INTERNATIONAL REGULATIONS ON ARTIFICIAL INTELLIGENCE IN THE MILITARY

Adequate or Outdated?

The 21st century has seen many new threats to national and international security, such as terrorism. In response, states have begun investing greatly in artificial intelligence to complement their respective defence systems. Given that public international law has remained largely unchanged over the past half a century, it may appear that public international law is ill-equipped to address the advent of artificial intelligence. However, this article argues that the existing legal framework is largely adequate in regulating the future use of artificial intelligence by states. This article also proposes certain modifications to the existing framework to enhance the framework's effectiveness.

LOW Yan Lin[1]
LLB (Singapore Management University).

I.   Introduction

1   At the border of the Gaza strip, Israel's famed Iron Dome system has proven able to detect and counter incoming enemy missiles for the past decade.[2] In South Korea, the Super aEgis II surveils the demilitarised zone between North Korea and South Korea and can decide whether to shoot intruders.[3] In short, states have already been employing intelligent machines along their borders. These machines *can* function without any human supervision as they are equipped with artificial intelligence ("AI"): the science allowing computer systems to perform tasks which

---

2   Ellen Loanes, "The US Military is Buying Israel's Battle-proven Iron Dome that Destroys Rockets" *Business Insider* (14 August 2019).
3   Daniele Amoroso & Guglielmo Tamburrini, "The Ethical and Legal Case against Autonomy in Weapons Systems" (2017) 18 *Global Jurist* 1 at 3.

normally require human intelligence.[4] However, given the limitations of current technology, these machines still need to be manned by human operators who can either override the system's firing decision or make the final command to fire;[5] as far as public knowledge is concerned, the technology allowing states to deploy fully autonomous weapons with confidence does not yet exist. United Nations ("UN") Special Rapporteur Christof Heyns has defined fully autonomous weapons as those which can select and engage targets without further intervention by a human operator.[6] As more states continue to invest in the use of advanced AI in their militaries,[7] it is a matter of time before fully autonomous weapons are deployed.

2   Many have argued that current public international law, typically grounded in consent and consensus amongst the international community, does not allow for the deployment of fully autonomous weapons. The 2018 Group of Governmental Experts Meeting on Lethal Autonomous Weapons, for instance, was indicative of a general distrust of fully autonomous weapons – numerous states had stressed at the meeting the need for "human control, supervision, oversight, or judgment [over] the use of force" by autonomous weapons.[8] Moreover, in the last two years, the European Union and 30 other states have also called for a treaty ban on fully autonomous weapons.[9] However, this article argues that current public international law rules can accommodate the introduction of fully autonomous weapons and can set the parameters for exactly which types of fully autonomous weapons are permitted.

3   As aptly put by Judge Lachs, "the great acceleration of social and economic change, combined with that of science and technology, have confronted law with a serious challenge: one it must meet, lest it lag

4   Oxford Dictionaries Online, "English Definition of 'Artificial Intelligence'" https://www.lexico.com/en/definition/artificial_intelligence (accessed 11 August 2019).

5   Daniele Amoroso & Guglielmo Tamburrini, "The Ethical and Legal Case against Autonomy in Weapons Systems" (2017) 18 *Global Jurist* 1 at 3–4.

6   United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* UN Doc A/HRC/23/47 (9 April 2013) at para 38.

7   Sandeep Singh, "Global Military Robots Market by Platform, Application, Operation, Payload and Region 2014–2025: Growth Opportunity and Business Strategy" *Medium* (23 March 2019).

8   *Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System* UN Doc CCW/GGE.1/2018/3 (23 October 2018) Annex III, para 12.

9   Stop Killer Robots Organisation, "Campaign to Stop Killer Robots: Country Views on Killer Robots" (25 October 2019) <https://www.stopkillerrobots.org/wp-content/uploads/2019/10/KRC_CountryViews_25Oct2019rev.pdf (accessed 1 January 2020); European Parliament, *European Parliament Resolution of 12 September 2018 on Autonomous Weapon Systems* 2018/2752(RSP) (12 September 2018).

even farther behind events than it has been wont to do".[10] Since public international law develops very slowly,[11] it is accepted that there are currently no rules specifically dealing with the use of militaristic AI by states. Yet, this does not necessarily mean that a "serious challenge" in the form of a lacuna in public international law has emerged.

4   This article shall focus on AI used in military robots in times of armed conflict, one type of which exists where there is "protracted armed violence between governmental authorities and organised armed groups within a State".[12] As the applicable law in times of armed conflict is international humanitarian law ("IHL"),[13] Part II[14] discusses whether robot can be regulated by the basic rules of IHL. As these basic rules of IHL are applied when states carry out their domestic weapons reviews, Part III[15] evaluates the effectiveness of this weapons review obligation. After passing the weapons review, the military robot will be deployed. Part IV[16] thus recommends specific ways to increase both state and individual accountability after deployment. Part V[17] concludes.

II.   Complying with international humanitarian law

5   Assume that an armed conflict of sufficient intensity has arisen in State A. Over the past year, a group of religious fanatics from State A had morphed into a terrorist organisation called the "X Militias". With a military-like command structure and a huge amount of resources, the X Militias have managed to procure a considerable stockpile of weapons. These militias have also begun threatening the lives of State A's citizens. To subdue the X Militias, State A has deployed its armed forces against them.

---

10   *North Sea Continental Shelf (Federal Republic of Germany v Netherlands)* (1969) ICJ Rep 3 (Dissenting Opinion of Judge Lachs) at 230.

11   George Galindo & Cesar Yip, "Customary International Law and the Third World: Do Not Step on the Grass" (2017) 16 *Chinese Journal of International Law* 251 at 262–263; Nico Krisch, "International Law in Times of Hegemony: Unequal Power and the Shaping of the International Legal Order" (2005) 16 *European Journal of International Law* 369 at 377.

12   *Prosecutor v Tadic* IT-94-1-A, Decision on Defense Motion for Interlocutory Appeal on Jurisdiction (2 October 1995) at para 70.

13   *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion* (1996) ICJ Rep 226 at [51].

14   See paras 5–31 below.

15   See paras 32–42 below.

16   See paras 43–62 below.

17   See paras 63–65 below.

6   As a result of their wealth and technological expertise, State A has successfully incorporated AI into its military robots. Three types of these robots have received great international attention – Type 1, Type 2, and Type 3 robots. Type 1 robots can target enemy objects without human involvement. During a counter-insurgency operation in one of State A's urban slums, State A's soldiers received news that armed rebels may be hiding inside a civilian household. The soldiers, together with a Type 1 attack robot, approach the house. Two men in the house are carrying *kirpan* daggers for purely religious reasons. Unbeknownst to the soldiers, no insurgents are present. Due to the commotion caused by the soldiers outside the house, the family's dogs race outside and start barking. The two men run outside and start shouting to get the dogs to calm down. At first glance, the men are two quickly approaching targets carrying weapons and running outside the house in an agitated manner. Would the robot be able to distinguish a religious artefact from a weapon and properly interpret the situation? If not, will this robot be deemed unlawful under IHL?

7   Separately, State A's Type 2 robots incorporate nanotechnology and can fly. This "flying nanobot" can enter a victim's nose or mouth and subsequently kill the victim with a micro-explosion in the lungs. When the explosion occurs, the victim will experience prolonged excruciating pain due to the damaged blood vessels in the lungs. Can such a robot ever be deployed by State A?

8   As for State A's Type 3 robots, through analysing large amounts of data and continuous self-learning, they are able to advise on the conduct of warfare (that is, choose a route to reach the target area, decide whether to deploy weapons, and, if so, decide which weapon system to deploy). Thus far, these robots have given effective advice which allowed State A to take over an X Militia command centre. However, the advice has also led to incidental civilian harm. When giving advice, is it possible for these robots to adhere to the existing guidelines for warfare under IHL?

9   As fully autonomous weapons do not yet exist, the above hypothetical scenario will be used to illustrate the perceived gaps in public international law. One must not forget that "war is distinguishable from murder and massacre only when restrictions are established on the reach of battle".[18] IHL lays out these restrictions on states and individuals and seeks to minimise suffering during wartime.[19] The International Court of Justice ("ICJ") has also made clear that in evaluating the legality

---

18   Michael Walzer, *Just and Unjust Wars* (New York: Basic Books, 2000) at p 42.
19   Laurie Blank *et al*, *International Law and Armed Conflict: Fundamental Principles and Contemporary Challenges in the Law of War* (Wolters Kluwer, 2016) at p 5.

of new weapons the applicable law is the fundamental principles of IHL.[20] Accordingly, in order to establish that current IHL rules are sufficient to regulate fully autonomous weapons, this part will examine the applicability of (a) the rule of distinction; (b) the rule against unnecessary suffering; (c) the rule of proportionality and (d) the Martens Clause.

A.   *Distinction*

10   This article argues that the rule of distinction is capable of functioning as a standard for military deployment of robots. As codified in the Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of International Armed Conflicts[21] ("AP I"), distinction requires that parties to an armed conflict distinguish between combatants and civilians, as well as between military and civilian objects.[22] However, if civilians "take a direct part in hostilities", such as by transporting combatants to the fight, they can be targeted as well.[23] This issue of distinction is especially pertinent in situations of modern warfare as combatants rarely wear uniforms and seek to blend in with the civilian population.[24]

11   In the above hypothetical situation, it is uncertain whether State A's Type 1 robots would be able to identify the men carrying the *kirpan* daggers as non-combatants. What is more certain is that, in that scenario, the human soldiers would be able to identify those men as civilians and be guided by the rule of distinction. This illustrates why several commentators, including a UN Special Rapporteur, have argued that fully autonomous weapons should not be trusted to carry out targeting activities.[25] However, the necessary conclusion reached by

---

20   *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion* (1996) ICJ Rep 226 at [34].
21   (Protocol I) 1125 UNTS 3 (8 June 1977; entry into force 7 December 1978) (hereinafter "Additional Protocol I").
22   Additional Protocol I, Arts 48 and 51.
23   Additional Protocol I, Art 51(3); *Public Committee against Torture in Israel v Government of Israel* HCJ 769/02 (13 December 2006) at para 35.
24   Kjølv Egeland, "Lethal Autonomous Weapon Systems under International Humanitarian Law" (2016) 85 *Nordic Journal of International Law* 89 at 98.
25   Amanda Sharkey, "Autonomous Weapons Systems, Killer Robots and Human Dignity" (2019) 21(2) *Ethics and Information Technology* 75 at 76; Robert Sparrow, "Robots and Respect: Assessing the Case against Autonomous Weapon Systems" (2016) 30 *Ethics in International Affairs* 93 at 98–100; United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* UN Doc A/HRC/23/47 (9 April 2013) at para 67; Marcello Guarini & Paul Bello, "Robotic Warfare: Some Challenges in Moving from Noncivilian to Civilian Theaters," in *Robot Ethics: The Ethical and Social Implications of Robotics* (Patrick Lin, Keith Abney & George Bekey eds) (MIT Press, 2011) ch 8
(cont'd on the next page)

these commentators is that all fully autonomous weapons would never be able to adhere to the principle of distinction. With respect, this article argues that a broad-brush approach should not be taken. Instead, fully autonomous weapons should be evaluated based on their technical competence.

12   Fully autonomous weapons can be specially designed to enable them to adhere to the rule of distinction. For example, automated target recognition is already a feature in modern weapons (which are not fully autonomous). Such a function may be used to identify, acquire, track, cue, or prioritise targets for a human operator. This target recognition technology is being used in automated sentry guns, such as those in the demilitarised zone between South Korea and North Korea;[26] and sensor-fused munitions, such as Sweden's BONUS System.[27] While humans deploy these systems, the machines themselves can fire at targets or detonate automatically based on predetermined parameters.[28]

13   Robots can adhere to the rule of distinction. It would all depend on the robot's specific algorithm and the *environment* it is used in.[29] Regarding algorithms, robots can be designed to recognise bodily postures and gestures (for example, hands raised in surrender), understand emotional expressions, and look past deceptive intentions and actions.[30] Hence, these types of robots are clearly more suited to be deployed in *urban areas* with civilians. In contrast, robots with basic sensors or analytic capabilities can easily be deployed in *underwater combat* where there are no civilians. Crucially, when we engage in the analysis of what is the appropriate environment for robots, then we already are applying the rule of distinction to regulate the use of robots.

14   As stated earlier,[31] opponents of fully autonomous weapons have stressed the need for human control and oversight. Indeed, in the

---

        at p 138; Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Routledge, 2009) at p 99.

26   Alexander Velez-Green, "The Foreign Policy Essay: The South Korean Sentry – A 'Killer Robot' to Prevent War" *Lawfare* (1 March 2015).

27   BAE Systems, "Swedish Army Selects Bofors 155mm BONUS Munition" (2 November 2017) <https://www.baesystems.com/en-us/article/swedish-army-selects-bofors-155mm-bonus-munition> (accessed 10 August 2019).

28   Vincent Boulanin, "Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapons Systems" *Stockholm International Peace Research Institute* (November 2015).

29   Kenneth Anderson *et al*, "Adapting the Law of Armed Conflict to Autonomous Weapon Systems" (2014) 90 *International Law Studies* 386 at 401 and 406.

30   Daniele Amoroso & Guglielmo Tamburrini, "The Ethical and Legal Case against Autonomy in Weapons Systems" (2017) 18 *Global Jurist* 1 at 6.

31   See para 2 above.

above hypothetical situation, the human soldiers would ideally stop the attack robot from firing at the two men. However, this article argues that it is not a must to rely on humans to fetter a robot's discretion. If a human operator is allowed to second-guess a robot's decisions, then the robot will be less effective as it would not be able to react immediately to threats. The crux is in educating humans to use these robots in the *correct environment*. Alternatively, states could just delay deploying fully autonomous weapons until they are sure that the robots can reliably distinguish between combatants and non-combatants.

B.   *Unnecessary suffering*

15   The rule against unnecessary suffering can also provide clear guidance for states delving into autonomous weapons deployment. This rule is codified in Art 35(2) of AP I and the ICJ has clarified unnecessary suffering to mean "a harm greater than that avoidable to achieve legitimate military objectives".[32] However, instead of preventing the suffering inflicted on civilians, this rule is meant to restrain the suffering inflicted on opposing combatants.[33]

16   The above hypothetical situation raised State A's Type 2 robots: the nanobots. If those nanobots were deployed into an X Militia command centre, this would be a clear violation of the rule against unnecessary suffering. The excruciating pain that the nanobots can inflict is clearly unnecessary when the mere aim is to incapacitate the leaders of the X Militia. The flying nanobots will likely belong to the same category of banned weapons as glass-filled projectiles, "dum-dum" bullets, poisoned weapons, and bayonets with serrated edges. All these weapons increase suffering without increasing military advantage.[34]

17   Hence, like the principle of distinction, the principle against unnecessary suffering can regulate the use of robots and ensure that warfare will be conducted in a humane manner. Currently, weapons which cause debilitating effects like mustard gas and blinding laser weapons are prohibited under specific treaties: the Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction[35] and the Convention on Prohibitions

---

32   *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion* (1996) ICJ Rep 226 at para 78.

33   Gary D Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (Cambridge University Press, 2010) at p 270.

34   Gary D Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (Cambridge University Press, 2010) at p 270.

35   1975 UNTS 45 (3 September 1992; entry into force 29 April 1997).

or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects,[36] respectively. Since there is no treaty directly regulating robots, this principle against unnecessary suffering is especially important and must be adhered to. However, the rule of proportionality in IHL may not be as clearly applied to robots.

C.   *Proportionality*

18   As codified in Art 51 of AP I, the principle of proportionality prohibits an attack if the incidental civilian harm is excessive in relation to the concrete and direct military advantage anticipated by the attack.[37] Unlike proportionality in self-defence which looks at the impact of the use of force against the target of the attack,[38] proportionality here is concerned with the incidental casualties after the attack.

19   If State A's Type 3 advisory robot had spotted one X Militia member in a remote and isolated desert hiding spot, it could advise the military commander to drop a 2,000-pound bomb on the militia member. It would be a gross waste of munitions, but there would be no proportionality issue because there are no civilians or civilian objects with which to be concerned. To kill an X Militia member in a crowded orphanage with a mere hand grenade could easily be a violation of proportionality, however, because of the close presence of so many civilians.[39] While this illustration seems to suggest that it is possible for robots to adhere to the existing proportionality guidelines under IHL, this article takes a contrary position. It is argued that State A's Type 3 robots will have great difficulty in determining what is a proportionate response in armed conflict. The primary reason is that the rule of proportionality is a complex one and meant to be applied by *humans*.

20   To determine what is considered proportionate, international law has developed a legal test of a "reasonable military commander". This standard has been recognised in multiple publications such as the 2000 International Criminal Tribunal for the Former Yugoslavia ("ICTY")

---

36   1342 UNTS 137 (10 October 1980; entry into force 2 December 1983).
37   Additional Protocol I, Art 51(5)(b); Jean-Marie Henckaerts *et al*, *Customary International Humanitarian Law, Volume I: Rules* (Cambridge University Press, 2009) at p 46 (Rule 14).
38   James Crawford, *Brownlie's Principles of Public International Law* (Oxford University Press, 2019) at p 722.
39   Gary D Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (Cambridge University Press, 2010) at p 277.

Final Report to the Prosecutor,[40] a 2014 International Law Association Study Group report,[41] and a 2016 Human Rights Watch report.[42]

21   In *Prosecutor v Galić*,[43] the ICTY held that:[44]

> In determining whether an attack was proportionate it is necessary to examine whether a reasonably well-informed person in the circumstances of the actual perpetrator, making reasonable use of the information available to him or her, could have expected excessive civilian casualties to result from the attack.

The reference to the "circumstances of the actual perpetrator" alludes to the military commander as he is usually the one responsible for authorising an attack.

22   Moreover, looking at the language of Art 51 of AP I – it refers to "anticipated" military advantage and "expected" injury to civilians. This careful choice of words shows that the analysis must be taken in a prospective manner from the viewpoint of the commander at the time of the attack.[45] Currently, we can analyse if the commander should have expected excessive civilian casualties relative to the military advantage he anticipated gaining, based on what he knew or should have known at the time of the decision to attack. However, it is unlikely that robots will follow this same thought process. Even if robots could think like a military commander, humans would not be able to tell.

23   At present, there are two commonly known forms of AI: symbolic and connectionist. For symbolic AI, the rules are created by humans and then coded into a program.[46] In contrast, for connectionist AI, the programmer does not specify the rules beforehand. The network "discovers the rules from training data"[47] and goes through machine

---

40   United Nations International Criminal Tribunal for the Former Yugoslavia, "Final Report to the Prosecutor by the Committee Established to Review the NATO Bombing Campaign Against the Federal Republic of Yugoslavia" (2000) 39 *International Legal Materials* 1257 at 1271.

41   International Law Association Study Group on the Conduct of Hostilities in the 21st Century, "The Conduct of Hostilities and International Humanitarian Law: Challenges of 21st Century Warfare" (2017) 93 *International Law Studies* 322 at 369.

42   Bonnie Docherty *et al*, "Making the Case: The Dangers of Killer Robots and the Need for a Preemptive Ban" *Human Rights Watch* (9 December 2016).

43   IT-98-29-T, Trial Judgment and Opinion (5 December 2003).

44   *Prosecutor v Galić* IT-98-29-T, Trial Judgment and Opinion (5 December 2003) at para 58.

45   Laurie Blank *et al*, *International Law and Armed Conflict: Fundamental Principles and Contemporary Challenges in the Law of War* (Wolters Kluwer, 2016) at p 204.

46   Pathmind website https://pathmind.com/wiki/symbolic-reasoning (accessed 16 August 2019).

47   Josef Bajada, "Symbolic *vs* Connectionist AI" *Towards Data Science* (8 April 2019).

learning. There are already algorithms which have capitalised on the strengths of both connectionist and symbolic reasoning like AlphaGo, the computer program that plays Go.[48] It is assumed that future military robots will incorporate both symbolic and connectionist AI as well. However, even with both forms of AI, it is unlikely that robots can undertake a proportionality analysis.

24   First, it is very difficult to code the rule of proportionality into robots as every analysis is very context-specific and cannot be simplified into a series of calculations. It is also tough to put a value on "military advantage" and "civilian lives".[49] Calculations may sometimes be possible, for example, when a planned attack will destroy a military target where 100 enemy combatants are present and is expected to cause the death of three civilians. But often, the expected military advantage and collateral damage are dissimilar, for example, in armed conflict, when the expected destruction of vital civilian infrastructure needs to be compared to the destruction of one enemy tank with four enemy soldiers.[50] Then, the autonomous weapon would be conducting a futile exercise of comparing apples and oranges.

25   Secondly, if a robot relies solely on inputted training data, it is highly unlikely that a robot could immediately learn to handle the infinite number of scenarios it might face.[51] Today's neural networks require large volumes of training data to be effective.[52] Since the introduction of the UN Charter,[53] there have been slightly over 60 incidents which are internationally recognised as qualifying as armed attacks.[54] Hence, there is a lack of training data for the military robots to determine what qualifies as a reasonable response in light of the military advantage gained. There is thus a need for the robot to continuously learn from new and real-world data.

---

48   Michael Fu, "AlphaGo and Monte Carlo Tree Search: The Simulation Optimization Perspective" in *Proceedings of the 2016 Winter Simulation Conference* (Theresa M Roeder *et al* eds) (IEEE Press, 2016).

49   United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* UN Doc A/HRC/23/47 (9 April 2013) at para 56.

50   Jeroen Van Den Boogaard, "Proportionality and Autonomous Weapons Systems" (2015) 6 *Journal of International Humanitarian Legal Studies* 247 at 268.

51   Bonnie Docherty, "Losing Humanity: The Case against Killer Robots" *Human Rights Watch* (19 November 2012).

52   Marta Garnelo & Murray Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations" (2019) 29 *Current Opinion in Behavioral Sciences* 17 at 17.

53   1 UNTS XVI (26 June 1945; entry into force 24 October 1945).

54   Tom Ruys *et al*, *The Use of Force in International Law: A Case-Based Approach* (Oxford University Press, 2018).

26   Finally, it is impossible for humans to check the reasoning process of the robots as many machine-learning AI systems operate as "black boxes".[55] The systems can learn from experience by "memorising" successful solutions[56] but even their own designers may be unable to explain what internal process led to a particular output.[57] It is thus impossible to use the principle of proportionality to evaluate State A's Type 3 robot.

27   Given the limitations of proportionality in setting a standard and regulating robots, State A is better off using robots to identify hostile targets and interpret signals that humans cannot, such as radar and sonar signals. Until the day when robots can think like humans,[58] or when humans can understand a robot's thought process,[59] robots will never be able to have regard to proportionality and advise on the means and methods of attack.

28   Alternatively, if robots were used to give advice, a stricter standard should apply to situations where an individual relies and acts upon advice given by a robot. A reasonable military commander is afforded a certain margin of appreciation in making decisions with regard to what constitutes acceptable collateral damage.[60] However, a robot is not subject to the same human fallibility (for example, being too emotional, making calculation errors).[61] This stricter standard also ensures that the individual either must fully trust in the robot's programming or must have done sufficient checks to ensure the proposed attack would not violate the proportionality requirement.

55   Marta Garnelo & Murray Shanahan, "Reconciling Deep Learning with Symbolic Artificial Intelligence: Representing Objects and Relations" (2019) 29 *Current Opinion in Behaviorial Sciences* 17 at 17.

56   Armin Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Routledge, 2009) at p 57.

57   Jacob Turner, *Robot Rules: Regulating Artificial Intelligence* (Palgrave Macmillan, 2018) at p 325.

58   Richard Waters, "Huge Computing Power 'Can Deliver Human-level AI in 5 Years'" *Financial Times* (3 August 2019).

59   Matt Turek, "Explainable Artificial Intelligence (XAI)" *Defense Advanced Research Projects Agency* https://www.darpa.mil/program/explainable-artificial-intelligence (accessed 16 December 2019).

60   Jeroen Van Den Boogaard, "Proportionality and Autonomous Weapons Systems" (2015) 6 *Journal of International Humanitarian Legal Studies* 247 at 278.

61   Kenneth Payne, *Strategy, Evolution, and War: From Apes to Artificial Intelligence* (Georgetown University Press, 2018) at p 187.

D. *Martens Clause*

29   The Martens Clause is found in Additional Protocols I[62] and II[63] to the Geneva Conventions and states that in cases "not covered by the law in force", persons remain under the protection of the "principles of humanity and the dictates of the public conscience". While some have argued that this clause is merely aspirational in nature,[64] the ICJ has recognised this clause as an effective means of addressing the rapid evolution of military technology.[65]

30   Like other commentaries,[66] this article argues that the Martens Clause has a substantive component to it and requires new weapons to adhere to the two guiding principles of humanity and the dictates of public conscience.[67] On the first principle, the ICJ has recognised that one basic consideration of humanity is the upholding of human dignity.[68] Under the "right to life" in international law, individuals are entitled to enjoy the right to "life with dignity"[69] and not to be debased of one's honour and self-respect.[70] Opponents of autonomous weapons have argued that delegating the decision to harm humans to fully autonomous weapons violates this entitlement.[71] As for the second principle, the dictates of public conscience are opinions and values that shape the conduct of the parties to a conflict.[72] Those who have raised ethical concerns over

---

62   Additional Protocol I, Art 1(2).

63   Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) 1125 UNTS 609 (8 June 1977; entry into force 7 December 1978) Preamble.

64   Yoram Dinstein, *Israel Yearbook on Human Rights* vol 49 (Martinus Dijhoff Publishers, 2019) at p 108; Antonio Cassese, "The Martens Clause: Half a Loaf or Simply Pie in the Sky" (2000) 11(1) *European Journal of International Law* 187 at 201–202.

65   *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion* (1996) ICJ Rep 226 at [78].

66   Knut Dormann *et al*, *ICRC Commentary on the First Geneva Convention* (Cambridge University Press, 2016) at p 1187; Theodor Meron, "The Martens Clause, Principles of Humanity, and Dictates of Public Conscience" (2000) 94(1) Am J Int'l L 78 at 82–83.

67   *Legality of the Threat or Use of Nuclear Weapons Advisory Opinion* (1996) ICJ Rep 226 at [74]–[84].

68   *Corfu Channel* (*United Kingdom of Great Britain and Northern Ireland v Albania*) (1949) ICJ Rep 2 at 22.

69   United Nations, Human Rights Committee, *General Comment No 36* UN Doc CCPR/C/GC/36 (3 September 2019) at paras 3, 26 and 62.

70   *Prosecutor v Furundzija*, IT-95-17/1-T, Trial Judgment (5 June 1998) at [183].

71   United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* UN Doc A/HRC/23/47 (9 April 2013) at para 89–97.

72   Knut Dormann *et al*, *ICRC Commentary on the First Geneva Convention* (Cambridge University Press, 2016) at p 1186; Theodor Meron, "The Martens Clause, Principles

allowing fully autonomous weapons to kill without human involvement include a UN Special Rapporteur and the International Committee of the Red Cross ("ICRC").[73] The collective belief is that the sanctity of life is disrespected when fully autonomous weapons are deployed against humans.[74]

31   However, the presence of the Martens Clause does not mean that fully autonomous weapons can never be deployed. As mentioned, the Martens Clause applies to protect persons in cases "not covered by the law in force". Hence, State A should only be considering the principle of humanity and the dictates of public conscience if it finds that the other rules of IHL had failed to regulate fully autonomous weapons. As earlier argued,[75] there is no problem with a robot adhering to the principles of distinction and avoiding unnecessary suffering. State A's Type 1 robots would be lawful if deployed in the correct environment while its Type 2 robots are clearly unlawful. The Martens Clause would thus not need to come into play. It is accepted that based on current AI technology, the Type 3 robots cannot be clearly evaluated by the principle of proportionality. But as AI develops, it is possible that robots will be able to think like humans. Therefore, for the most part, robots can be assessed by the same IHL principles which govern other weapons.

III.   Weapons review obligation in international humanitarian law

32   Thus far, this article has discussed whether fully autonomous robots can be adequately regulated by specific rules in IHL. However, when exactly do states have to check for a weapon's compliance with these IHL rules? The answer is found in Art 36 of AP I:[76]

---

of Humanity, and Dictates of Public Conscience" (2000) 94(1) Am J Int'l L 78 at 82–83.

73   International Committee of the Red Cross, "ICRC views on LAWS, Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?" (3 April 2018)   <https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control> (accessed 27 July 2019); Theodor Meron, "The Martens Clause, Principles of Humanity, and Dictates of Public Conscience" (2000) 94(1) Am J Int'l L 78 at 85.

74   International Committee of the Red Cross, "ICRC Views on LAWS, Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?" (3 April 2018)   <https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control> (accessed 27 July 2019); Theodor Meron, "The Martens Clause, Principles of Humanity, and Dictates of Public Conscience" (2000) 94(1) Am J Int'l L 78 at 85.

75   See paras 10–14 and 15–17 above.

76   Additional Protocol I, Art 36.

> In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.

33   To keep up with constant and rapid evolution of weapons, public international law requires states to conduct legal reviews of new weapons *before their use*.[77] This obligation is breached when the envisaged use of the new weapon breaches public international law binding on the state (including IHL).[78] A failure to carry out the Art 36 weapons review renders a state party internationally responsible for a breach of its obligations *vis-à-vis* the other parties to AP I. However, this article argues that the current weapons review obligation contained in Art 36 of AP I is (a) limited in its effectiveness in ensuring the legality of fully autonomous weapons; and (b) proposes a solution to encourage more states to conduct comprehensive weapons reviews.

A.   *Limitations to weapons review*

34   There are four limitations to the effectiveness of Art 36. First, this legal review is only mandatory for the 174 state parties to AP I.[79] As the extent to which states comply with Art 36 is not clear,[80] commentators are divided over whether the Art 36 weapons review is a rule of international custom and binding on all states.[81] A rule of international custom must

---

77   International Committee of the Red Cross, "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" (2006) 88 *International Review of the Red Cross* 931 at 933.

78   International Committee of the Red Cross, "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" (2006) 88 *International Review of the Red Cross* 931 at 933; Claude Pilloud *et al*, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Martinus Nijhoff Publishers, 1987) at para 1469.

79   Parties to Additional Protocol I of the Geneva Convention https://treaties.un.org/ Pages/showDetails.aspx?objid=08000002800f3586 (accessed 23 July 2019).

80   Marie Jacobsson, "Modern Weaponry and Warfare: The Application of Article 36 of Additional Protocol I by Governments" in *The Law of War in the 21st Century: Weaponry and the Use of Force* (Anthony Helm ed) (Naval War College Press, 2007) ch 9 at pp 184–185.

81   Proponents of international custom: Michael Schmitt & Jeffrey Thurnher, "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict" (2013) 4 Harv Nat'l Sec Journal 231 at 271; Darren Stewart, "New Technology and the Law of Armed Conflict" in *International Law and the Changing Character of War* (Raul Pedrozo & Daria Wollschlaeger eds) (Military Bookshop, 2011) ch 11 at p 283; Duncan Blake & Joseph Imburgia, "'Bloodless Weapons?' The Need to Conduct
(cont'd on the next page)

be supported by widespread, consistent, and representative state practice and *opinio juris*.[82]

35   However, in November 2015, the Stockholm International Peace Research Institute determined that "only a limited number of states are known to have a weapons review mechanism in place".[83] Ten states appear to have formal review processes, the details of which are publicly available: Australia, Belgium, Germany, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the UK, and the US.[84] Interestingly, the US conducts weapons reviews even though it is not a party to the AP I. There are other states that have indicated that they may have informal or formal review processes but have not made sufficient information available to determine whether this is the case. They are Austria, Brazil, Canada, Croatia, Czech Republic, Denmark, Finland, France, Mexico, Poland, Portugal, Russian Federation and South Africa.[85] Therefore, this article agrees that it is difficult to conclude if the obligation in Art 36 of AP I is part of international custom.

36   Secondly, Art 36 fails to encourage a sense of accountability. Article 36 reviews are national procedures beyond any kind of international oversight, and there only exists soft guidelines from the ICRC.[86] Further, "other contracting States have no possibility of verifying

---

Legal Reviews of Certain Capabilities and the Implications of Defining Them as 'Weapons'" (2010) 66 *Air Force Law Review* 157 at 161; International Committee of the Red Cross, "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" (2006) 88 *International Review of the Red Cross* 931 at 933. Opponents of international custom: Natalia Jevglevskaja, "Weapons Review Obligation under Customary International Law" (2018) 94 *International Law Studies* 186 at 220; Chantal Grut, "The Challenge to Autonomous Lethal Robotics to International Humanitarian Law" (2013) 18 *Journal of Conflict & Security Law* 5 at 10.

82   *North Sea Continental Shelf* [1969] ICJ Rep 3 at [77].
83   Vincent Boulanin, "Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapons Systems" *Stockholm International Peace Research Institute* (November 2015).
84   Vincent Boulanin & Maaike Verbru, "SIPRI Compendium on Article 36 Reviews" *Stockholm International Peace Research Institute* (December 2017); Isabelle Daoust *et al*, "New Wars, New Weapons? The Obligation of States to Assess the Legality of Means and Methods of Warfare" (2010) 84 *International Review of the Red Cross* 345 at 359.
85   Brian Rappert *et al*, "The Roles of Civil Society in the Development of Standards around New Weapons and Other Technologies of Warfare" (2012) 94 *International Review of the Red Cross* 765 at 781.
86   International Committee of the Red Cross, "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" (2006) 88 *International Review of the Red Cross* 931.

whether the obligation laid down [in Art 36] is complied with".[87] In accordance with Art 84 of AP I, other state parties to the Protocol may ask to be informed about the *procedure* for legal review. However, there is no requirement for states to publish the *results* of their national reviews and states generally restrict access to them.[88] Without more transparency, national review processes might just become a means to legitimise all sorts of weapons, rather than to filter out unlawful systems.[89]

37   Thirdly, there is a low likelihood of repercussions for non-compliance. In theory, if a weapons review is not carried out, the state will be responsible in any case for any wrongful damage ensuing,[90] and likely asked to give compensation.[91] However, this punishes the state for the wrongful use of force or deployment of the weapon. No state has been punished for failing to carry out a *domestic review* of new weapons. This article argues that if a state is found to possess weapons already deemed unlawful under IHL, the ICJ could order cessation[92] and all the state's research and development effort would be wasted. This threat of cessation would be the only factor motivating states to comply with the weapons review requirement.

38   The final limitation is especially relevant to the use of advanced robots in the military: the lack of resources to conduct testing. Evaluation of performance and the level of risk associated with the use of a weapon must be done in the context of various realistic operational scenarios. The more complex the operating environment, the harder and more expensive it becomes to create a representative and realistic scenario for the purpose of testing and evaluation.[93] For robots built on neural

---

87   Antonio Cassese, "Means of Warfare: The Traditional and the New Law" in *The Human Dimension of International Law: Selected Papers of Antonio Cassese* (Paola Gaeta & Salvatore Zappalà eds) (Oxford University Press, 2008) ch 10 at p 232.

88   Natalia Jevglevskaja, "Weapons Review Obligation under Customary International Law" (2018) 94 *International Law Studies* 186 at 199. Many organisations have appealed to states to be more open about their procedures: see "Call for Increased Transparency in UK Weapons Reviews" *Article 36* (27 February 2012); see also Darren Stewart, "New Technology and the Law of Armed Conflict" (2011) 87 *International Law Studies* 271 at 284.

89   *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)* CCW/CONF.V/2 (10 June 2016) at para 50.

90   Claude Pilloud *et al*, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (Martinus Nijhoff Publishers, 1987) at para 1466.

91   Additional Protocol I, Art 91.

92   *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Art 30.

93   Vincent Boulanin & Maaike Verbruggen, "Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies" *Stockholm International Peace Research Institute* (December 2017).

networks, the procedure will be extremely complex. A neural network of nodes learns by being provided data from which the system will generalise conclusions and develop rules.[94] In a machine learning system, there may be hundreds of millions of nodes and weighted values.[95] There may even be instances of robots working on top of other systems. Consequently, conducting a weapons review of robots will be extremely expensive and states might not always have the resources to fund, or the expertise to conduct, the necessary types of testing and evaluation.[96] This is especially true for states who did not develop the weapon but purchased the technology from a third party.[97]

B.   *Proposed solution*

39   To combat the limitations previously discussed, this article proposes the creation of a committee under the ICRC which currently receives its mandate from the four Geneva Conventions of 1949 and AP I. It is noted that in various states, weapons reviews are often conducted by not only the Ministry of Defence or the armed forces but also a variety of experts from other fields, including medical, technical, and environmental experts.[98] Similarly, here, a multidisciplinary approach should be adopted. The committee should consist of lawyers experienced in IHL, military experts, weapon designers and health professionals. This committee should not be evaluating the new weapons of each country as this is administratively burdensome and is essentially performing a state's obligation for them. Instead, the committee should receive the results of each weapons review and scrutinise them. If this committee is indeed set up, there may be a need for the state parties to AP I to sign an agreement[99] acknowledging that the Art 36 obligation includes a duty to submit testing results to the ICRC. This will then ameliorate the problem of accountability.

---

94   Larry Hardesty, "Explained: Neural Networks" *MIT News* (14 April 2017).
95   James Farrant & Christopher Ford, "Autonomous Weapons and Weapon Reviews: The UK Second International Weapon Review Forum" (2017) 93 Int'l L Stud 389 at 401.
96   Vincent Boulanin & Maaike Verbruggen, "Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies" *Stockholm International Peace Research Institute* (December 2017).
97   Vincent Boulanin, "Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapons Systems" *Stockholm International Peace Research Institute Insights on Peace and Security* (November 2015).
98   International Committee of the Red Cross, "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" (2006) 88 *International Review of the Red Cross* 931 at 950–951.
99   Vienna Convention on the Law of Treaties 1155 UNTS 331 (23 May 1969; entry into force 27 January 1980) Art 31(3)(a).

40   In the initial stages, this committee can rely on the ICRC's guide[100] to implementing Art 36 of AP I and consider if the tests conducted by each state were sufficiently comprehensive and reliable. If the committee is not satisfied with the tests, it can ask the state to conduct additional testing to prove that the new weapon complies with the basic IHL principles. This solution preserves the flexibility in which countries approach their obligations under Art 36. Each state is still able to determine how such a testing process can best be integrated into its own weapons acquisition process, a key factor for the successful implementation of Art 36.[101]

41   This committee could also maintain a database of all weapons-testing results from each state. This will enable states to cut costs when testing new robots. For example, there is no longer a need for two states to conduct their own separate tests when they both plan to buy a fully autonomous weapon from the same manufacturer. The committee can publish a list of approved weapons, means, or methods of warfare on a platform accessible to all states to ensure that there is no repeat testing. Further, the committee can facilitate international co-operation in the field of testing and evaluation for states through the sharing of best practices.[102] This will ultimately bring down testing costs.[103]

42   However, it is noted that many states wish to keep the results of their weapons review confidential, citing national security concerns.[104] The proposed committee should respect this and keep the results of each weapons review confidential unless the specific state opts to make this information accessible to other states. Robots are just like any other new weapon. If this solution is implemented, the weapons review obligation will be effective in preventing states from developing or acquiring robots which fail to adhere to the basic principles of IHL.

---

100 International Committee of the Red Cross, "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" (2006) 88 *International Review of the Red Cross* 931.

101 Justin McClelland, "The Review of Weapons in Accordance with Article 36 of Additional Protocol I" (2003) 85 *International Review of the Red Cross* 397 at 414.

102 *Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)* UN Doc CCW/CONF.V/2 (10 June 2016) at para 51.

103 Vincent Boulanin, "Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapons Systems" *Stockholm International Peace Research Institute Insights on Peace and Security* (November 2015).

104 Michael Meier, "Lethal Autonomous Weapons Systems – Is It the End of the World as We Know It … Or Will We Be Just Fine?" in *Complex Battlespaces: The Law of Armed Conflict and the Dynamics of Modern Warfare* (Winston Williams & Christopher Ford eds) (Oxford University Press, 2018) ch 10 at p 302.

IV.   Accountability for mishaps

A.   *State responsibility*

43   Assuming that a military robot has successfully passed the domestic weapons review, this robot would be deployed by the state. Since these robots are capable of making their own decisions, who would be responsible for any unlawful use of force? This article argues that it is possible for the state to be responsible for the robots' actions.

44   To aid in our analysis, let us consider the following scenarios:

> (a)   A robot with machine learning capabilities chose its own targets and the attack could have been but was not stopped by State A's military personnel.

> (b)   A robot with advanced processing and analytical abilities gave advice to State A's military commander and the commander acted on that advice.

> (c)   A fully autonomous robot was programmed to fire at incoming missiles but misfired once and hit a civilian aircraft.

> (d)   A sudden blizzard caused a military robot to make errors in targeting.

It is assumed that all four scenarios resulted in an unlawful use of force.

45   Under international law, the conduct of any state organ is considered an act of the state.[105] As the armed forces is a state organ,[106] decisions made by the military commanders can be attributed to the state. In the first and second scenarios, State A's military personnel were partially responsible for the unlawful use of force as they either indirectly or directly approved the robot's actions.

46   In the third scenario, though the robot made a mistake and misfired, the state can still be held responsible. In general, autonomous weapons serve an identical function as the armed forces of a state, and actions by the armed forces, are attributable to the state even when this

---

105 *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Art 4.

106 Jean-Marie Henckaerts *et al*, *Customary International Humanitarian Law* Vol I: Rules (Cambridge University Press, 2009) at p 530 (Rule 149(a)); Christopher Greenwood, "State Responsibility and Civil Liability for Environmental Damage Caused by Military Operations" (1996) 69 Int'l L Stud 397 at 399.

entity had acted in excess of authority or contrary to instructions.[107] This pseudo "strict liability" treatment would encourage military leadership to make informed decisions and incorporate additional contextual information which the machine might not be aware of (for example, rumours of a stock-piling of weapons). Human–machine collaboration could employ the strengths of the two types of cognition; with humans adding contextual understanding of the sort that the fast-deciding, pattern-recognising AI cannot.[108]

47   What about the fourth scenario relating to the blizzard? Ostensibly, it seems like the *force majeure* defence can apply to that scenario. Under international custom, this defence requires an irresistible force or an unforeseen event beyond a state's control which will make it materially impossible for that state to perform its obligations.[109] However, under the law of state responsibility, the *force majeure* defence cannot apply to peremptory norms.[110] The illegal use of force is widely recognised as a peremptory norm.[111] The three common justifications for the use of force are self-defence,[112] United Nations Security Council ("UNSC") authorisation,[113] and consent[114] from the other state. Why then can these three justifications be used?

48   The first two justifications can be found in the UN Charter while consent is elaborated on in Art 20 of the International Law Commission's ("ILC") Draft Articles on State Responsibility. As argued

---

107 Carrie McDougall, "Autonomous Weapon Systems and Accountability" (2019) 20 *Melbourne Journal of International Law* 58 at 64; *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Arts 4 and 7.

108 Kenneth Payne, *Strategy, Evolution, and War: From Apes to Artificial Intelligence* (Georgetown University Press, 2018) at p 209.

109 *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Art 23(1).

110 *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Arts 23 and 26.

111 James Crawford, *Brownlie's Principles of Public International Law* (Oxford University Press, 2012) at p 595; Bruno Simma *et al, The Charter of the United Nations: A Commentary* (Oxford University Press, 2012) at p 203; *International Law Commission, Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law – Report of the Study Group of the International Law Commmission* UN Doc A/CN.4/L.702 (18 July 2006) at para 33; *Military and Paramilitary Activities in and against Nicaragua* (1986) ICJ Rep 14 at [188]–[192];

112 United Nations Charter (1 UNTS XVI) (26 June 1945; entry into force 24 October 1945) Art 51.

113 United Nations Charter (1 UNTS XVI) (26 June 1945; entry into force 24 October 1945) Art 42.

114 *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Art 20.

---

by James Crawford, when either of the three justifications are present, the use of force would be "render[ed] … intrinsically lawful",[115] or even an "inherent right".[116] In other words, the use of force will no longer be *illegal*. When a state successfully proves the elements of self-defence or consent, or shows evidence of UNSC authorisation, there can be no breach of public international law. Any other types of use of force will be considered illegal and a breach of a peremptory norm. Consequently, the *force majeure* defence is not available as a justification for the use of force.

49   Even if it is assumed that the state is allowed to rely on *force majeure*, it is likely unable to rely on it to preclude the wrongful use of force by the military robots. Regarding the requirement of "an unforeseen event", the ILC Commentaries stress that "the event must have been neither foreseen nor of an easily foreseeable kind".[117] States cannot rely on this defence if they had assumed the risk of the event occurring, or if proper precautions could have prevented the event from materialising.[118] Through the testing of the military robots, State A would have been able to foresee the possible complications in the robots' performance like how extreme weather conditions may affect a robot's targeting accuracy. Additionally, due to the unpredictability of self-learning robots,[119] State A should know the risks of deploying such robots without a human exercising some control over it. Here, the robot's error in targeting can be treated the same as the malfunction of any other weapon. For example, in May 2018, problems were detected in the US's M4A1 assault rifle:[120]

> The M4A1 was charged and the weapon's selector was positioned between 'semi' and 'auto.' When the weapon's trigger was pulled, the weapon did not fire. When the selector was then moved into either 'semi' or 'auto,' the weapon fired without a trigger pull.

In such cases of malfunction, it is unlikely that states can rely on the *force majeure* defence.

---

115   James Crawford, "Revising the Draft Articles on State Responsibility" (1999) 10 *European Journal of International Law* 435 at 445.

116   James Crawford, "Revising the Draft Articles on State Responsibility" (1999) 10 *European Journal of International Law* 435 at 444.

117   *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Art 23, Commentary, para 2.

118   *Draft Articles on Responsibility of States for Internationally Wrongful Acts with Commentaries* UN Doc A/56/10 (2001) Art 23(2)(b).

119   Ashley Deeks, Noam Lubell & Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States" (2019) 10(1) *Journal of National Security Law & Policy* 1.

120   Todd South, "This Malfunction Has the Army Inspecting Its M4s and M16s" *Army Times* (31 May 2018).

50   However, the consequences of malfunction when it relates to a robot capable of targeting hundreds of military objects at once will be far greater than the malfunction of a mere handgun. Hence, this further reinforces the need for states to conduct multiple tests on robots before they are placed on the battlefield.

B.   *Individual responsibility*

51   From the previous part, it can be seen that there is generally no problem in finding a state responsible when a fully autonomous weapon malfunctions. But what if insufficient testing was carried out by the manufacturer or programmer of a robot before it was sold to a state? Any misfiring here would not be caused by an act or omission by a member of a state's military. It is thus appropriate to discuss the criminal and civil responsibility of the private individual who had produced a faulty weapon.

52   Preliminarily, even with state responsibility, there is still a need for individual responsibility. It has been argued that if military robots malfunction, there is no need for individual responsibility because this would not be any different from accidental killings,[121] for which human soldiers are not held accountable for.[122] However, this reduces the weight and consequence of a loss of human life as the state only must make financial compensation for the wrongful act. Any opportunities for retributive justice and deterrence are also reduced.[123]

53   This part will show how it is difficult to find individuals to be criminally and civilly responsible. Under international criminal law, it is unlikely that the programmer or manufacturer of the robot will be held liable. This is as most international crimes require the individual to intend to commit the crime. For example, for an individual to be prosecuted for committing a Crime of Aggression under the Rome Statute of the International Criminal Court,[124] he must have "planned, prepared, initiated, or executed" an act of aggression. Hence, if there were only mere negligence of the programmer or manufacturer of the robot, this would not amount to a crime under international law. In contrast, if

---

121 Jens Ohlin, *Research Handbook on Remote Warfare* (Edward Elgar, 2017) at pp 361–362.

122 Michael Schmitt & Jeffrey Thurnher, "Out of the Loop: Autonomous Weapon Systems and the Law of Armed Conflict" (2013) 4(2) Harv Nat'l Sec J 231 at 242; *Prosecutor v Delalic* IT-96-21-T, Trial Judgment (16 November 1998) at para 468.

123 United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* UN Doc A/HRC/23/47 (9 April 2013) at paras 60 and 75.

124 2187 UNTS 38544 (17 July 1998; entry into force 1 July 2002) Art 8 bis(1).

the programmer of the robot, a racist, programmed the military robots to target and strike all the homes of race X, this programmer may be successfully prosecuted under international law. However, such clear intent to commit a crime would be extremely rare.

54   On a separate note, in a situation where a state's military commander follows a robot's advice and this resulted in an unlawful use of force, the commander is unlikely to be individually liable for an international crime. This is as the commander likely cannot understand why the robot recommends a certain strategy.[125] For instance, the neural networks of robots will weigh different data components, like the number of enemy troops, number of residential properties, and the distance of the attack from the military base, before proposing a military strategy. The military commander would not know the weight the algorithm gave to these specific input points.[126] It is assumed that due to its algorithm, the robot's "thought process" would be very different from a human commander's. For example, AlphaGo made moves that seasoned observers saw as radically different and of the sort that humans would not conventionally make.[127] It would be unsurprising if a robot comes up with a military strategy and weighs risks in a very different way from humans.

55   Additionally, the very use of machine learning algorithm may hinder a court's analysis of whether the attack was intentional or accidental.[128] As highlighted above,[129] even experts will find it extremely difficult to understand the thought processes of machine learning AI systems. Decisions made by robots are virtually unexplainable and impossible to be understood by a fact-finding court.[130] This is akin to how certain scientific and mathematical theories – for instance, in fields such as quantum physics – are impossible to describe fully in natural language

---

125 Ashley Deeks, Noam Lubell & Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States" (2019) 10(1) *Journal of National Security Law & Policy* 1.

126 See paras 22–26 above.

127 Kenneth Payne, *Strategy, Evolution, and War: From Apes to Artificial Intelligence* (Georgetown University Press, 2018) at p 175.

128 Ashley Deeks, Noam Lubell & Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States" (2019) 10(1) *Journal of National Security Law & Policy* 1.

129 See para 26 above.

130 Christof Heyns, "Autonomous Weapons in Armed Conflict and the Right to a Dignified Life" (2017) 33(1) *South African Journal on Human Rights* 46 at 57; International Committee of the Red Cross, "AI and Machine Learning in Armed Conflict" (6 June 2019) <https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach> (accessed 1 January 2020).

without resorting to numbers and symbols in equations. This difficulty is compounded where machine learning systems update themselves as they operate, through a process of backpropagation and re-weighting their internal nodes so as to arrive at better results each time.[131] As a result, the thought process which led to one result may not be the same as one used subsequently.[132]

56   Then, we are only left with civil liability to deal with issues like negligent programming or testing of the robots. This option would treat any unlawful use of force as a legal accident, and manufacturers would be required to pay for any damage caused and to compensate victims or their families.[133] However, there are a few problems with this product liability approach.

57   First, weapon manufacturers are rarely held accountable for problems in performance after they notify the purchasers of the possible risks when using the weapons.[134] Secondly, as earlier argued,[135] the robots' future actions can be quite unpredictable. This is as the robots' algorithm develops by learning data patterns rather than learning a specific task.[136] It is possible that the system learns from its experience and surroundings in a way that is not predictable by the developer.[137] In a similar vein, systems can be put to work on top of other software systems, which may well be produced by different developers, and unforeseen incompatibilities between them may result in harmful events.[138] In these situations, a programmer cannot be said to be negligent.

---

131 James Farrant & Christopher Ford, "Autonomous Weapons and Weapon Reviews: The UK Second International Weapon Review Forum" (2017) 93 Int'l L Stud 389 at 400–401.

132 Jacob Turner, *Robot Rules: Regulating Artificial Intelligence* (Palgrave Macmillan, 2018) at p 326.

133 Daniel N Hammond, "Autonomous Weapons and the Problem of State Accountability" (2015) 15(2) Chi J Int'l L 652 at 665.

134 Jack Beard, "Autonomous Weapons and Human Responsibilities" (2014) 45 Geo J Int'l L 617 at 647; Robert Sparrow, "Killer Robots" (2007) 24 *Journal of Applied Philosophy* 62 at 65.

135 See para 55 above.

136 International Committee of the Red Cross, "ICRC views on LAWS, Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?" (3 April 2018) <https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control> (accessed 27 July 2019).

137 Daniel N Hammond, "Autonomous Weapons and the Problem of State Accountability" (2015) 15(2) Chi J Int'l L 652 at 667.

138 Daniele Amoroso *et al*, "Who Is to Blame for Autonomous Weapons Systems' Misdoings?" in *Use and Misuse of New Technologies: Contemporary Challenges in International and European Law* (Elena Carpanelli & Nicole Lazzerini eds) (Springer International Publishing, 2019) ch 11 at p 215.

C.   *Proposed solutions*

58   Attribution of responsibility to the state is clear-cut when there is a member of the military involved in the decision to deploy the weapon.[139] However, this article accepts that the position is not as clear when a fully autonomous weapon misfires without any human interference. As argued above,[140] there should be state responsibility though it has not been explicitly recognised in public international law that autonomous weapons serve an identical function as the armed forces of a state.

59   Therefore, to further clarify the law, states can consider including the below provision in any future treaty governing the use of robots:[141]

> Each State Party to the Treaty shall bear international responsibility for all the actions taken by its military, whether such activities were directed by humans or by artificial intelligence, and for assuring that activities carried out in conformity with the provisions set forth in the present Treaty and other principles of international law applicable between the Parties.

60   For individual responsibility, the focus should not be on lowering the threshold for the crime of aggression to require less than "intention" to commit the crime. Instead, the focus should be on ensuring that negligent robot programmers and manufacturers can be suitably dealt with under civil law.

61   First, to ensure that someone will be responsible for any possible malfunction of the robot, responsibility can be assigned in advance to the programmer or manufacturer.[142] Hence, even if the robot developer had already notified the state about the possible risks, this should not protect him from liability. This is similar to how certain exclusion of liability clauses are unenforceable under domestic statutes.[143] However, the purpose of such allocation of responsibility is not to protect the state as a consumer with a weaker bargaining position. Instead, it ensures

---

139 International Committee of the Red Cross, "Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS): Views of the ICRC on Autonomous Weapon Systems" (11 April 2016) <https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system> (accessed 23 July 2019).

140 See para 46 above.

141 Adapted from The Convention on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and other Celestial Bodies 610 UNTS 205 (27 January 1967; entry into force 10 October 1967) Art VI.

142 United Nations General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns* UN Doc A/HRC/23/47 (9 April 2013) at para 81.

143 *Eg*, Unfair Contract Terms Act (Cap 396, 1994 Rev Ed); Unfair Contract Terms Act 1977 (c 50) (UK).

that private individuals involved in the manufacture of military robots conduct proper testing.

62   Secondly, to resolve the problem of unpredictability in the robots, states can limit the parameters within which an algorithm can act.[144] They could direct programmers to prescribe the area that the sensors must search and the time limitations of such a search or set the number of points of similarity that there must be for acceptable recognition of a lawful target.[145] Alternatively, an alarm can be triggered when certain conditions are present.[146] For example, each time a robot targets an area where there is civilian presence, a warning will be given, and a person must manually flip a switch before the robot fires the shot. However, these solutions would greatly limit the robot's learning and flexibility. These key traits help robots identify targets faster than a human possibly could in real-time.[147] Hence, states must effectively balance the risks of a robot's unpredictability and the benefits of speed and accuracy when using AI.

V.   Conclusion

63   With most debate in international law now surrounding the legality of fully autonomous military robots, this article has sought to consider the issue from the other angles of human–robot collaboration and whether the current IHL rules can regulate robots. This article has found that most of the existing public international rules are adequate to deal with the introduction of robots into the military. For the most part, robots can and should be treated like any other weapon.

64   It is accepted that the current state of AI is not advanced enough to ensure that IHL principles such as proportionality will be adhered to by fully autonomous weapons. However, as technology develops, such weapons could be designed to abide by public international law. It is too premature to conclude that all fully autonomous weapons should

---

144 Ashley Deeks, Noam Lubell & Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States" (2019) 10(1) *Journal of National Security Law & Policy* 1.

145 William Boothby, "Dehumanization: Is There a Legal Problem Under Article 36?" in *Dehumanization of Warfare: Legal Implications of New Weapon Technologies* (W Heintschel von Heinegg, Robert Frau & Tassilo Singer eds) (Springer, 2017) ch 3 at p 25.

146 Ashley Deeks, Noam Lubell & Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States" (2019) 10(1) *Journal of National Security Law & Policy* 1.

147 Ronald Arkin, "Lethal Autonomous Systems and the Plight of the Non-combatant" in *The Political Economy of Robots: Prospects for Prosperity and Peace in the Automated 21st Century* (Ryan Kiggins ed) (Springer 2017) ch 15 at p 320.

be banned. Further, the alleged difficulties regarding the attribution of responsibility to states or individuals have been overstated. This article has suggested solutions which do not involve drastic changes like the creation of a new committee under the ICRC, modifying specific private law rules, and setting parameters.

65   Recently, there has been news of humans potentially being able to control robots with their minds.[148] If this technology succeeds, a whole other set of problems will surface. In the meantime, there is no harm in starting to evaluate the current international legal frameworks. As Shakespeare wrote (albeit in a different context), better three hours too soon than a minute late.

———————————

---

148 Julia Carrie Wong, "Elon Musk Unveils Plan to Build Mind-reading Implants: 'The Monkey Is Out of the Bag'" *The Guardian* (17 July 2019).