantage from AWS. The risk of 'creep' here requires tight regulation and close monitoring.

Question 6: 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?

See response to Question 4.

Dr Elliot Winter
March 2023

[31] Ministry of Defence (n 27) 3.
[32] ibid.