2021 13th International Conference on Cyber Conflict
Going Viral
T. Jančárková, L. Lindström, G. Visky, P. Zotz (Eds.)
2021 © NATO CCDCOE Publications, Tallinn

The Role of Artificial Intelligence in Kinetic Targeting from the Perspective of International Humanitarian Law

Anastasia Roberts*
Lt Col, UK Army Legal Services
Office of Legal Affairs
SHAPE
Belgium

Adrian Venables
Senior Researcher
Department of Software Science
Tallinn University of Technology
Estonia
adrian.venables@taltech.ee

Abstract: The use of artificial intelligence (AI) in kinetic targeting is an emotive issue. Human Rights Watch (HRW) is a prominent campaigner against Lethal Autonomous Weapons Systems (LAWS) and has expressed concern these systems are fundamentally at odds with the international humanitarian law (IHL) framework for armed conflict. This framework places human control over the use of lethal force at the very heart of the targeting process. HRW asserts that the ceding of human control to AI-enabled capabilities may undermine and gradually erode the IHL framework, leaving the battlespace legally ungoverned and civilians unprotected. Concerns about the military use of AI have been exacerbated by the actions and narratives of some nations that are perceived as competing in an AI arms race. However, the debate about AI has been clouded by the fact that it focuses excessively on LAWS and human control. As a result, very little consideration is given to other potentially positive uses of AI technology in targeting. These include AI's role in Intelligence, Surveillance and Reconnaissance and Information Operations. This paper seeks to present a more nuanced examination of the role of AI in kinetic targeting and how it may affect compliance with IHL. The legal, ethical and technical arguments against and in favour of the use of AI will be examined. Finally, a way forward on this complex and emotive issue is proposed that offers a means to reinforce IHL whilst accepting that advances in technology will continue.

Keywords: *international humanitarian law, artificial intelligence, armed conflict, targeting*

* The lead author is a serving member of the British Army, currently working in NATO. The views expressed in this paper are those of the author alone and not of the Army, the UK Ministry of Defence or the UK Government or of NATO.

1. INTRODUCTION

There are many potential military uses for artificial intelligence (AI) enabled technology in armed conflict.[1] However, the one that arguably attracts the most attention is its use in kinetic targeting and, in particular, the employment of Lethal Autonomous Weapons Systems (LAWS). The use of LAWS is an emotive issue as demonstrated by the high-profile Human Rights Watch (HRW) coordinated Campaign to Stop Killer Robots.[2] This campaign has been calling for an outright ban on LAWS since 2012 and continues to gather momentum.

HRW's concern is that the use of LAWS may supplant the human role in the targeting process. It sees this as being fundamentally at odds with the international humanitarian law (IHL) framework for targeting in armed conflict that centres on human control over the use of lethal force.[3] HRW asserts that ceding human control to machines may undermine and gradually erode the IHL framework, leaving the battlespace legally ungoverned and civilians unprotected.[4] Unfortunately, this focus on LAWS and human control has clouded the broader debate regarding the use of autonomous AI capabilities in kinetic targeting and distracted attention from other potentially positive uses of the technology. Some states and commentators have argued that AI could, in fact, strengthen IHL compliance in armed conflicts.[5] Regrettably, these assertions are either lost in the emotion surrounding LAWS or are met with distrust and dismissed.

Fears that future conflicts will be dominated by autonomous AI technology have been exacerbated by what commentators are now referring to as an inter-state AI arms race.[6] This is being led by the United States, China, Russia and South Korea.[7] This rivalry is reflected in the narratives of the competing nations that assert that they must

1 'Artificial intelligence': the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages (*Oxford Reference*, 2020) <www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960> accessed 4 March 2021.

2 See campaign website at <www.stopkillerrobots.org/> accessed 4 March 2021.

3 Human Rights Watch and Harvard Law School International Human Rights Clinic, 'Killer Robots and the Concept of Meaningful Human Control – Memorandum to Convention on Conventional Weapons (CCW) Delegates' (April 2016) <https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control> accessed 4 March 2021.

4 Human Rights Watch and Harvard Law School International Human Rights Clinic, 'Losing Humanity: The Case against Killer Robots' (HRW, 2012) 36 <www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots> accessed 5 March 2021.

5 US Working Paper, 'Implementing International Humanitarian Law in the Use of Autonomy in Weapon Systems' (CCW/GGE.1/2019/WP.5, 2019) paras 13–15 < https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2019/gge/Documents/2019GGE.2-WP5.pdf > accessed 4 March 2021; Peter Marguelies, 'The Other Side of Autonomous Weapons: Using Artificial Intelligence to Enhance IHL Compliance' (2018) Roger Williams Univ Legal Studies Paper No. 182 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3194713> accessed 4 March 2021.

6 Matt Bartlett, 'The AI Arms Race in 2020' (*Towards Data Science*, 16 June 2020) <https://towardsdatascience.com/the-ai-arms-race-in-2020-e7f049cb69ac> accessed 29 November 2020.

7 Justin Haner and Denise Garcia, 'The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development' (2019) 10:3 Glob Policy <https://onlinelibrary.wiley.com/doi/full/10.1111/1758-5899.12713> accessed 29 November 2020.

develop AI technology before their adversaries do so, fuelling a sense of urgency.[8] These provocative narratives raise the concern that states will develop capability first and then deal with the legal and moral issues afterwards.

This paper seeks to present a more nuanced examination of the use of autonomous AI in kinetic targeting and how it may affect compliance with IHL. Three potential uses of AI will be reviewed against the IHL framework: LAWS, Intelligence, Surveillance and Reconnaissance activities and Information Operations. The legal, ethical and technical arguments against and in favour of using autonomous AI technology in kinetic targeting will then be examined. Finally, a way forward on this complex and emotive issue is proposed that offers a means to reinforce IHL whilst accepting that advances in technology will continue.

2. IHL FRAMEWORK FOR TARGETING

Fundamental Principles

The International Committee of the Red Cross (ICRC) describes IHL as 'a set of rules which seek, for humanitarian reasons, to limit the effects of armed conflict'.[9] This protects those who are not, or are no longer, taking part in hostilities and reduces the suffering of those who are, for example, by proscribing weapons that cause superfluous injury or unnecessary suffering.[10] IHL is founded on four fundamental principles that underpin the targeting process: necessity, humanity, distinction and proportionality.[11] It is the latter two that raise the most challenges for the use of AI technology in targeting. The principle of distinction requires a clear difference to be drawn between civilians and civilian objects and combatants and military objects. This is necessary because only the latter may be targeted, as civilians and civilian objects are protected under IHL. The principle of proportionality requires that the incidental civilian losses resulting from an attack, known generally as collateral damage, must not be excessive in relation to the expected military advantage. This requires the use of military judgement to assess and weigh the competing military and civilian impact before an attack is authorised.

Precautionary Measures

IHL's fundamental principles and detailed rules for their application in targeting, known as 'precautions in attack', are codified in Additional Protocol I to the Geneva

8 Edwin Mora, 'Pentagon: U.S. "in Danger" of Losing Dominance in Artificial Intelligence' (*Breitbart*, 11 December 2018) <www.breitbart.com/national-security/2018/12/11/pentagon-u-s-in-danger-of-losing-dominance-in-artificial-intelligence/> accessed 29 November 2020.

9 ICRC, 'What is International Humanitarian Law?' (ICRC, 2004) 1 <https://www.icrc.org/en/document/what-international-humanitarian-law> accessed 4 March 2021.

10 *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I)* (adopted 8 June 1977, entered into force 7 December 1979) 1125 UNTS 3 art 35(2).

11 ibid art 35(1); art 1(2); art 48; art 51(5)(b), art 57.

Conventions of 12 August 1949 (API).[12] Not all states are party to API but many of its provisions are considered to be customary international law (CIL) and are applicable in international and non-international armed conflict. API is clear that responsibility for applying the IHL principles and precautionary measures rests with those who plan or decide on an attack. Accordingly, whilst military commanders are supported by specialist personnel to inform their decision-making, including intelligence and legal officers, they are ultimately accountable.

When planning an attack, military commanders must do everything feasible to confirm that a selected target is military and not civilian. Furthermore, feasible precautions must be taken to avoid collateral damage, which will dictate how and when an attack is conducted. If collateral damage cannot be avoided altogether, it must be weighed against the anticipated military advantage. In this, the military commander is given a 'fairly broad margin of judgement'.[13] If the military commander assesses that the collateral damage is excessive, the attack cannot proceed. All these issues must be kept under constant review before and during a military operation. If the ongoing evaluation recognises that collateral damage is or will be excessive in relation to the military advantage expected, the attack must be cancelled or suspended. API is clear that issues of distinction and proportionality are subjective and 'must above all be a question of common sense and good faith for military commanders'.[14] It is this absence of human judgement and experience which makes the concept of autonomous AI capability so difficult to reconcile with the IHL framework.

Weapon Reviews

Article 36 of API requires state parties to review new weapons, means or methods of warfare (which are undefined) to ensure their compliance with IHL. There is no consensus as to whether this specific provision has the status of CIL thereby binding non-API states. The ICRC's view is that it does.[15] Other commentators assess that CIL requires at least a legal review of new weapons and means of warfare, if not methods.[16] In any event, according to the ICRC only a limited number of states are known to conduct legal reviews of weapons.[17]

The API commentary suggests that weapons and means are synonymous and distinguishes them from methods, which are narrowly defined as referring to how

12 Protocol 1 (n 10) art 57.

13 ICRC, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Yves Sandoz and others (eds), Martinus Nijhoff Publishers, 1987) para 2210.

14 ibid para 2208.

15 Kathleen Lawand, 'A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977' (ICRC, January 2006) 4.

16 Jeffrey T Biller and Michael N Schmitt, 'Classification of Cyber Capabilities and Operations as Weapons, Means, or Methods of Warfare' (2019) 95 INT'L L STUD 179, 186; William Boothby (ed), *New Technologies and the Law in War and Peace* (CUP 2018) 17.

17 Lawand (n 15) 5.

weapons are used.[18] However, it has been argued in the context of cyber operations that the term 'methods' is broader than this. It encompasses all tactics, techniques and procedures (TTPs) for carrying out military operations involving the conduct of hostilities, not just targeting. This decouples methods and weapons.[19] The state practice of Germany and Belgium seems to support this broader assessment but, more generally, states do not appear to have addressed this issue.[20]

Defining what constitutes a method of warfare is essential to determining whether a capability that is not obviously a weapon falls within the review process. Whether autonomous AI capability used in kinetic targeting is subject to legal review is an important element in considering whether such capability can be reconciled with the IHL framework. A legal review would need to ensure that the capability is not inherently indiscriminate and that it can apply the targeting rules, as applicable to its specific function.[21]

3. POTENTIAL MILITARY USES OF AI IN TARGETING

This section will explore three potential military uses of AI in the targeting process: LAWS, Intelligence, Surveillance and Reconnaissance (ISR) activities and Information Operations (IO).

LAWS

There is no internationally recognised definition of LAWS. The UN Group of Governmental Experts on Lethal Autonomous Weapons Systems has yet to agree on the issue.[22] In this paper, we define LAWS as a weapons system that, through the use of AI technology, can independently select and use force against targets without human control. This is likely to be achieved by the application of machine learning (ML).[23] This self-learning ability distinguishes LAWS from the so-called semi-autonomous or automated weapons systems already in use with the military. These weapons systems respond in a predefined and programmed manner to certain stimuli and are generally used in narrow defensive roles such as close anti-aircraft defence systems. A recent study of the military application of AI suggests that fully autonomous weapons

18 API commentary (n 13) para 1957.
19 Biller (n 16) 200.
20 Vincent Boulanin and Maaike Verbruggen, 'SIPRI Compendium on Article 36 Reviews' (SIPRI Background Paper, 2017) 3, 6 <https://sipri.org/publications/2017/sipri-background-papers/sipri-compendium-article-36-reviews> accessed 4 March 2021.
21 Boothby (n 16) 139.
22 Chair, 2020 Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 'Commonalities in National Commentaries on Guiding Principles' (UN 2020) para 5 <https://reachingcriticalwill.org/disarmament-fora/ccw/2020/laws/documents> accessed 8 March 2021.
23 'Machine learning': the capacity of a computer to learn from experience, i.e. to modify its processing on the basis of newly acquired information (*Oxford Reference*, 2020) <https://www.oxfordreference.com/view/10.1093/acref/9780195314496.001.0001/acref-9780195314496-e-1161> accessed 29 November 2020.

systems have not yet been developed. However, both China and the US have built systems that could assume this function with simple software modifications.[24]

Intelligence, Surveillance and Reconnaissance (ISR)

Surveillance is the persistent monitoring of a target. Reconnaissance is information gathering conducted to answer a specific military question. Intelligence is the final product derived from these activities, fused with other information, which is then used to support military decision-making, including targeting.[25] It is reported that ISR is one of the areas attracting the most investment in military AI and that AI will enable dramatic improvements in this area.[26]

It is anticipated that AI will enable large amounts of information from multiple data sources to be processed and synthesised more quickly and effectively.[27] Considerable advances have already been made in image processing with some automated image-recognition and object-detection capabilities that now surpass human ability.[28] Such tools will be key to positive target identification through facial recognition but also by identifying whether observed conduct is or is not hostile. For example, is the object next to an individual digging at the side of the road an IED or a drainage pipe? Similarly, facial expression analysis could help identify hostile intent such as in the case of suicide bombers.

Information Operations (IO)

IO involves the military use of information to create a desired effect on the will, understanding and capability of adversaries and other approved parties.[29] The Internet now plays a dominant role in IO, supporting more traditional influence methods such as leaflet campaigns and radio broadcasts. It is reported that AI is already able to analyse large amounts of open-source online information to understand how to influence target audiences and tailor messaging to them.[30] As AI develops, it will also be used increasingly to create influence effects by generating, for example, autonomous online agents to engage with target audiences through social media.[31] Given the increasing prevalence of AI-generated deepfakes on the Internet, AI is also likely to be used to create and disseminate disinformation.[32] Through these means, IO

24 Forrest E Morgan and others, 'Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World' (RAND Corporation, 2020) 61.

25 'Joint Intelligence, Surveillance and Reconnaissance' (NATO, March 2021) <https://www.nato.int/cps/en/natohq/topics_111830.htm> accessed 9 April 2021.

26 Morgan (n 24) 20.

27 Paul Scharre, 'Artificial Intelligence: Risks and Opportunities for SOF' in Zachary S Davis and others (eds), *Strategic Latency Unleashed: The Role of Technology in a Revisionist Global Order and the Implications for Special Operations Forces* (LLNL CGSR 2021).

28 Morgan (n 24) 13–14, 17.

29 NATO, Allied Joint Publication 3.10 – Allied Joint Doctrine for Information Operations (NATO 2009) para 0107 <https://info.publicintelligence.net/NATO-IO.pdf> accessed 7 March 2021.

30 Morgan (n 24) 20.

31 ibid.

32 Robert Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 CalLRev 1753.

may be used to facilitate the kinetic targeting process. This may involve ensuring that a target is in a desired location at a particular time or that an area is clear of civilians.

Method of Warfare?

Neither IO nor ISR activities involve the use of force unless they are integral to a weapons system. However, as noted, they may facilitate the targeting process or support targeting decisions. On this basis, the issue of whether they must remain under human control to satisfy IHL is as relevant as it is for LAWS. An autonomous capability may, for example, incorrectly identify a civilian as a target or may perfidiously feign protected status under IHL to entice a target to a particular location. Without a degree of human control to identify and prevent such occurrences, violations of IHL may result. There is a danger that the possible IHL implications of these capabilities may be missed due to the focus on LAWS.

An argument could perhaps be made that AI-enabled ISR and IO capabilities are methods of warfare and should be subject to Article 36 legal review, at least for state parties to API. This is based on the broader definition of methods of warfare as TTPs for carrying out military operations involving the conduct of hostilities, rather than simply relating to how a weapon is used. Thinking beyond LAWS would allow for a clearer discussion on the scope of the legal review process.

4. THE CASE AGAINST AI

In setting out the arguments for and against the use of autonomous AI capability in targeting, three areas are examined: legal, ethical and technical.

Legal Arguments

The primary legal concern is whether autonomous AI capabilities could even be capable of compliance with the IHL framework for targeting because they lack the requisite human judgement and experience that underlie the application of the legal tests.[33]

Distinction is an increasingly complex issue at a time when adversaries are often indistinguishable from the civilian population and will habitually alternate between targetable and non-targetable status. Often the only way to make this identification on the ground is by assessing someone's activity to discern if they are directly taking part in hostilities at a particular time, rendering them targetable. This is challenging as there is no precise definition of what constitutes direct participation in hostilities.

33 HRW (n 4) 30–34.

The ICRC CIL study[34] proposes a definition that has not been accepted by all states.[35] Moreover, while the ICRC study is helpful at a doctrinal level, the situation on the ground is often informed by the operational context and intelligence picture.

It has been suggested that the ability to discern hostility requires an understanding of an individual's mental state, which in turn relies on emotional intelligence.[36] One example of this might be celebratory weapons fire, a cultural practice in many countries. Without understanding the cultural and emotional context, autonomous AI may interpret this weapons fire as hostile activity. Moreover, as there is no clear consensus on what constitutes direct participation in hostilities and noting the key variables of operational context and intelligence, it is difficult to see how an AI capability can be programmed to learn to identify it. This would apply equally to LAWS or to standalone ISR capabilities that identify and track targets.

Dual use issues are also problematic. This is the use by the adversary of a protected civilian object for hostile purposes. In these circumstances, it must be determined whether the object has lost its protection and become a legitimate military objective. This occurs when it is making 'an effective contribution to military action' and targeting it will accordingly provide 'a definite military advantage'.[37] Again, there are no clear criteria to assess this: it is a matter of the military commander's own judgement and experience.

This is also the case with proportionality. It is difficult to see how an autonomous AI capability could conduct the required balancing exercise between military advantage and collateral damage. How will it assess the military value of the target, noting that it will be different in every attack? Similarly, how will it ascribe a value to the human life or lives involved in the context of the wider operation? These are more than ethical arguments; these are issues about compliance with the legal framework. While computer modelling and simulation are now an integral part of a collateral damage estimate for targeting, the software does not make the proportionality decision. It simply informs the military commander's decision, as does the advice received from other specialist personnel, such as legal and intelligence officers. The ICRC's position is that 'preserving human control and judgement will be an essential component for ensuring legal compliance'.[38]

Related to legal compliance is the issue of legal accountability. International criminal law provides an established framework for dealing with violations of IHL by

34 Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law Volume I Rules* (ICRC, CUP 2005).

35 John B Bellinger and William J Haynes, 'A US government response to the International Committee of the Red Cross study Customary International Humanitarian Law' (2007) 89:866 IIRC 4.

36 HRW (n 4) 31.

37 Protocol 1 (n 10) art 52.

38 ICRC, 'Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach' (ICRC, 2019) 9 <www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach> accessed 7 March 2021.

individuals. As with IHL, this framework is human-centric. If a military commander deliberately orders an attack on civilians, this is a war crime and is subject to criminal prosecution. But who is accountable for such an attack carried out or decided on by an autonomous AI capability? In cases where an autonomous AI system is intentionally manipulated by humans to commit a war crime, such as its deliberate programming to target civilians, accountability is clear. This is described as the Perpetration-by-Another liability model.[39] In these circumstances LAWS are no different to any other weapons used to commit an offence. However, perhaps a more likely and more problematic scenario is the unintended malfunction of a capability, whether LAWS or an IO or ISR capability used in support of targeting.

In the Natural-Probable-Consequence liability model, if the malfunction was the natural or probable consequence of someone's conduct, and was therefore foreseeable, that person will be held criminally accountable.[40] However, this model may be too simplistic to account for what are likely to be complex situations. In the case of the developer, for example, liability will hinge upon their level of involvement in the capability development process. If the developer was not given enough detail of the likely operational environment, including use cases and the IHL framework, it is difficult to see how foreseeability could be established.

Another suggestion is to distribute criminal accountability between key stakeholders in the creation and use of the AI capability.[41] This could include the operator, military commander, programmer, manufacturers, defence personnel involved in the acquisition process, and senior politicians. However, such an approach is likely to be evidentially challenging, politically charged and protracted, and unlikely to satisfy the victims' families. It also risks distributing accountability so widely that no individual can be held responsible for a failure under the criminal standard of proof.

Finally, the Direct-Liability model holds the AI capability itself criminally accountable.[42] Even if this was legally possible, which is debatable, it is likely to offend victims' families, making a mockery of the legal framework, and does not merit further discussion.

The lack of a clear accountability framework for IHL violations by autonomous AI capabilities is a significant impediment to the use of this technology by the military. Human accountability is a cornerstone of the IHL framework for targeting in armed conflict, and any dilution of this principle will undermine that framework.

39 Gabriel Hallevy, 'The Basic Models of Criminal Liability of AI Systems and Outer Circles' (11 June 2019) 1–4 <https://ssrn.com/abstract=3402527> accessed 7 March 2021.

40 ibid 4–8.

41 Tetyana (Tanya) Krupiy, 'Regulating a Game Changer: Using a Distributed Approach to Develop an Accountability Framework for Lethal Autonomous Weapon Systems' (2018) 50, GJIL, 45–70.

42 Hallevy (n 39) 8–15.

Ethical Arguments

Even if it could be demonstrated that an autonomous AI capability can comply with IHL, there is still significant opposition to the use of such capability on ethical grounds. It is argued that ceding life and death decisions to machines would deprive people of their inherent dignity and result in the dehumanisation of warfare because military decision-making would be stripped of emotion.[43] It is asserted that even when it would be lawful to use force, conscience often acts as a final barrier against killing civilians.[44] The role of human emotion over and above legal compliance was demonstrated by the International Security and Assistance Force (ISAF) policy of 'courageous restraint' in Afghanistan in 2009.[45] This encouraged military personnel to refrain from the use of force, even when legally permissible, to spare the civilian population even if at a cost to themselves or other ISAF personnel. An autonomous AI capability will not have a conscience or the human emotion to instinctively know when restraint should be exercised.

Technical Arguments

Three technical issues have emerged which suggest that AI may be unable to comply with the IHL framework. The first issue is bias. While this is not a trait usually associated with machines, it has been demonstrated that ML technology can display preferences. This is thought to be caused by the data sets it is trained with if they are unrepresentative or reflect prejudice.[46] This issue could have serious implications in the targeting process. By way of example, the training data for an ISR capability might contain a disproportionate number of images of individuals with a particular ethnicity. As a result, the capability may learn that persons of this group are *prima facie* adversaries, and therefore targetable. The question of skewed training data sets may be of particular concern where AI technology is developed internally by the military, noting that the demographic of most Western militaries is predominantly white male.

The second technical issue is the linked problems of predictability and reliability. AI/ML technology is not programmed to make decisions in a particular way but rather develops its own decision-making process by analysing and modelling its training data. As a result, developers are often unable to explain how AI technology arrived at a decision because of its complex evolving internal processes. This is known as the black-box effect.[47] If the AI decision-making process is not fully understood, it is impossible to predict how it will respond in any given situation, which reduces confidence in the system. This problem is exacerbated by the fact that the very nature

43 HRW (n 3).
44 HRW (n 4) 37–39.
45 Joseph H Felter and Jacob N Shapiro, 'Limiting Civilian Casualties as Part of a Winning Strategy: The Case of Courageous Restraint' (2017) 146:1 AAAS 44.
46 Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?* (HL 2017-19, 100) paras 107–121.
47 ibid paras 89–94.

of ML means that training is not finite and that a capability may keep learning from external environmental factors even when deployed.[48]

This leads to the third and most important issue of explainability; that is, the ability to describe how a decision has been made. This is a key aspect of the targeting process as a military commander must be able to explain their decision-making process to demonstrate IHL compliance. This clearly links to the issue of accountability. Accordingly, the output of any autonomous AI technology must include an analysis of its decision-making process and the factors relied on. The black-box effect described above suggests that this may be technologically impossible at this time.

5. THE CASE FOR AI

The arguments against the use of autonomous AI capability by the military in targeting must be recognised. However, the potential for AI technology to also strengthen IHL compliance is often overlooked.

Legal Arguments

Targeting in armed conflict can be fast-moving and pressured, with short decision-making windows and an imperfect intelligence picture. This is why targeting decisions are judged by the standard of a reasonable military commander and are made on the known circumstances at the time and the information available. However, if AI-enabled ISR capability develops as predicted, it will result in the faster production of a more accurate intelligence picture. The reported advances in image and facial recognition and expression analysis will provide greater certainty in distinction, positive target identification and more precise collateral damage estimates. Even the use of LAWS may in fact strengthen precautions, as when using unguided or conventional munitions, a military commander has no control over an attack once it has commenced. This means that it may not be possible to stop the attack if the collateral damage estimate changes or target identification is lost. Even with modern precision-guided munitions, control in flight remains limited. In contrast, it is suggested that LAWS will be able to abort or delay an attack as soon as it identifies a change in conditions.[49] As for IO, AI-enabled capabilities could strengthen IHL by identifying and facilitating non-kinetic alternatives to kinetic targeting, which informs the consideration of the principles of necessity and proportionality. Moreover, they may aid in reducing collateral damage by providing an effective means of warning civilians of an attack or otherwise ensuring that they are out of the target area.

In terms of the concern that the use of AI is incompatible with the concept of legal accountability, it is true that international and domestic criminal law do not appear to

48 ICRC (n 38) 10–11.

49 Ryan Khurana, 'In Defence of Autonomous Weapons' (*The National Interest*, 14 October 2018) <https://nationalinterest.org/feature/defense-autonomous-weapons-33201> accessed 7 March 2021.

provide a readily adaptable framework. This concern could be addressed by focusing on the accountability of the state rather than individuals. However, this approach poses its own challenges as a state cannot be held directly accountable under international criminal law. The International Humanitarian Fact-Finding Commission's ability to investigate alleged violations of IHL by state parties to an armed conflict is dependent on the consent of those parties, which is also required for any disclosure of the investigation report.[50] The injured state could refer the alleged violation to the UN but this is a political rather than legal route and the UN's response will be dictated accordingly.

Another possible option is to develop the law of state responsibility to address a state's negligent deployment, in an armed conflict, of untested or inadequately tested AI capability that operates in breach of IHL. For example, directly attacking and killing civilians in violation of the principle of distinction. Under the Draft Articles on Responsibility of States for Internationally Wrongful Acts (ASR),[51] where a state seriously violates a peremptory norm, which includes the basic rules of IHL, the legal interest of the whole international community is affected. This empowers any state to invoke the responsibility of the offending state before the International Court of Justice (ICJ), not just the injured state.[52] Indeed, there is arguably an obligation on other states to do so.[53] In the context of ICJ proceedings, whether or not the state conducted a legal review in accordance with CIL or Article 36 of API may be an important feature.

Accordingly, instead of trying to formulate new accountability models to accommodate autonomous AI capability, it may be more productive to focus on strengthening existing mechanisms for holding states to account for the development and use of such capabilities. Noting that few states appear to comply with either Article 36 of API or CIL legal review obligations, this should include strengthening legal review compliance. This could include clarifying the scope of the review process in terms of methods of warfare. A clearer accountability framework may well provide a natural brake on the rapid development of autonomous AI capability and an incentive to demonstrate compliance with the legal review process.

Ethical Arguments

It has been suggested that ethical concerns about the dehumanisation of warfare ignore the fact that IHL is deliberately structured to counter rather than endorse the effects of human emotion on the battlefield.[54] Conflict is inherently fast-paced, physically

50 See the International Humanitarian Fact-Finding Commission's website <https://www.ihffc.org/index.asp?page=home> accessed 7 March 2021.

51 ILC, 'Report of the International Law Commission (ILC) on the Work of its Fifty-third Session' (2001) UN Doc A/56/10 29.

52 ibid ILC commentary to art 40 ASR para 5; ILC commentary to art 48 ASR paras 8–9; Marco Sassòli, 'State responsibility for violations of international humanitarian law' (2002) 84: 846 IRRC 401, 413–414.

53 ILC (n 51) art 41(1), (2).

54 William H Boothby, *Weapons and the Law of Armed Conflict* (2nd edn, OUP 2016) 343.

demanding, mentally draining and stressful. Humans are likely to experience emotions such as fear, exhaustion and anger that may adversely influence their decision-making. The loss of comrades on the battlefield may affect their judgement and increase the risk of unlawful conduct. However, by imposing rules on the conduct of warfare, IHL seeks to control the impact of these emotions. As autonomous AI capabilities will be immune to emotion and respond to events objectively in accordance with their programming, they could actually provide exactly what IHL is seeking to achieve; that is, the best possible protection for civilians and combatants.

In light of this, it has been argued that it is no longer necessary 'to cling to a human-centred approach' to IHL on the assumption that this protection is best achieved by people. The peremptory rejection of autonomous AI technology cannot be justified by arguments about human dignity when this technology offers an alternative and potentially superior means to achieve IHL's humanitarian goals.[55] It could also be argued that someone facing death in armed conflict is more likely to be concerned about the actual loss of their life rather than who or what decides to take it. In fact, given the often-remote nature of targeting, it is likely that the origin of the decision will not be clear in any event. However, the key point is accountability in the event of the unlawful taking of life and this is perhaps where ethical arguments should focus.

Technical Arguments

It could be argued that the technical concerns about AI technology are equally applicable to humans. Humans can be biased, unpredictable and unreliable and make seemingly 'illogical and impenetrable' decisions.[56] Military commanders are only held to a reasonable standard so why would we expect more from machines? Moreover, the potential technical advantages of AI are undeniable. If realised, these will improve support to the targeting process and IHL compliance.

However, the potential benefits in terms of IHL compliance will depend entirely on how autonomous AI capabilities are programmed and utilised. Careful and conscientious development practices and compliance with the legal review process are necessary to ensure that new capabilities are legally compliant and technologically protected against interference and misuse. Technological advances will enable the imposition of constraints and increasingly complex rule sets to control the behaviour of AI-based systems. In this respect, it is important that lawyers, both military and private sector, are involved in the development of autonomous AI capability as early as possible. If legal compliance issues are identified early, rule sets to control the behaviour of the capability can be incorporated into the design, becoming an integral feature of the capability, rather than an afterthought.

55 Masahiro Kurosaki, 'Toward the Special Computer Law of Targeting: "Fully Autonomous" Weapons Systems and the Proportionality Test' in Claus Kreß and Robert Lawless (eds), *Necessity and Proportionality in International Peace and Security Law* (The Lieber Studies Series Book 5, 2021).

56 Kenneth Anderson and others, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems' (2014) 90 INT'L L STUD 386, 393.

6. CONCLUSIONS AND RECOMMENDATIONS

As noted in the introduction, the purpose of this paper was to examine the role of autonomous AI capability in kinetic targeting and its potential impact on IHL compliance. The aim was to present a more balanced analysis than is often seen because of the narrow focus on LAWS. The arguments for and against the use of such capability have been presented, hopefully demonstrating that there are in fact two sides to this debate. On the one hand, it is difficult to see how autonomous AI capability can comply with the IHL rules for targeting, given the centrality of the human role. The question is also whether this should even be attempted for ethical reasons. Let us be honest, the concept of a Terminator-style machine holding human life in its hands fundamentally feels wrong. However, on the other hand, autonomous AI technology presents clear opportunities for strengthening IHL compliance. To appreciate this fact, it is necessary to look beyond 'killer robots' to the wider use of the technology.

This leads to the fact that the current polarised debate about autonomous AI capability is unhelpful. If there is to be any meaningful control over its development, a sensible and informed middle ground must be found. Calls to ban the use of autonomous AI in military systems are both unrealistic and naïve. Any such ban will push capability development into an ungoverned space with no possibility of control or debate. The reality is that capabilities are being developed now and, without safeguards, there is a real risk of technological development outpacing IHL. This reality leads us to make several recommendations to ensure the ongoing relevance of, and respect for, IHL.

First, it should be accepted that there will never be sufficient state consensus to secure a total ban on LAWS or to introduce any new legal controls on the role of AI in military systems. The desire to employ new technology to achieve an advantage on the battlefield has been a constant feature of conflict and will not change. Instead, international organisations such as the UN should focus their efforts on supporting the development of non-binding guidance on how states should apply the existing IHL framework to this complex area.

Second, states and international organisations should specifically seek to strengthen compliance with the legal review process. Development of the non-binding guidance suggested above could be a vehicle for this. This includes clarification as to when capabilities that are not obviously weapons, but do support the kinetic targeting process, should fall under the review process as methods of warfare. A great strength of the IHL framework for armed conflict is that it is inherently flexible and is designed to adapt to incorporate new technology. States, international organisations and the public must use and trust this framework or risk losing it altogether.

Third, and linked to the above, the role of lawyers in the development of AI is key. This relates to both military and civilian lawyers, noting that much capability development takes place in the private sector. Lawyers can inform a capability's development by providing the detail of the legal framework it will need to operate within. This will allow for the development of technical rules to control the capability's behaviour. Legal compliance will then be an integral part of the design, rather than an afterthought.

Fourth, states and organisations should seek to clarify the legal accountability framework. If a clear accountability framework can be identified and, if necessary, strengthened, this may provide a deterrent effect and also slow the development of autonomous AI capability to allow for the proper consideration of legal, ethical and technical issues.

Finally, the sense of urgency being ascribed to AI development by some states, fuelled by the media and references to an AI arms race, needs to be tempered. These narratives are not helping to achieve a balanced debate on this issue. States are likely to secure greater public support and trust in their AI development initiatives if they adopt a measured, rational and open approach.