West, Leah

Working Paper

Moving toward best practices in accountability and military use of AI

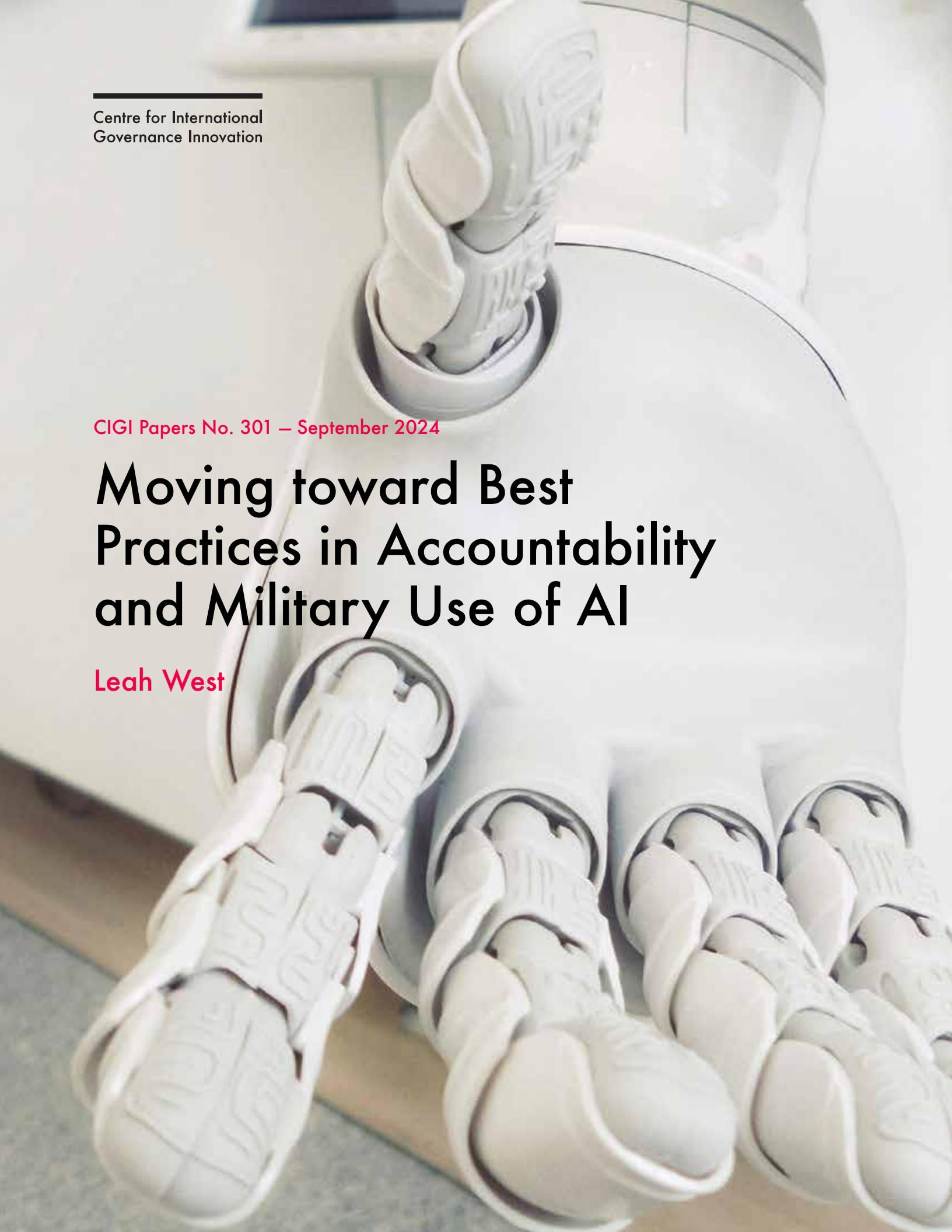
CIGI Papers, No. 301

Provided in Cooperation with:
Centre for International Governance Innovation (CIGI), Waterloo, Ontario

Suggested Citation: West, Leah (2024) : Moving toward best practices in accountability and military use of AI, CIGI Papers, No. 301, Centre for International Governance Innovation (CIGI), Waterloo, ON, Canada

This Version is available at:
https://hdl.handle.net/10419/303158

Centre for International
Governance Innovation



CIGI Papers No. 301 — September 2024

Moving toward Best Practices in Accountability and Military Use of AI

Leah West

Centre for International
Governance Innovation

CIGI Papers No.301 — September 2024

Moving toward Best Practices in Accountability and Military Use of AI

Leah West

About CIGI

The Centre for International Governance Innovation (CIGI) is an independent, non-partisan think tank whose peer-reviewed research and trusted analysis influence policy makers to innovate. Our global network of multidisciplinary researchers and strategic partnerships provide policy solutions for the digital era with one goal: to improve people's lives everywhere. Headquartered in Waterloo, Canada, CIGI has received support from the Government of Canada, the Government of Ontario and founder Jim Balsillie.

À propos du CIGI

Le Centre pour l'innovation dans la gouvernance internationale (CIGI) est un groupe de réflexion indépendant et non partisan dont les recherches évaluées par des pairs et les analyses fiables incitent les décideurs à innover. Grâce à son réseau mondial de chercheurs pluridisciplinaires et de partenariats stratégiques, le CIGI offre des solutions politiques adaptées à l'ère numérique dans le seul but d'améliorer la vie des gens du monde entier. Le CIGI, dont le siège se trouve à Waterloo, au Canada, bénéficie du soutien du gouvernement du Canada, du gouvernement de l'Ontario et de son fondateur, Jim Balsillie.

This research paper has been made possible in part by the Mobilizing Insights in Defence and Security (MINDS) program at the Department of National Defence.

Credits

Managing Director and General Counsel Aaron Shull
Director, Program Management Dianna English
Program Manager and Research Associate Kailee Hilt
Publications Editor Christine Robertson
Senior Publications Editor Jennifer Goyder
Graphic Designer Sepideh Shomali

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

About the Author

Leah West is a CIGI senior fellow, an associate professor at the Norman Paterson School of International Affairs and a leading expert in national security law.

Leah has written extensively on the law as it relates to information and intelligence collection, the privacy implications of using new technology to collect personally identifying information, and intelligence policy. She is frequently consulted by government agencies such as Public Safety Canada, the Communications Security Establishment, and the Canadian Security Intelligence Service on national security law and policy matters, and has testified before the House of Commons, the Senate and the European Parliament. She serves on the editorial board of the *Journal of National Security Law & Policy* and *Terrorism and Political Violence*.

Additionally, Leah is counsel with Friedman Mansour LLP, supporting the firm's criminal, quasi-criminal and administrative law practice. She has appeared before the Ontario Superior Court, the Federal Court of Canada, the Security Intelligence Review Committee and the Supreme Court of Canada. Her academic literature has also been cited by the Supreme Court. She previously served as counsel with the Department of Justice National Security Litigation and Advisory Branch. Before being called to the Ontario Bar in 2016, Leah clerked for the Honourable Justice Mosley of the Federal Court of Canada. Prior to attending law school, Leah served in the Canadian Armed Forces for 10 years as an armoured officer and was deployed to Afghanistan in 2010.

Leah completed her doctorate of juridical science at the University of Toronto. Her dissertation examined the application of constitutional, criminal and international law to online conduct by state intelligence and security agencies, for which she was awarded a Canada Graduate Scholarship to Honour Nelson Mandela. Leah was also the anti-terrorism law fellow at the University of Ottawa Faculty of Law from 2015 to 2017 while she completed her master of laws. She also has a master of arts in intelligence studies from American Military University, a juris doctorate from the University of Toronto Faculty of Law and a B.A. (honours) in politics from the Royal Military College of Canada.

Acronyms and Abbreviations

AI	artificial intelligence
AWS	autonomous weapons systems
IAC	international armed conflict
ICL	international criminal law
ILC	International Law Commission
ICRC	International Committee of the Red Cross
IHL	international humanitarian law
LAWS	lethal autonomous weapons systems
NIAC	non-international armed conflict
REAIM	Responsible Use of Artificial Intelligence in the Military Domain
UN CCW	UN Convention on Certain Conventional Weapons

Executive Summary

In February 2023, the Netherlands hosted the inaugural global summit on the Responsible Use of Artificial Intelligence in the Military Domain (REAIM), culminating in the endorsement of the "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy" by 32 states, which has since grown to 54 endorsing states. This declaration, developed by the United States, is non-binding and aims to foster consensus on norms for the military deployment of artificial intelligence (AI).

This paper supports Canada's leadership of the Accountability Working Group, one of three international working groups formed to elaborate on compliance with the declaration's principles. It delves into the complex legal discourse surrounding accountability for AI actions in armed conflict, particularly focusing on lethal autonomous weapons systems (LAWS) and decision support systems used in targeting.

The first part outlines the Political Declaration's principles on accountability, emphasizing the requirement for commanders to "exercise appropriate care" in deploying AI systems. It suggests that this term captures the need for commanders and operator to make conscious, context-specific decisions about AI systems that are informed by its function, their training on the system, their knowledge of the target and environment and the requirements of international humanitarian law (IHL).

The second part sets out the existing international legal framework that regulates the conduct of hostilities. It begins with a description of the core IHL principles that govern the conduct of military operations and the debate regarding if and how those principles can be upheld when using AI as a decision support tool or as part of a weapons system. This part concludes that the existing literature convincingly argues that the principles of IHL can be adhered to when using AI at least as well as when using other forms of modern technology on the battlefield.

The third part addresses concerns about potential accountability gaps in IHL and international criminal law (ICL) due to AI use. and argues that existing doctrines of command and state responsibility are sufficient to maintain accountability.

Finally, having established that the doctrines of command and state responsibility are crucial to maintaining accountability for AI use by armed forces, the fourth part argues that the real work is defining what military commanders require to rely on and deploy an AI system for which they bear legal liability. The answer is threefold: predictability, training and discipline. To that end, it is recommended that Canada should focus on developing or reconfiguring existing doctrine to meet these requirements when developing and deploying AI.

Introduction

The most significant impact of the February 2023 REAIM summit was the endorsement by 32 states of the "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy" (hereafter Political Declaration) developed by the United States. At the time of writing, the list of endorsing states has grown to 54, and work is under way to build upon the principles through three international working groups.

The statement of principles is not legally binding, nor does it attempt to set out in any detail a legal framework for the use of autonomous systems in the context of armed conflict. Rather, it serves as a road map to building consensus around norms for the use and deployment of AI by armed forces around the world.

Canada and Portugal have stepped up to lead one of the three international working groups made up of endorsing and observer states tasked with further defining what compliance with the principles set out in the declaration entails. This working group is focused on the question of "accountability," a notoriously contentious issue surrounding the use of autonomous systems, especially LAWS in armed conflict (US Department of State 2023). This paper is an attempt to unravel the many layers of the decade-long legal debate regarding the question of accountability for the consequences of actions undertaken by AI in the context of armed conflict, with the aim of supporting Canada's work as the chair of the Accountability Working Group.

This paper relies on the definition of AI as set out in the Political Declaration: "artificial intelligence may be understood to refer to the ability of machines to perform tasks that would otherwise require human intelligence. This could include recognizing patterns, learning from experience, drawing conclusions, making predictions, or generating recommendations. An AI application could guide or change the behavior of an autonomous physical system or perform tasks that remain purely in the digital realm. Autonomy may be understood as a spectrum and to involve a system operating without further human intervention after activation" (US Department of State 2023). Most of the subsequent discussion focuses on the narrow subset of AI systems that raise concerns regarding accountability and legal liability for violations of IHL: LAWS and decision support systems used for targeting. The definition of LAWS is borrowed from a working paper submitted by the United States and Canada as those systems in which the "operator relies on autonomous functions to select and engage targets with lethal force and, before activation, the system operator does not identify a specific target or targets for intended engagement."[1] Decision support systems are "computerized tools that are designed to aid humans in making complex decisions by presenting information that is relevant for the decision or proposing options for the decision maker to choose from in order to achieve a goal" (Roff 2024).

The Declaration's Measures to Ensure Accountability

The Political Declaration is an effort to move the discussion of responsible use of AI in the military domain in a more pragmatic direction, away from the debate over the need to ban LAWS or create a new legal regime governing their use toward a discussion about the development of international consensus and, eventually, norms of best practice around the development, deployment and use of military AI.

The text of the declaration contains an introductory paragraph setting out key principles, a list of 10 foundational measures of responsible behaviour and six commitments that endorsing states agree to undertake to further the objectives of the declaration: "Military use of AI must be in compliance with applicable international law. In particular, use of AI in armed conflict must be in accord with States' obligations under international humanitarian law, including its fundamental principles. Military use of AI capabilities needs to be accountable, including through such use during military operations within a responsible human chain of command and control" (US Department of State 2023).

As leaders of the Accountability Working Group, Canada and Portugal are tasked with leading discussions to develop best practices that operationalize and implement three of the foundational measures most closely linked to a state's ability to maintain a responsible human chain of command and control:

> E. States should ensure that relevant personnel exercise appropriate care in the development, deployment and use of military AI capabilities, including weapon systems incorporating such capabilities.
>
> F. States should ensure that military AI capabilities are developed with methodologies, data sources, design procedures, and documentation that are transparent to and auditable by their relevant defense personnel.
>
> G. States should ensure that personnel who use or approve the use of military AI capabilities are trained so they sufficiently understand the capabilities and limitations of those systems in order to make appropriate context-informed judgments on the use of those systems and to mitigate the risk of automation bias. (ibid.)

Notably, measure E calls on all relevant personnel to "exercise appropriate care" in their use of AI, giving rise to the question: what does appropriate care entail?

The concept of "appropriate care" is not a legal term or standard used in either IHL or ICL. It also differs from a requirement for "meaningful human control," a term that emerged from the 2014 meeting of the UN Convention on Certain Conventional Weapons (UN CCW) and now

1 *Principles and Good Practices on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, 8 August 2022, UN Doc CCW/GG#.1/2022/WP

frequently used to reinforce the need for human accountability in the deployment of LAWS.[2] Over the past decade, there has been much debate within the international community and academic circles on what meaningful human control requires in practice (Roff 2024; Shany 2024). While there is no consensus, scholars have identified common elements in the positions advanced by various parties, namely, that human operators make informed and conscious decisions; have sufficient and accurate information about the functioning of the weapon and the context in which it is deployed; and are well trained on weapons systems that are designed to perform predictably (Crootof 2016; Roff and Moyes 2016; Horowitz and Scharre 2015). Furthermore, many humanitarian organizations and advocacy groups cite the need for a positive action by a human operator in authorizing direct attacks (Human Rights Watch 2016; International Committee of the Red Cross [ICRC] 2021).

The United States has not adopted the "meaningful human control" language, calling the focus on defining and implementing the term "misplaced."[3] Instead the Department of Defense Directive *Autonomy in Weapon Systems*, first published in 2012 and later updated in 2023, requires that "persons who authorize the use of, direct the use of, or operate autonomous and semi-autonomous weapon systems must do so with *appropriate care* and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement" (US Department of Defense 2023, emphasis added). It also stipulates that the design and development of AI, as well as training regarding a system's capabilities, doctrine and tactics, techniques and procedures, allow a commander and operator to "*exercise appropriate levels of human judgment* over the use of force" (ibid., emphasis added).

A working paper submitted by the United States in 2018 to the UN CCW Group of Governmental Experts outlined that "the key issue for human-machine interaction in emerging technologies in the area of LAWS is ensuring that machines help effectuate the intention of commanders and the operators of weapons systems."[4] The critical risk of AI use is that these systems may act in ways unintended or unanticipated by the humans who deploy them. However, so long as machines do what commanders and operators intend for them to do, their use can be compliant with the laws of war.

The paper goes on to explain that "human judgement" is a distinct and broader concept from "human control" because even if an operator exercises complete control over a weapon system, that does not guarantee that they will exercise good judgement when doing so. Thus, human judgement requires "broader human involvement in decisions about how, when, where, and why the weapon will be employed" (Congressional Research Service 2024, 1). Moreover, what amounts to appropriate human judgement is context specific. The standard of appropriateness is flexible and "reflects the fact that there is not a fixed, one-size-fits-all level of human judgment that should be applied to every context. What is 'appropriate' can differ across weapon systems, domains of warfare, types of warfare, operational contexts, and even across different functions in a weapon system."[5] In short, the US position is that weapons systems that make autonomous decisions about targeting may be deployed so long as the commander of that system is capable and has exercised appropriate human judgement when choosing to deploy it.

Heather Roff argues that the requirement for meaningful human judgement actually creates two positive obligations. First, "that humans deploying the systems must understand how they will operate in realistic environments so that humans can make informed decisions regarding their use"(Roff 2016, 3). Second, that "autonomous weapon systems require adequate levels of operational testing, verification, validation and evaluation. This step is required to ensure not only compliance with IHL, but also to provide empirical evidence of system reliability and predictability that informs human decision makers."[6]

Given the long-standing debate about the need for "meaningful human control" versus "appropriate human judgement," it is not surprising that neither term was included in a declaration aimed at building international consensus. That said, the core elements of each standard are set out in measures E through G (while setting

2 *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, 2nd Sess, 28 August 2018, UN Doc CCW/GGE.2/2018/WP.4.

3 *Ibid*.

4 *Ibid*, at para 1.

5 *Ibid*, at para 9.

6 Ibid, at 3.

aside the question of where "in the loop" a human must be when deploying LAWS).

While not definitive, this background provides a good framework for outlining what the standard of "appropriate care" in measure E encompasses: the need for commanders and operators to make conscious, context-specific decisions about an AI system that are informed by its function, their training on the system, knowledge of the target and environment and the requirements of IHL. Moreover, as the subsequent discussion will reveal, this standard mirrors existing expectations of unit commanders under the principles of IHL and the legal and professional doctrines of command responsibility.

The Compatibility of Autonomous Weapons Systems with IHL

IHL, or the law of armed conflict, regulates the conduct of hostilities. This body of law only applies after an armed conflict arises and for the duration of the conflict. Once an armed conflict develops, be it a non-international or international armed conflict (NIAC or IAC), it applies to all states and non-state parties to the conflict. IHL comprises both customary rules and treaties, most significantly the Hague Regulations of 1907 and the four Geneva Conventions and their Additional Protocols. The former sets out the rules for conducting war, while the latter focuses on protecting the victims of war. Only a limited number of these treaty rules apply to NIACs, but the general principles codified in these treaties apply in all conflicts.

Distinction

The core principle of IHL is distinction, which requires that military operations be directed at military objectives. As per article 52(2) of *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol 1)* (hereafter AP 1), military objectives include "objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[7]

This customary principle is codified in articles 48, 51(2) and 52(2) of AP 1. Article 48 stipulates that, "in order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[8]

Put simply, the principle of distinction partitions people into two categories: combatants and non-combatants. Combatants are members of the armed forces party to an armed conflict, except medical and religious personnel. Non-combatants are civilians (unless and for as long as they directly participate in hostilities), persons *hors de combat* and medical and religious military personnel. Combatants may target and kill other combatants without that conduct constituting a war crime. Conversely, civilians may not be targeted, although they do not enjoy absolute protection against being killed.

IHL also requires belligerents to distinguish themselves from the civilian population. Nevertheless, it is not uncommon in modern warfare for members of organized armed groups, especially in NIACs, to not wear uniforms or identify themselves as combatants. Armed groups who do not comply with IHL may also purposely seek to gain tactical advantage by blending in with the civilian population. Additionally, members of armed groups may only support or participate in hostilities intermittently, giving rise to the complicated dilemma of the "baker by day soldier by night" (Forcese and West Sherriff 2017, 168). Civilians are only protected against attack *unless and for such time as they take a direct part in hostilities*. What level of participation and for how long civilians lose their protected status after they put down their rolling pin and pick up a weapon is the subject of detailed guidance by the International Committee of the Red Cross (ICRC) but remains contested by the international community.

7 OHCHR, *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, 8 June 1977 1125 UNTS 3, art 52(2) (entered into force 7 December 1978) [*Protocol to Geneva Conventions*].

8 *Ibid*, art 48.

IHL also prohibits indiscriminate attacks, which means that attacks not specifically directed at a precise military objective are unlawful. The same is true for attacks that employ a method or means of combat that cannot be targeted or whose effects cannot be limited to a military objective or combatants.[9] In other words, any attack must be narrowly tailored to the military objective.

That said, even weapons that are less accurate or incapable of distinguishing between objects could be lawfully deployed if there is no risk of targeting non-combatants or civilian objects. Think, for example, of a naval battle or aviation dog fight where the only objects present in the battlespace are military objectives: the fact that a maritime drone could not distinguish between a naval vessel and a civilian vessel would not prohibit its use in such circumstances. As Michael Schmitt (2012, 2) explains, "The inability of the weapons system to distinguish bears on the legality of their use in particular circumstances (such as along a roadway on which military and civilian traffic travels), but not their lawfulness, *per se*."

Proponents for some form of international ban on the use of LAWS often begin their critique here, insisting that AI is incapable of distinguishing between combatants and non-combatants. The ICRC, for example, effectively makes the argument that war zones are too chaotic and the identification of civilians who are not or no longer taking direct part in hostilities, or a combatant who is *hors de combat*, are too complex for autonomous weapons systems (AWS) to be deployed against persons in a manner that can respect the principle of distinction (ICRC 2021, 9). This, the ICRC states, is because the ways a civilian may take direct part in hostilities or the ways in which a person may surrender or react to being wounded is extremely diverse, making it difficult to standardize for programming purposes. Moreover, "these legal characterizations can change quickly, meaning that an assumption about the targetability of persons within an AWS' area of operation made by a commander upon launching an attack are subject to change before the AWS strikes" (ibid.). Similarly, Human Rights Watch has argued that AWS "do not have the ability to sense or interpret the difference between soldiers and civilians" (Docherty 2012, 2). This is because certain assessments, like whether a civilian is directly participating in hostilities, are, in part, a question of intent, and "fully autonomous weapons would not possess human qualities necessary to assess those intentions" (ibid., 9).

There are several weaknesses in this position. First, as noted above, not every battlespace is populated by civilians or civilian objects. In such circumstances, the identified concerns are irrelevant (Anderson and Waxman 2013; Schmitt 2012).

Second, article 41 (2) of AP 1 clearly stipulates three circumstances in which a person is *hors de combat*: "(a) he is in the power of an adverse Party; (b) he clearly expresses an intention to surrender; or c) he has been rendered unconscious or is otherwise incapacitated by wounds or sickness, and therefore is incapable of defending himself."[10] Arguably, only b) lends itself to considerable ambiguity, but on this point the ICRC commentary explains that "in general, a soldier who wishes to indicate that he is no longer capable of engaging in combat, or that he intends to cease combat, lays down his arms and raises his hands. Another way is to cease fire, wave a white flag and emerge from a shelter with hands raised, whether the soldiers concerned are the crew of a tank, the garrison of a fort, or camouflaged combatants in the field. If he is surprised, a combatant can raise his arms to indicate that he is surrendering, even though he may still be carrying weapons" (Sandoz, Swinarski and Zimmermann 1987, 487). So long as a weapon system is capable of identifying this kind of activity in a dynamic battlefield environment, it could be deployed consistently with IHL. Literature suggests that this is not yet the case although efforts to develop AI technology with this capability are under way (Winter 2022; Seixas-Nunes 2022).

Third, identifying whether a civilian is directly participating in hostilities is often a highly contextual and challenging assessment for even the most experienced soldier, commander and legal officer (Heller 2023). Fortunately, there is a default position: when in doubt, the law requires that a civilian be treated as a non-combatant (Melzer 2009). The same is true for objects. In cases of doubt about whether an object is making an effective contribution to military action, the presumption must be made that it is not being used

9 *Ibid*, art 51(4).

10 *Ibid*, art 41(2).

in this manner (AP 1, art. 52(3)). Assessing doubt is a matter of weight or probability, which is how most AI systems render decisions. To limit the risk of miscalculating that a civilian is targetable, AI systems could be programmed to weigh their targeting decisions heavily in favour of a finding of non-combatant. Alternatively, AI systems could be programmed to only target civilians engaged in specific or obviously belligerent activities where intent can be inferred directly from their actions (Schmitt 2012). Such actions could include firing on the enemy, planting an improvised explosive device or driving an ammunition truck toward the front lines (Melzer 2009). Systems could also be programmed to not recognize certain objects or people as targets, or to require human approval before engaging them. For example, AI could be programmed to never recommend or engage children as targets, even though they may be lawfully engaged if they are directly participating in hostilities. Similar limitations could be imposed on certain objects such as hospitals, ambulances and religious buildings.

Finally, regardless of the weapon system used, be it a drone, artillery, grenade or personal rifle, there is always a risk that the targetability of a person will change after they are engaged. What matters from a legal standpoint is whether a person was a lawful target when the strike was launched or the trigger was pulled, based on the information available to the commander or operator at the time (Seixas-Nunes 2022). If so, the attack is consistent with the principle of distinction. The same is true if the target is selected or engaged by an autonomous system.

All this is to say that there is nothing inherent in the use of AI that makes it incapable of being deployed consistently with the principle of distinction. In the future, AI systems may prove to be far superior at battlespace recognition than even the best trained analysts (Margulies 2019; Winter 2022). Nevertheless, as we discussed in the third part of the paper, the commander choosing to rely on an AI decision support system or deploying a LAWS is ultimately responsible for complying with the principle of distinction, and not the AI system itself. Thus, before relying on AI to make targeting decisions, a military commander must be armed with the following information:

- 1. to what extent is the battlefield populated by objects or persons whose status as a valid military objective is in doubt;
- 2. to what extent is the AI system capable of distinguishing between combatants and non-combatants and military or civilian objectives;
- 3. how often is the data relied upon to make targeting decisions updated, and how does the system weigh new versus older information;
- 4. to what extent is the system weighted or programmed to error on the side of not engaging ambiguous targets; and
- 5. whether that system has or can be programmed to refrain from engaging lawful targets to comply with the rules of engagement (such as not targeting children, schools, hospitals, cultural sites, and so on, even where they are valid military objectives without additional authority).

While criteria 4 and 5 are not necessarily required to comply with the principle of distinction, understanding how the system will respond to identified targets will help ensure the commander deploys the weapon system with consistency in terms of the overall mission.

Proportionality and Precaution

The second fundamental principle in IHL that critics suggest is at odds with the use of LAWS and AI decision support systems is proportionality. Article 51(5)(b) of AP 1 codifies this customary principle by prohibiting "an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[11]

This rule of proportionality "establishes a link between the concepts of military necessity and humanity" (Office of the Judge Advocate General 2001, 2–2). It affirms that civilians and civilian objects may not be targeted in an armed conflict but accepts that they may be injured or killed as collateral damage when engaging a target whose destruction or neutralization presents a proportionate military advantage.

Critics of AWS argue that determining the concrete and direct military advantage of an attack and then measuring that advantage against the anticipated

11 *Ibid*, art 51(5)(b).

collateral damage is so multifaceted and contextual that it is a decision only humans are capable of making (Docherty 2015; ICRC 2021). While it is certainly true that some targeting decisions will present complex questions regarding the proportionality of an attack, this is certainly not the case in all, and arguably not in most, decisions regarding the use of force in armed conflict.

Questions of proportionality only present themselves when a strike may result in collateral damage to civilians and civilian objects. Some battlespaces, such as those at sea, in the air, in the desert, in the Arctic or in cyberspace, may present little to no chance of collateral damage. Moreover, the destruction of certain targets — such as military bases, command centres, naval vessels, artillery positions, air defence systems, a column of advancing tanks, and so on — may present such an overwhelming military advantage that, depending on the means and method of attack, there is very little doubt about the proportionality of a strike. In more complex environments, AI systems could be programmed so as not to select or engage a target if the collateral damage assessment reached a certain threshold or unless a proportionality assessment calculated a minimum military advantage (Schmitt 2012).

Thus, as with the question of distinction, there is nothing inherent about AWS that makes their deployment incompatible with the principle of proportionality in every instance. In some cases, the data analytics capacity of an autonomous system may enhance a commander's capacity to assess the potential military and collateral damage; indeed, many states already use computer programs to conduct collateral damage assessments.

This leads to the precautionary principle, another customary rule of IHL codified in article 57 of AP 1, which sets out a number of obligations for targeting. In particular, this principle requires that "in the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects" and with respect to attacks, "those who plan or decide upon an attack *shall do everything feasible* to verify that the objectives to be attacked are neither civilians nor civilian objects."[12] Additionally, those who plan or decide upon an attack must take "all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects."

According to the treaty commentary on article 57, before states adopted the provision, the phrase "everything feasible" was discussed at length (Sandoz, Swinarski and Zimmermann 1987). Some delegations, including the United Kingdom, understood the words to mean "everything that was practicable or practically possible, taking into account all the circumstances at the time of the attack, including those relevant to the success of military operations" (ibid., 680). The commentary suggests this last requirement is too broad as it could give the success of the mission precedence over humanitarian obligations. Instead, what is required is that necessary identifications must be carried out in a timely manner to spare the civilian population to the furthest extent possible.

As Jean-François Quéguiner (2006, 809–10) explains, "when taking precautions in attack, armed forces cannot be required to do the objectively impossible, nor can they be content with merely doing what is possible." Compliance with the precautionary principle is, therefore, largely reliant on the collection and analysis of information about potential targets, which is dependent on the capabilities and technical resources of a party to the conflict (ibid., 797).

Consequently, where a state has access to decision support systems that can improve its ability to identify the nature of a target, the military advantage presented by a target's destruction or the anticipated collateral damage, the precautionary principle requires states to leverage that technology where feasible. The same would be true if a LAWS proved to be better at making these assessments than a human operator. Conversely, in any circumstance where the use of an AI system is less capable of making these assessments than the available analysts, operator or commander, the law prohibits their use. As Schmitt explains (2012, 20–21), "the only situation in which an autonomous weapon system can lawfully be employed is when its use will realize military objectives that cannot be attained by other available systems that would cause less collateral damage."

Ultimately, military commanders must base their decisions to leverage an AI system in any given situation on their assessment of the complexity of the battlespace; the capabilities and limitations

12 *Ibid*, art 57 (emphasis added).

of that particular system; and their obligations under IHL. If the AI system in question is unable to distinguish combatants from non-combatants or military objectives from civilian ones, or to calculate collateral damage and assess military advantage at least as well as the men and women in their chain of command in that situation, they may not rely on this system. Should a commander nevertheless choose to deploy an AI system in such circumstances, they are liable for violating the core principles of IHL, regardless of whether it results in civilian casualties.

Accountability for IHL Violations

The concern over the ability to hold states liable for violations of IHL resulting from their use of AI in armed conflict is another long-standing concern among critics. The section of the paper that follows argues that the concerns over an AI "accountability gap" are overblown. Together, the doctrine of command responsibility and the law of state responsibility provide the means of holding those involved in the design, development and deployment of AI systems implicated in violations of IHL accountable at least as well as those responsible for violations resulting from the use of traditional weapon systems; the prospect of accountability is not dependent on where a human stands in the "loop." Thus, the focus of the Accountability Working Group should not be on the question of whether states or their armed forces can be held accountable or the level of human involvement required in all cases of AI deployment, but instead on what policies and procedures states should put in place to ensure compliance with IHL and what training and information military commanders need to feel comfortable relying on or deploying a weapon system for whose actions they are liable.

Command Responsibility

Under the legal doctrine of command (or superior) responsibility, a military commander is responsible for not only their acts and omissions (direct command responsibility), but the actions of their subordinates (indirect command responsibility) (Gunawan et al. 2022). This customary law obligation creates an affirmative duty for all commanders to ensure that everyone under their command complies with IHL during armed conflict (AP 1, art. 86; International Criminal Court 2021, art. 28). Moreover, commanders are duty-bound to investigate allegations and punish violations of IHL perpetrated by their subordinates; failure to do so is itself a violation of international law and may result in criminal proceedings against a commander. The ICRC articulates this customary rule as follows: "Commanders and other superiors are criminally responsible for war crimes committed by their subordinates if they knew, or had reason to know, that the subordinates were about to commit or were committing such crimes and did not take all necessary and reasonable measures in their power to prevent their commission, or if such crimes had been committed, to punish the persons responsible" (Henckaerts and Doswald-Beck 2005, 153). This legal doctrine is linked to but differs from the broader military concept of responsible command or commander accountability. The professional doctrine dictates that a unit commander (regardless of the size of the unit) is responsible for the effectiveness of their unit in executing its mission (Corn 2014). This responsibility is all-encompassing, including all aspects of unit training, organization, direction and coordination to accomplish its mission as well as the health, welfare, morale and discipline of the unit's members (Kraska 2021).

As Geoffrey Corn explains (2014, 904), these two concepts are technically distinct:

> The IHL notion of responsible command is … inextricably linked to these responsibilities. This is because IHL is unquestionably and intuitively premised on the expectation that the proper exercise of command responsibility is essential to enhancing the probability of IHL compliance in the most physically and morally challenging martial situations. Thus, "responsible command" in the IHL sense does not connote a distinct command function, such as the responsibility to train soldiers, or provide clear and effective orders, or ensure equipment is properly maintained, or manage the expenditure of finite unit resources. Instead, the IHL notion of "responsible command" inherently connotes an expectation that *all* command responsibilities will be conceived and

> executed in a manner that advances the core objectives of IHL. Preparing a military unit to execute its combat function within the bounds of IHL is therefore an inherent expectation of responsible command, and as such, IHL permeates the entire concept of command and every function performed in the execution of command responsibilities.

Nothing about the use or deployment of AWS changes the legal or professional responsibilities of unit commanders. Their obligation to exercise command responsibility persists regardless of how proximate or remote they are from the weapon system that engages a target. However, critics argue that due to the complexity and nature of LAWS, commanders will not be in a position to meet these obligations. As such, there is a gap in a state's ability to hold individuals accountable for violations of IHL resulting from the use of AI. The only resolution, some suggest, is ensuring that a human pulls the trigger or pushes the button that deploys a weapon system (Roff 2024).

Underlying this argument is the presumption that in all cases of armed conflict where there are civilian casualties, there exists the possibility of criminal liability for someone in the chain of command. This is simply not true. Rather, the crucial question is whether the use of AI reduces our existing ability to hold states and commanders liable for violations of IHL.

Consider a situation where a programmer, operator or commander intends to use a LAWS to commit a war crime or deploys a LAWS knowing that it is likely to or could result in the commission of a war crime (Dunlap 2016). In such an instance, the fact that the weapon used to commit the war crime relied on AI in no way alters the human actor's liability. At the very least, this actor will be criminally liable if they do not take reasonable measures to prevent the crime or investigate and punish those responsible (ibid.).

This scenario leads critics to suggest that it may not be possible for an investigation to identify who is responsible for an IHL violation given the number of people involved in the process of bringing a LAWS online, and then programming and deploying it (Human Rights Watch 2016; Amoroso and Tamburrini 2019). Alternatively, it may be challenging to prove that a programmer, operator or commander intended, knew or should have known that the weapon's use would result in a war crime, making it impossible to establish the mental element required for criminal liability under international criminal law. These problems, however, are not unique to LAWS. There are many hands involved in the development, production and deployment of any weapons system.

The central questions from an accountability standpoint are whether the commander: reasonably understood how the weapon operates; authorized its deployment in a manner that was consistent with both its intended use and IHL; and met their legal obligations as a commander if the weapon malfunctioned, acted unpredictably or if a subordinate operated the AWS in an unauthorized manner. In other words, did the commander exercise appropriate care?

If the answer to any of these questions is no, and a war crime ensues, the commander will be either directly or indirectly liable for that crime. If the answer to any of the questions is yes, and an investigation initiated as required reveals that the conduct or crime was the result of negligence, recklessness or intent on the part of a manufacturer, validator, programmer, maintainer or operator, there are means for holding each accountable (Dunlap 2016).

Only in the rare instances in which an investigation would be unable to identify any punishable wrongdoing by a human leading to a war crime perpetrated by an AWS could there be an accountability gap (Heller 2023). Here, it is important to distinguish between a malfunction and what Afonso Seixas-Nunes classifies as an error (Seixas-Nunes 2022). Malfunctions are the result of an unforeseeable hardware failure that cannot be attributed to a human fault. Any weapon system is capable of malfunctioning and could conceivably result in civilian casualties or damage to civilian infrastructure. There is no reason to believe LAWS are any more prone to hardware malfunctions than similar semi-autonomous or manual weapon systems. However tragic, in the absence of human fault, a malfunction is not a war crime (International Criminal Court 2021, art. 28, art. 30); no accountability gap exists.

An error is caused by an unforeseeable software failure, such as "the missile is not able to adapt to the circumstances of the battlefield, so hits the wrong target, the munition engages the target accurately but causes more damage than

initially foreseen, [or] the algorithm for situation management does not function as planned" (Seixas-Nunes 2022, 210). While errors are possible for any weapon system that relies on software, the risk of an error for a weapon system that leverages AI is greater. This increased risk, roboticists warn, stems from the possible strength of AI, which could overturn a LAWS programmed mission or constraints, leading to truly unforeseeable and unlawful outcomes. As with a malfunction, there is no human fault when an error arises, meaning no criminal liability is attached (ibid.).

Fortunately, while there will be some circumstances in which no individual is criminally liable for violations of international law, states may be held accountable under the law of state responsibility.

State Responsibility

The International Law Commission's (ILC) Draft Articles, published in 2001, codified the law on state responsibility and are widely understood to reflect customary international law (ILC 2013). The underlying principle of state responsibility is set out in article 1: "every international wrongful act of a state entails the responsibility of that state" (ibid.). An international wrongful act exists when an act or omission is "(a) attributable to the state under international law; and (b) constitutes a breach of an international obligation of the State" (ibid.). Per article 2, if an injured state invokes responsibility, absent a valid justification of self-defence, consent, distress, necessity, countermeasures or force majeure (ibid., art. 20–25), the responsible state is obligated to cease its wrongful conduct and make reparations (ibid., art. 30–31).

The law of state responsibility is a set of secondary rules, the primary rules being the obligations a state commits itself to uphold by treaty or by virtue of its statehood under customary international law. In the context of a state's use of AWS during an armed conflict, the primary rules are a state party's treaty and customary IHL obligations, which, in accordance with Common Article 1 of the Geneva Conventions, they must respect and ensure respect for "in all circumstances."[13] The first question in the law of state responsibility is always whether the conduct in question is attributable to a state. The starting point for attribution is that "the conduct of any state organ shall be considered an act of that state" (ILC 2013, art. 2). This rule applies to entities or officials belonging to the executive, legislative or judicial branches of government operating at the federal, provincial or municipal levels. A state's armed forces is one of the most obvious organs of the executive branch, and its actions are, as James Crawford (2013, 216) explains, "in the context of armed conflict…in all cases attributable to and engage the international responsibility of the state in question." The resulting implication is that "all acts of designing, programming, maintaining and activating/deactivating and AW [autonomous weapons] are attributable…to the deploying state" (Seixas-Nunes 2022, 249). This fact remains true regardless of whether or not the organ exceeds its authority or contravenes instructions (ILC 2013, art. 7).

The decision to deploy an AWS that subsequently acts unpredictably does not vitiate the attribution of the resulting conduct to the state. Given the nature of autonomous systems, there is always some level of risk that it will perform unpredictably. "The nature of AWS, being systems whose operations will be based on probabilities and not on pre-given deterministic data, demands an implicit acceptance of the inherent risk that devolves upon the deploying state" (Seixas-Nunes 2022, 250). In other words, the risk of error is inherent in LAWS, and unlike a malfunction, a violation of IHL may be the result of a deliberate action or actions undertaken by the system itself.

Thus, if states want to benefit from the use of LAWS, they must accept responsibility if their use results in indiscriminate or disproportionate attacks on civilians or civilian objects. Per articles 30 and 31 of the ILC's Draft Articles, those consequences include the cessation of the system's deployment, appropriate guarantees of non-repetition and full reparations for the injury caused. As Seixas-Nunes (2022, 250) writes, "The challenge that 'foreseeability' presents is that it requires the state to accept responsibility for a less direct or proximate act than it has had to do before."

13 OHCHR, *Geneva Convention Relative to the Protection of Civilian Persons in Time of War*, 12 August 1949, 75 UNTS 287 (entered into force 21 October 1950).

Predictability, Training and Discipline

As Geoffrey Corn (2014) has explained, a commander's professional responsibilities are inextricably linked to their responsibilities under IHL. Before relying on AI to make targeting decisions or deploying a LAWS, they must be confident that in relying on that system they are upholding the principles of distinction, proportionality and precaution; if not, they will be liable for any resulting war crimes.

Responsible commanders will be hesitant to deploy AI systems in battlespaces where there is a risk of civilian casualties or damage to civilian objects, especially those systems that operate fully autonomously. Yet, in some cases, relying on AI rather than humans to make assessments and targeting decisions may be the responsible thing to do. The key then is to identify what a responsible commander requires so that they, their superiors and their subordinates (who may also be liable for resulting IHL violations or whose very lives might depend on a LAWS) are confident relying on AI in armed conflict.

This question is highly contextual. The specific requirements will vary considerably depending on the type of AI system or weapon, the battlespace, the nature of the conflict and the capabilities of the adversary. That said, the Accountability Working Group can begin to establish baseline best practices to ensure that the responsible military commander is confident that all relevant personnel have and will exercise appropriate care in the development, deployment and use of military AI. It bears repeating that where a commander lacks this confidence, the law prohibits the deployment of LAWS or reliance on AI decision support systems.

The starting point for LAWS and AI decision support systems is the same as any other surveillance or intelligence tool or weapon system: predictability, training and discipline.

Predictability

Predictability refers both to the weapons system itself and to how it will be used. As explained in the third part of the paper, a LAWS powered by machine learning will never be entirely predictable. However, through training, testing, validation, programming constraints and policy limitations on use, a LAWS may behave predictably enough for a responsible commander and their subordinates to feel confident that its authorized use will not violate IHL (United Nations Office for Disarmament Affairs 2017). The same is true for decision support systems.

Of course, there is an important distinction to made between conventional weapons or decision support tools versus a system that continuously learns. The goal for the latter, in most instances, will be for the system to become more predictable over time, but this will not always be the case. This reality leads Peter Margulies (2017) to advocate for what he calls "dynamic diligence." He suggests that a unit that deploys LAWS must have the technical capabilities to continuously monitor its actions, if not its decision making, adapt the system to meet the changing realities of the operating environment and override or shut down the system at any time. To ensure effective oversight and to maintain situational awareness over the deployment of LAWS, the use of these systems must be limited in terms of time, scale and space. An associated task is ensuring that the data relied upon by the AI system is up to date and not corrupted in any way.

Units, explains Margulies, should also have the technical and procedural capacity to conduct after-action reviews of a LAWS performance, and operators should perform regular tests of the system to confirm compliance with set parameters. For these reviews and tests to be meaningful, AI target identification or "nomination decisions" must be transparent and interpretable.

Through these processes and procedures, commanders can build and maintain the confidence that they can predictably deploy an AI system consistently with IHL. Still, building up the technical capacity and competence for commander and their units to exercise dynamic diligence in an armed conflict requires training. Similarly, building confidence that a unit and its members will deploy an AI system predictably requires both training and discipline.

Training

Training "is the essential component in preparing soldiers and military units for success in battle and other military operations"(Corn 2014, 914). Military training is a complex process. It is not enough for a soldier to have a high level of competency in a weapon or system, or for a commander to

acquire an understanding of a weapon or system's capacity and limitations. They must then learn how to leverage that competency and capacity within their small unit through simulations that demand problem solving under various constraints and stress, and finally integrate those skills into collective battlefield simulations that mimic the operating environment. Corn writes that "all of this contributes to developing almost instinctual reactions and responses to the myriad of situations that will confront the individual and collective military assets of an armed force," including "those related to IHL itself" (ibid.). As such, training compliance with IHL is necessary at every phase of training: "only by training compliance with this law will soldiers be genuinely prepared for their inevitable moments of decision" (ibid.).

In preparing to deploy and operate a LAWS, commanders must integrate some unique knowledge and skill sets into their training. In particular, those who deploy or rely on AI systems need to understand and be trained to counter automation bias (Kwik and Van Engers 2021). According to Nema Milaninia (2020, 215), "automation bias refers to the human tendency to favour results generated by automated or computer systems over those generated by non-automated systems, irrespective of the error rates of each." This bias poses a serious problem from an IHL perspective because it may lead a commander or operator to believe an AI's nomination decisions, proportionality assessment or collateral damage calculation is more likely to be correct and not intervene when it appears to have made an error.

There is also a second element of training that is somewhat unique to LAWS and AI decision support, and that is the need to train the system itself. Here again, training must replicate the type of data, operational conditions and physical environment that the system will encounter in the battlespace.

Discipline

One final element that is crucial to ensuring AI systems are deployed predictably by a unit, as well as consistently with a soldier's training and the requirements of IHL, is discipline. Military discipline is not about blindly following orders. Rather, as Corn (2014, 904) explains, meaningful military discipline produces two outcomes: "an unhesitating willingness to subject themselves to mortal danger in order to execute superior orders" and the full commitment "to respecting the IHL-based legal limitations on their power." In short, a disciplined solider would be willing to die, and a commander would be prepared to lose the soldiers under their command, before deploying a LAWS in violation of IHL. Equally, a disciplined solider will not comply with an unlawful order or failure to intervene where they believe the use of a LAWS by another member of their unit may result in a war crime.

Building unit discipline is dependent on predictability and training. Failure to make decisions consistently with IHL in training must be addressed in a timely and predictable manner, and punishment for such failures must be routine and transparent. This means that units must have policies and procedures in place to report, investigate and punish incidents of AI use that violates not only IHL, but the rules of engagement, as well as state or unit obligations and limitations regarding their use.

Conclusion

The Political Declaration and Canada and Portugal's leadership of the Accountability Working Group present an opportunity to move past debate about whether LAWS and AI decision support tools can be deployed consistently with IHL toward building consensus about how best to exercise appropriate care to ensure their lawful use. This paper has made the argument that there is nothing inherent about the deployment of AI systems in armed conflict that is irreconcilable with the principles of distinction, proportionality and precaution. There are various means of assuring LAWS are deployed consistently within these principles, and where the technology or conditions are such that compliance cannot be assured, IHL prohibits their use. There is no gap in the law governing LAWS just as there is no accountability gap should the autonomous use of AI result in a war crime or violate a state's obligations under IHL.

Where gaps are most likely to exist is between responsible commanders and their willingness to rely on LAWS and AI decision support systems to make decisions for which they or their troops may be criminally liable. Closing that gap requires predictability, training and discipline. Developing or retooling best practices from existing doctrine

to ensure these elements are met, including policies and procedures for training, testing, validating, monitoring, reviewing and correcting AI systems and those who operate them in complex military environments, should be the focus of Canada and Portugal's leadership.

Works Cited

Amoroso, Daniele and Gugliemo Tamburrini. 2019. "What makes human control over weapons 'meaningful'?" International Committee for Robot Arms Control Report to the CCW GGE. August. www.icrac.net/wp-content/uploads/2019/08/Amoroso-Tamburrini_Human-Control_ICRAC-WP4.pdf.

Anderson, Kenneth and Matthew C. Waxman. 2013. "Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can." Stanford University, The Hoover Institution (Jean Perkins Task Force on National Security and Law Essay Series), American University, WCL Research Paper 2013-11, Columbia Public Law Research Paper 13-351. https://doi.org/10.2139/ssrn.2250126.

Congressional Research Service. 2024. *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems.* https://crsreports.congress.gov/product/pdf/IF/IF11150.

Corn, Geoffrey S. 2014. "Contemplating the true nature of the notion of 'responsibility' in responsible command." *International Review of the Red Cross* 96 (895–96): 901–17. https://international-review.icrc.org/sites/default/files/irrc-895_896-corn.pdf.

Crawford, James. 2013. *State Responsibility: The General Part.* Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139033060.

Crootof, Rebecca. 2016. "A Meaningful Floor for 'Meaningful Human Control.'" *Temple International & Comparative Law Journal* 30 (1): 53–62. https://sites.temple.edu/ticlj/files/2017/02/30.1.Crootof-TICLJ.pdf.

Docherty, Bonnie. 2012. "Losing Humanity: The Case against Killer Robots." Human Rights Watch, November 19. www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

———. 2015. "Mind the Gap: The Lack of Accountability for Killer Robots." Human Rights Watch, April 9. www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots.

Dunlap, Charles J., Jr. 2016. "Accountability and Autonomous Weapons: Much Ado about Nothing?" *Temple International & Comparative Law Journal* 30 (1): 63–76. https://sites.temple.edu/ticlj/files/2017/02/30.1.Dunlap-TICLJ.pdf.

Forcese, Craig and Leah West Sherriff. 2017. "Killing Citizens: Core Legal Dilemmas in the Targeted Killing Abroad of Canadian Foreign Fighters." *Canadian Yearbook of International Law* 54 (October): 134–87. https://doi.org/10.1017/cyl.2017.13.

Gunawan, Yordan, Muhamad Haris Aulawi, Rizaldy Anggriawan and Tri Anggoro Putro. 2022. "Command responsibility of autonomous weapons under international humanitarian law." *Cogent Social Sciences* 8 (1): 2139906. https://doi.org/10.1080/23311886.2022.2139906.

Heller, Kevin Jon. 2023. "The Concept of 'The Human' in the Critique of Autonomous Weapons." *Harvard Law School National Security Journal* 15. https://harvardnsj.org/2023/12/15/the-concept-of-the-human-in-the-critique-of-autonomous-weapons/.

Henckaerts, Jean-Marie and Louise Doswald-Beck. 2005. *Customary International Humanitarian Law — Volume I: Rules.* Cambridge, UK: Cambridge University Press.

Horowitz, Michael and Paul Scharre. 2015. "Defining 'Meaningful Human Control' Over Autonomous Weapons." Just Security, March 19. www.justsecurity.org/21244/defining-meaningful-human-control-autonmous-weapon-systems/.

Human Rights Watch. 2016. "Killer Robots and the Concept of Meaningful Human Control: Memorandum to Convention on Conventional Weapons Delegates." April 11. www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control.

ICRC. 2021. "ICRC Position on Autonomous Weapon Systems." May 12. Geneva, Switzerland: ICRC. www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

ILC. 2013. *Materials on the Responsibility of States for Internationally Wrongful Acts.* UN Legislative Series. New York, NY: United Nations. https://doi.org/10.18356/1b3062be-en.

International Criminal Court. 2021. *Rome Statute of the International Criminal Court.* The Hague, Netherlands: International Criminal Court. www.icc-cpi.int/sites/default/files/2024-05/Rome-Statute-eng.pdf.

Kraska, James. 2021. "Command Accountability for AI Weapon Systems in the Law of Armed Conflict." *International Law Studies* 97 (1): 408–45. https://digital-commons.usnwc.edu/ils/vol97/iss1/22/.

Kwik, Jonathan and Tom Van Engers. 2021. "Algorithmic fog of war: When lack of transparency violates the law of armed conflict." *Journal of Future Robot Life* 2 (1–2): 43–66. https://doi.org/10.3233/FRL-200019.

Margulies, Peter. 2017. "Making autonomous weapons accountable: command responsibility for computer-guided lethal force in armed conflicts." In *Research Handbook on Remote Warfare,* edited by Jens David Ohlin, 405–42. Cheltenham, UK: Edward Elgar. https://doi.org/10.4337/9781784716998.00024.

———. 2019. "The Other Side of Autonomous Weapons: Using Artificial Intelligence to Enhance IHL Compliance." In *The Impact of Emerging Technologies on the Law of Armed Conflict,* edited by Ronald T. P. Alcala and Eric Talbot Jensen, 147–74. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oso/9780190915322.003.0006.

Melzer, Nils. 2009. *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law*. Geneva, Switzerland: ICRC.

Milaninia, Nema. 2020. "Biases in machine learning models and big data analytics: The international criminal and humanitarian law implications." *International Review of the Red Cross* 102 (913): 199–234. https://doi.org/10.1017/S1816383121000096.

Office of the Judge Advocate General. 2001. *Law of Armed Conflict at the Operational and Tactical Levels.* Ottawa, ON: Department of National Defence. www.fichl.org/fileadmin/_migrated/content_uploads/Canadian_LOAC_Manual_2001_English.pdf.

Quéguiner, Jean-François. 2006. "Precautions under the law governing the conduct of hostilities." *International Review of the Red Cross* 88 (864): 793–821. https://doi.org/10.1017/S1816383107000872.

Roff, Heather. 2016. "Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons." Briefing paper for delegates at the Review Conference of the Convention on Certain Conventional Weapons. Geneva, Switzerland, December 12–16. https://article36.org/wp-content/uploads/2016/12/Control-or-Judgment_-Understanding-the-Scope.pdf.

———. 2024. "Magnifying human confusion: Meaningful Human Control and the ongoing debate on autonomous weapons." *The Rule of Law Post* (blog), May 6. www.penncerl.org/the-rule-of-law-post/magnifying-human-confusion-meaningful-human-control-and-the-ongoing-debate-on-autonomous-weapons/.

Roff, Heather and Richard Moyes. 2016. "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons." Briefing paper for delegates at the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, Switzerland, April 11–15. https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf.

Sandoz, Yves, Christophe Swinarski and Bruno Zimmermann, eds. 1987. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*. ICRC. Geneva, Switzerland: Martinus Nijhoff Publishers.

Schmitt, Michael N. 2012. "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics." *Harvard National Security Journal* 4: 1–37. https://centaur.reading.ac.uk/89864/1/Schmitt-Autonomous-Weapon-Systems-and-IHL-Final.pdf.

Seixas-Nunes, Afonso. 2022. *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective*. Cambridge, UK: Cambridge University Press.

Shany, Yuval. 2024. "Red Herring, Meaningful Human Control and the Autonomous Weapons Systems Debate." Institute for Ethics in AI, University of Oxford. March 18. www.oxford-aiethics.ox.ac.uk/blog/red-herring-meaningful-human-control-and-autonomous-weapons-systems-debate.

United Nations Office for Disarmament Affairs. 2017. "Perspectives on Lethal Autonomous Weapon Systems." UNODA Occasional Papers No. 30. New York, NY: United Nations.

US Department of Defense. 2023. "DoD Directive 3000.09: Autonomy in Weapon Systems." Office of the Under Secretary of Defense for Policy, January 25. www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.PDF.

US Department of State. 2023. "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy." Bureau of Arms Control, Deterrence, and Stability, November 9. www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/.

Winter, Elliot. 2022. "The Compatibility of Autonomous Weapons with the Principles of International Humanitarian Law." *Journal of Conflict and Security Law* 27 (1): 1–20. https://doi.org/10.1093/jcsl/krac001.

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org