COMPLIANCE WITH INTERNATIONAL HUMANITARIAN LAW IN THE DEVELOPMENT AND USE OF AUTONOMOUS WEAPON SYSTEMS

What Does IHL Permit, Prohibit and Require?

LAURA BRUUN, MARTA BO AND NETTA GOUSSAC

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

The Governing Board is not responsible for the views expressed in the publications of the Institute.

GOVERNING BOARD

DIRECTOR

sipri

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

COMPLIANCE WITH INTERNATIONAL HUMANITARIAN LAW IN THE DEVELOPMENT AND USE OF AUTONOMOUS WEAPON SYSTEMS

What Does IHL Permit, Prohibit and Require?

LAURA BRUUN, MARTA BO AND NETTA GOUSSAC

sipri

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

March 2023

Contents

Acknowledgements

SIPRI and the authors express their sincere gratitude to the Foreign, Commonwealth & Development Office of the United Kingdom, for its generous financial support for this project.

In addition, the authors thank all the experts who shared their knowledge and experience in the workshop SIPRI organized in Stockholm on 10–11 November 2022: Maj. Odyne Berger, Tabitha Bonney, Leonardo Cardote, Michelle Carr, Lt Col. Duco Le Clercq, Damian Copeland, Heather Harrison Dinniss, Jeannine Drohla, Laurence Degraeve, Ola Engdahl, Laurent Gisel, Julia Grignon, Marco Grill, Sigrid Redse Johansen, Natalia Krutko, Dustin Lewis, Diane D. Lipana, Hamidou Mahfoud, Matthew McCormack, Hitoshi Nasu, Javiera Nuñez Ortega, Nada Tarbush, Charles P. Trumbull, Taylor Woodcock and Iben Yde.

The authors are also grateful for the comments provided by Merel Ekelhof, as well as SIPRI colleagues Vincent Boulanin, Dustin Lewis and Nivedita Raju.

Finally, the authors acknowledge the invaluable editorial work of Linda Nix and the SIPRI Editorial Department.

Responsibility for the views and information set forth in this report lies entirely with the authors.

Summary

In the decade-long policy conversation on autonomous weapon systems (AWS), it stands undisputed that the development and use of AWS must comply with international law, including international humanitarian law (IHL). However, how the rules of IHL should be interpreted and applied as relates to AWS remains, in some respects, unclear or disputed. Many states have, therefore, called for more focused work on how IHL applies to AWS. Besides helping to promote respect for IHL, a deeper understanding of these issues may serve important regulatory efforts by helping to identify the types and uses of AWS that are prohibited or regulated under existing IHL, and whether there are additional issues of concern which should be further clarified and developed.

This report aims to support those efforts. It aims to help states better understand what certain key principles and rules of IHL permit, prohibit and require in the development and use of AWS, especially in terms of human–machine interaction. Informed by contributions to the international policy discussion on AWS and an expert workshop that SIPRI convened in Stockholm on 10–11 November 2022, the report maps areas of common ground and identifies aspects that warrant further clarification. The key findings and recommendations of the report can be summarized as follows.

First, IHL permits, at least in part, states, through their human agents, to rely on AWS to perform IHL obligations. Specifically, states seem to agree that there is nothing in existing IHL treaties or customary IHL that prevents human agents from using autonomous functions to support, inform and implement aspects of the evaluative decisions and judgements required to comply with the rules on distinction, proportionality and precautions in attack. However, it also seems widely agreed that the *extent* to which IHL permits such reliance is not unlimited. To this end, workshop participants invoked four reasons, all of which link to the need to retain human agency in decisions to apply force: (*a*) some rules of IHL are specifically addressed to humans; (*b*) accountability cannot be transferred to machines; (*c*) some rules demand context-based decisions and judgements that cannot be translated into technical indicators; and (*d*) unlimited reliance would contravene the Martens Clause (i.e., the principles of humanity and the dictates of public conscience). Thus, it appears well-established that IHL only permits limited reliance on AWS. However, *where* the limits lie remains debated.

Second, IHL prohibits certain design characteristics and uses of AWS through the prohibition against indiscriminate attacks. A common reading of this key rule is that the design and use of an AWS that cannot be directed at a specific military objective, or that cannot be limited in its effects as required by IHL, would be off-limits. Identifying such off-limits design characteristics or uses is, however, not a straightforward task and would involve clarifying two key issues. First, clarity is needed around how to interpret the notion of 'specific military objective' in the context of weapon systems that function on preprogrammed target profiles, especially how 'specific' it has to be. While some argue that 'specific military objective' can be interpreted as a class of targets, others argue that, in the specific context of AWS, it must be one specific target. Second, there is debate around the issue of unpredictability. It is commonly agreed that the design characteristics or uses of AWS that undermine the user's ability to reasonably 'predict' the effects may significantly impede the ability to comply with IHL. However, whether unpredictability can be used as a criterion to prohibit certain design characteristics and uses of AWS remains debated, particularly because this assessment will often depend on the context, as argued by several states.

Third, compliance with IHL rules guiding the conduct of hostilities requires that the humans legally responsible for using an AWS can reasonably foresee and limit the

behaviour and effects of the system. To this end, states and experts seem to agree that users should at least possess *some* knowledge of how the system works and how it will interact with the environment of use to determine whether the attack will be lawful. However, states have offered different interpretations on how far in advance AWS users can or must fulfil these obligations, as well as on what types and standards of knowledge, foreseeability and precautions are legally required. States seem to agree that these questions are hard to answer in the abstract, as they depend on the context, including characteristics of the systems and the environment of use.

The report's findings make it clear that compliance with IHL often depends highly on context. Therefore, a deeper understanding of how to ensure compliance with IHL in the context of AWS requires unpacking the 'context dependency problem'. Thus, the report's main recommendation is for states to **review the rules of IHL in relation to specific scenarios involving AWS to generate more focused and constructive discussions on what is permitted, prohibited and required under IHL in the development and use of AWS**.

Discussing scenarios could be useful for three reasons. First, it could help states delineate design characteristics and specific uses of AWS that are either prohibited or subject to regulations. Second, it could help states specify what IHL requires of the different people involved in the development and use of AWS. Such an exercise could help states specify the standards of behaviour and knowledge that IHL requires of humans in targeting decisions, and which tasks can lawfully be delegated to machines if technology allows. Finally, concrete scenarios would help states identify circumstances and conditions that may be off-limits, not as a matter of law but as a matter of policy or ethics.

1. Introduction

In 2023, intergovernmental discussions on the governance of autonomous weapon systems (AWS) entered their 10th year. Primarily held under the auspices of the Convention on Certain Conventional Weapons (CCW Convention), these discussions have since 2017 been led by a group of governmental experts (GGE).[1] After 10 years, it stands undisputed that the development and use of AWS must comply with international law, including international humanitarian law (IHL). It is also well-established that human–machine interaction must be taken into account in the development and use of AWS to ensure compliance with IHL. However, the interpretation and application of the rules of IHL in general—and particularly in relation to AWS—remain, in some respects, disputed or unclear. Several states participating in the GGE have called for more focused work on how IHL applies in the specific context of AWS.[2] Clarifying how the rules of IHL apply in the development and use of AWS is deemed critical to determining not only the types and uses of AWS that are—or should be—prohibited or regulated, but also what IHL compliance requires of states in the development and use of AWS.

This report is a response to this need. By mapping the spectrum of views regarding what IHL compliance entails in relation to several key issues concerning AWS—identifying areas of common ground as well as aspects that merit further clarification—this report aims to facilitate a deeper and more precise discussion of how IHL applies in the context of AWS. The report is structured around three questions that underpin critical issues for clarifying and developing the normative and operational framework governing AWS:

1. What does IHL permit with respect to reliance on autonomous functions in targeting-related decisions and judgements?
2. What does the IHL rule on indiscriminate attacks prohibit in the development and use of AWS?
3. What does IHL require with respect to knowledge, foreseeability, care, and precautions in the development and use of AWS?

The subject of this report is compliance with existing IHL in the context of AWS. Other legal frameworks applicable to AWS and specific ethical considerations are beyond the report's scope. The report purposefully focuses on the IHL rules that have been identified as critical for the development and use of AWS, notably those guiding the conduct of hostilities (table 1.1).[3] Other relevant obligations under IHL with respect to AWS—such as the conduct of legal review, provision of legal advice and dissemination

[1] Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (CCW Convention, or 'Inhumane Weapons' Convention), opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983.

[2] CCW Convention, Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS), 'Operationalising all eleven guiding principles at a national level', Commentary by Portugal, Aug. 2020; 'United Kingdom proposal for a GGE document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)', Proposal by the United Kingdom, Mar. 2022; and 'Roadmap towards new protocol on autonomous weapons systems', Proposal by Argentina, Costa Rica, Guatemala, Kazakhstan, Nigeria, Panama, Philippines, Sierra Leone, Palestine and Uruguay, Mar. 2022.

[3] Boulanin, V., Bruun, L. and Goussac, N., *Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human–Machine Interaction* (SIPRI: Stockholm, 2021); and Bo, M., Bruun, L. and Boulanin, V., *Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS* (SIPRI, Stockholm: Oct. 2022).

Table 1.1. Primary international humanitarian law obligations of particular relevance to the development and use of autonomous weapon systems

Obligation	IHL source
The principle of distinction obliges parties to an armed conflict to distinguish between the civilian population and combatants, between militarily active combatants and those *hors de combat* (e.g. those expressing an intention to surrender or who are wounded or sick) and between civilian objects and military objectives, and accordingly to direct attacks only against military objectives. The principle of distinction prohibits making a civilian population, as well as individual civilians, the object of attack.	AP I, Arts 41, 48, 51(2), 51(4), 51(5); CIHL, Rules 1, 6, 7, 13, 47
The prohibition against indiscriminate attacks prohibits attacks that are of a nature to strike military objectives and civilians or civilian objects without distinction, because such an attack: (*a*) is not directed at a specific military objective, (*b*) employs a method or means of combat which cannot be directed at a specific military objective, or (*c*) employs a method or means of combat the effects of which cannot be limited as required by IHL.	AP I, Art. 51(4)(5)(a); CIHL, Rules 11–13
The principle of proportionality prohibits the conduct of an attack that may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, that is excessive in relation to the concrete and direct military advantage anticipated.	AP I, Art. 51(5)(b); CIHL, Rule 14
The requirement to take precautions in attack requires taking constant care in military operations to spare the civilian population, civilians and civilian objects. Those who plan or decide on an attack must: (*a*) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives, (*b*) take all feasible precautions in the choice of means and methods of attack with a view to avoiding, or at least minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects, and (*c*) refrain from deciding to launch any 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. Moreover, an attack must be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection, or that the attack 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.	AP I, Art. 57; CIHL, Rules 15–19

AP I = Additional Protocol I (see note 7); CIHL = customary IHL (see note 7); IHL = international humanitarian law.

of IHL to armed forces, and suppression and repression of IHL violations—merit further attention elsewhere.

This report refers to AWS as well as 'autonomous functions' as relevant to the specific issue under analysis. While there is no internationally agreed definition of AWS, this report uses a working definition which defines AWS as weapons that, **once activated, can identify, select and apply force (lethal or non-lethal) to targets without human intervention**. This working definition recognizes that it is the *nature* of the tasks that are completed autonomously by a machine that primarily matters, not the *level* of autonomy of the systems as a whole. While autonomy may be applied to a number of functions in ways that can drastically affect the overall behaviour and effect of the systems (e.g., mobility, interoperability), this working definition further

Box 1.1. What are the unique characteristics of autonomous weapon systems?

Autonomous weapon systems (AWS) come in many shapes and forms. Yet, at their core, they share several distinctive sociotechnical characteristics. First, AWS function based on preprogrammed target profiles and technical indicators that AWS can recognize through their sensors and software. Second, since AWS are triggered to apply force partly by their environment of use (rather than a user's input), aspects of a decision to apply force can be made further in advance than with traditional weapons. Third, a human operator may supervise and retain the possibility of overriding the system, but the system's default functioning is that human input is not required to identify and select targets, nor to apply force against them. These characteristics entail that those who configure and deploy an AWS will not necessarily know the exact targets, location, timing or circumstances of the resulting use of force.

Sources: Moyes, R., 'Target profiles', Article 36 Discussion Paper, Aug. 2019; and Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017).

recognizes that legal questions arise primarily when autonomy is used for targeting tasks such as searching, identifying, tracking, prioritizing and engaging of targets (see box 1.1 for further elaboration of the unique characteristics of AWS).[4] The report uses the term 'humans' throughout as a shorthand to refer to those acting on behalf of a states and parties to an armed conflict and those whose conduct is attributable to that party. The term 'user' encompasses 'operators' and 'commanders', and refers to the person or group of persons who plans, decides on or carries out military action involving an AWS.

The report is informed by proposals submitted to the GGE as well as by an in-person workshop that SIPRI convened in Stockholm on 10–11 November 2022. The workshop, which was conducted under the Chatham House Rule, provided a unique opportunity for experts in the field to discuss the interpretation and application of IHL in the context of AWS. The 27 legal experts in attendance from states, international organizations and academia ensured diversity and representation across a wide spectrum of views, gender and geography. However, the format also had limitations, such as the limited physical capacity which restricted the number of experts who could attend, the inability of some smaller governments to make a legal adviser available, and the fact that the workshop was conducted in English.

The report contains three substantive chapters that are structured around the three key questions and that, for each topic, identify areas of common ground and ascertain aspects meriting further clarification. Chapter 2 examines what IHL permits in terms of reliance on autonomous functions in targeting-related decisions and judgements. Chapter 3 considers what design characteristics or uses of AWS IHL prohibits, through two aspects of the prohibition against indiscriminate attacks—weapons that are by nature indiscriminate, and the indiscriminate use of a weapon. Chapter 4 discusses what IHL, notably the rules concerning distinction, proportionality and precautions in attacks, requires with respect to knowledge, foreseeability, care, and precautions in the development and use of AWS. The final chapter summarizes the report's key findings and presents recommendations for the international policy conversation on the governance of AWS.

[4] Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017), p. 6.

2. What does IHL permit with respect to reliance on autonomous functions in targeting-related decisions and judgements?

After 10 years of intergovernmental discussions on AWS, it is well-established that IHL applies to humans and not machines.[5] The view that AWS are only 'additive to, but not a substitute for' humans in decisions to use force was supported by all workshop participants.[6] However, the extent to which humans can rely on AWS or autonomous functions when performing IHL obligations remains disputed.[7] Indeed, questions around the lawful (and ethical) division of labour between humans and technology in targeting decisions have been at the centre of the AWS debate since the beginning.[8] This chapter aims to facilitate a more detailed understanding of what continues to be a key issue in the AWS debate. In doing so, the chapter is structured around the following key question: **To what extent, if at all, is it legally permissible to use autonomous functions when making evaluative decisions and judgements?**

IHL permits aspects of decision-making to be informed, supported and implemented by autonomous functions

A common reading of IHL suggests that there is nothing in existing IHL treaties or customary IHL that prevents humans from relying on autonomous functions when performing obligations under IHL. For example, as reflected in some GGE contributions, and especially during the workshop, states increasingly recognize and accept AWS as 'tools' that, to varying degrees and subject to applicable restrictions and prohibitions, humans can develop and use to partly or wholly establish or validate the information on which IHL-mandated evaluative decisions and judgements are made.[9] The critical question is, however, to what extent and under what circumstances IHL permits users of AWS to rely on such autonomous functions in targeting.[10]

The extent to which IHL permits reliance on autonomous functions is not unlimited

While there is nothing per se in existing treaty or customary IHL that prevents humans from relying on autonomous functions to inform and support aspects of human decision-making in targeting decisions, states seem to agree that the extent

[5] This principle was crystallized in GGE discussions in 2019 which established that 'IHL imposes obligations on States, parties to armed conflict and individuals, not machines'. See, e.g. CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Report of the 2019 session, CCW/GGE.1/2019/3, 25 Sep. 2019, para. 17(b).

[6] Views expressed at the Expert workshop, Stockholm, 10–11 Nov. 2022.

[7] In particular the rules guiding the conduct of hostilities—notably the principles of distinction, proportionality and precautions in attack (table 1.1). See Protocol Additional to the 1949 Geneva Conventions, and relating to the Protection of Victims of International Armed Conflicts (AP I), opened for signature 12 Dec. 1977, entered into force 7 Dec. 1978, Arts 48, 51(2), 51(4), 51(5) and 57(1); and International Committee of the Red Cross (ICRC), Rules, Customary IHL Database, [n.d.], Rules 1, 7, 13, 14, 15, 16, 17, 18 and 19.

[8] United Nations, Human Rights Council, 'Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns', A/HRC/23/47, 9 Apr. 2013, paras 89–97; and Lewis, D. A., 'On "responsible AI" in war: Exploring preconditions for respecting international law in armed conflict', eds S. Voeneky et al., *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives* (Cambridge University Press: Cambridge, 2022), p. 492.

[9] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Implementing international humanitarian law in the use of autonomy in weapon systems', Submission by the United States, CCW/GGE.1/2019/WP.5, 28 Mar. 2019; 'United Kingdom proposal for a GGE Document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)' (note 2); and Views expressed at the Expert workshop (note 6).

[10] The decision-making process that leads to a use of force in military action, such as an attack with an AWS, involves different actors. This may mean that more than one person is considered the 'user' of an AWS.

to which existing norms of IHL permit such reliance is not unlimited. The following sections set out the main strains of rationales that dominate the debate and to which states subscribe to varying degrees.

Some rules are explicitly addressed to humans. One of the most recurring arguments made during the workshop for why the performance of IHL obligations cannot be fully delegated to AWS relates to the fact that some IHL obligations are addressed explicitly to humans. The most cited example was Article 57(2)(a) of the Protocol Additional to the 1949 Geneva Conventions, and relating to the Protection of Victims of International Armed Conflicts (Additional Protocol I, AP I), which specifically refers to the duties of 'those who plan and decide upon an attack'. As pointed out by several workshop participants, this is an important rule in the AWS context as it is one of the few IHL rules that expressly demands actions from humans in the performance of IHL obligations. While open questions remain as to *who* those who 'plan and decide' are in the context of AWS (discussed in chapter 4), it suffices for now to acknowledge the explicit *human* addressees of this rule.

Accountability cannot be transferred to machines. A connected rationale for why IHL does not permit unlimited reliance on autonomous functions relates to the shared view that humans must retain responsibility for the use of weapon systems as machines cannot be held accountable for violations of IHL. This view is already recognized by the GGE and was reiterated by most workshop participants.[11] For example, as one workshop participant put it, 'the law does not bite on machines'. Thus, as emphasized by many, if not all, workshop participants, if the use of force does not reflect human agency, it may be difficult, and potentially impossible, to hold states or individuals responsible for unlawful conduct or consequences connected with the attack.[12]

Some rules demand context-dependent judgements that arguably cannot be translated into technical indicators. A number of states, as well as the International Committee of the Red Cross (ICRC), argue that IHL does not permit unlimited reliance on autonomous functions because compliance with key IHL rules require judgements made on the basis of values and interpretations of a particular situation rather than numbers or technical indicators.[13] This is, arguably, especially the case with regard to the principles of distinction, proportionality and precautions in attack.[14] To sustain that view, workshop participants cited a number rules and standards guiding the conduct of hostilities, including:

[11] Views expressed at the Expert workshop (note 6); CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Principles and good practices on emerging technologies in the area of lethal autonomous weapons systems', Proposal by Australia, Canada, Japan, Republic of Korea, UK and USA, 7 Mar. 2022; 'Roadmap towards new protocol on autonomous weapons systems' (note 2); Working paper submitted by Finland, France, Germany, the Netherlands, Norway, Spain and Sweden to the 2022 Chair of the GGE on Emerging Technologies in the Area of LAWS, 13 July 2022; and CCW Convention, CCW/GGE.1/2019/3 (note 5), Annex IV, 'Guiding Principles', paras (b), (d).

[12] Human agency is a concept used across a number of disciplines and is, in certain respects, highly contested. It is assumed in this report that an exercise of human agency is a necessary condition to satisfactorily perform certain IHL obligations; such agency may be exhibited by, e.g., conducting deliberative reasoning processes and exercising volition in relation to targeting decisions. See Lewis (note 8); and ICRC, 'Ethics and autonomous weapon systems: An ethical basis for human control?', ICRC Report, 3 Apr. 2018.

[13] See 'the 'numbers' challenge' in Boulanin, V. et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (SIPRI and ICRC: Stockholm, June 2020), p. 5.

[14] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Joint "commentary" on Guiding Principles A, B, C and D', Submission by Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico and New Zealand, 2020; and Submission by Argentina, Costa Rica, Ecuador, El Salvador, Panama, Palestine, Peru, Philippines, Sierra Leone and Uruguay, Sep. 2021. See also ICRC, 'ICRC position on autonomous weapon systems', ICRC Position and Background Paper, 12 May 2021.

- the obligation to assess whether a person is taking a 'direct part in hostilities' (and thereby loses immunity from targeting in attack) and whether a combatant has been placed *hors de combat* (and thereby gains immunity from targeting in attack)[15]
- the presumption of civilian status in case of 'doubt'[16]
- the principle of proportionality, namely the obligation to assess whether an attack is expected to cause incidental loss of civilian life, injury to civilians or damage to civilian objects that would be 'excessive' in relation to the 'concrete and direct military advantage' anticipated[17]
- the obligation to cancel or suspend an attack if it becomes apparent that the objective is not a military one or is subject to special protection or that the attack would violate the principle of proportionality[18]
- the obligation to assess whether an object—by its nature, location, purpose or use—makes 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'[19]
- the prohibition on destruction of an enemy's property unless such destruction is rendered 'absolutely necessary' by military operations.[20]

During the workshop, one rule received particular attention, namely the obligation to presume civilian status in case of doubt. At least most (if not all) workshop participants agreed that while it may be technically possible to encode 'doubt' in an AWS, such code should not replace evaluative decisions made by humans. Rather, the encoding of 'doubt' should function as a fail-safe mechanism, serving as 'an additional layer of precision'.[21] Thus, this rule appeared to be a good example as to why IHL does not permit unlimited reliance on autonomous functions. Overall, a fundamental assumption forming this interpretation of the rules guiding the conduct of hostilities is that such context-based judgements required by humans must be made 'reasonably temporally proximate to an attack, to remain valid'.[22]

However, not all states share this reading of the rules guiding the conduct of hostilities. According to some states, such as the United States and the United Kingdom, there is nothing in the rules of distinction, proportionality and precautions in attack that prevents high degrees of reliance on autonomous functions when performing IHL-mandated decisions as long as decisions reflect human agency and human accountability is retained.[23] The first argument is that if the technology (and circumstances) allow, human judgement about decisions to apply force can be exercised a longer time in advance, and not necessarily when the system engages the target.[24] In the same vein,

[15] AP I (note 7), Arts 41 and 51(3); and ICRC, Rules, Customary IHL Database (note 7), Rules 6 and 47.

[16] See e.g. AP I (note 7), Arts 50(1) and 52(3).

[17] AP I (note 7), Art. 51(5)(b); and ICRC, Rules, Customary IHL Database (note 7), Rule 14.

[18] AP I (note 7), Art. 57(b).

[19] AP I (note 7), Art. 52; and ICRC, Rules, Customary IHL Database (note 7), Rule 8.

[20] Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War (GC IV), opened for signature 12 Aug. 1949, entered into force 21 Oct. 1950, Art. 53.

[21] View expressed at the Expert workshop (note 6).

[22] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Chairperson's Summary', CCW/GGE.1/2020/WP.7, 19 Apr. 2021, para. C.1.30; and View expressed at the Expert workshop (note 6).

[23] See, e.g. CCW Convention, GGE on Emerging Technologies in the Area of LAWS, CCW/GGE.1/2019/WP.5 (note 9); and 'United Kingdom proposal for a GGE document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)' (note 2).

[24] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'US commentaries on the Guiding Principles', Submission by the USA, 1 Sep. 2020; and 'United Kingdom proposal for a GGE document on the

some states have argued, at the workshop and elsewhere, that there is not necessarily an obligation to stay in control after activation and until the very moment of payload delivery. The second argument is that reliance on autonomous functions in fact contains the potential to enhance IHL compliance. To sustain this view, a handful of states and experts have asserted that, in general, there is no legal obstacle as such to using AWS that are programmed to compute elements relevant to concepts such as 'concrete and direct military advantage' and 'incidental loss' of civilian life, or to using AWS based on algorithms to help decision-makers identify gaps in available targeting information as part of the obligation to take feasible precautions.[25] For the UK, for example, the 'principle of military necessity' permits measures and weapons that engage autonomous functions where they 'are necessary to accomplish a legitimate military purpose' and are not prohibited by IHL for other reasons.[26]

However, as flagged by several workshop participants, the formulation and configuration of information proxies or technical indicators would require that the human users are willing to give numerical expression to the values they accord different factors. Also, as this interpretation holds that human agency can be exercised further in advance, how far in advance IHL permits obligations to assess, for example, 'specific and direct military advantage' and 'excessiveness' in relation to the anticipated military advantage, to be made, needs elaboration.

To advance the discussion, at least two aspects warrant further attention. First, workshop participants generally agreed that determining what the rules guiding the conduct of hostilities permit in terms of reliance on autonomous functions is highly context-dependent. As stressed by multiple experts, certain rules, circumstances of use and technical characteristics allow for greater or lesser reliance on machines than others.[27] States could, therefore, usefully identify circumstances where reliance on AWS would in principle be permitted, as the USA did in its submission to the GGE in 2019.[28] Second, states are divided on one fundamental item: whether the issue is considered a technical or legal one. At least, some workshop participants indicated that the current limitations on reliance on autonomous functions in targeting decisions are technical rather than legal. In discussions of the extent to which IHL permits reliance on autonomous functions, states need to specify whether (and which) limitations flow from legal principles concerning the human role or from current technological capabilities.

Ethical considerations. Finally, several states, as well as the ICRC, argue that IHL does not permit unlimited reliance on autonomous functions as it would contravene 'the principles of humanity and the dictates of public conscience' as reflected in the

application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (laws)' (note 2).

[25] Views expressed at the Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'US proposals', Submission by the USA, 11 June 2021, para. 3. Compliance with the principle of precautions in attack deserves particular attention in the context of AWS, and is subject to a deeper examination in chapter 4.

[26] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'United Kingdom proposal for a GGE document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)' (note 2), Annex A, p. 2.

[27] Views expressed at the Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Israel considerations on the operationalization of the eleven Guiding Principles adopted by the Group of Governmental Experts', Submission by Israel, 31 Aug. 2020; 'Contribution of Austria to the Chair's request on the Guiding Principles on emerging technologies in the area of LAWS', Submission by Austria, 2020; and ICRC, 'ICRC position on autonomous weapon systems' (note 14), p. 2 point 3.

[28] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, CCW/GGE.1/2019/WP.5 (note 9), para. 5.

Martens Clause.[29] According to this interpretation of the Martens Clause, IHL limits the use of autonomous functions in targeting decisions for ethical reasons.[30] However, this understanding is not shared by all and, as reflected in the workshop, the interpretation (and implications) of the Martens Clause, as well as broader ethical considerations, in the specific context of AWS remain an area that needs a deeper (common) understanding.

Human agency as a condition for IHL compliance

The four reasons invoked by states for limits on reliance on autonomous functions all share one important assumption about IHL compliance: **the exercise of human agency in decisions to apply force is a condition for compliance with IHL**. All workshop participants seemed to agree (albeit to varying degrees) that because of the need to retain human agency, the processes through which IHL obligations are performed and achieved cannot be fully automated. Related to this aspect, numerous states have already argued in the context of the GGE that human involvement/control/judgement (the exact term is debated) concerning the use of force is required to ensure compliance with IHL.[31] In sum, IHL permits aspects of decision-making to be informed, supported and implemented by autonomous functions. However, the extent is not unlimited. Where limits lie partly depends on what IHL prohibits and requires in terms of AWS development and use. These questions are examined in chapters 3 and 4, respectively.

[29] AP I (note 7) Art. 1(2)g (Martens Clause).

[30] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by Argentina et al. (note 14), pp. 3–4; and ICRC, 'ICRC commentary on the "Guiding Principles" of the CCW GGE on "lethal autonomous weapons systems"', July 2020.

[31] See e.g. CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Commonalities in national commentaries on guiding principles', Working paper by the chair, Sep. 2020; 'Roadmap towards new protocol on autonomous weapons systems', (note 2); Working paper submitted by Finland et al. (note 11); and 'United Kingdom proposal for a GGE document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)' (note 2).

3. What does the IHL rule on indiscriminate attacks prohibit in the development and use of AWS?

IHL treaties and customary IHL include rules that prohibit certain weapons' characteristics and effects. These include the prohibitions on indiscriminate attacks, weapons that are of a nature to cause superfluous injury or unnecessary suffering, and weapons that are intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment.[32] In the context of AWS, the prohibition against indiscriminate attacks is oft-debated and deserves particular attention.[33] This rule prohibits any attack that is of a nature to strike military objectives and civilians or civilian objects without distinction, because (*a*) the attack is not directed at a specific military objective; (*b*) the attack employs a method or means of combat which cannot be directed at a specific military objective; or (*c*) the attack employs a method or means of combat whose effects cannot be limited as required by IHL.[34] The rule thus covers at least two aspects related to the use of a weapon: attacks that involve an inherently indiscriminate weapon—that is, the weapon is indiscriminate by nature as it is incapable of being directed at specifically military objectives—and attacks that involve a weapon that could in principle be directed against a specific military objective but is used indiscriminately.

To support a deeper understanding of what characteristics and uses of AWS are (or should be) prohibited or subject to restrictions, this chapter maps views around what limits the prohibition against indiscriminate attacks places on the development and use of AWS. It examines both aspects of this prohibition, beginning with *inherently* indiscriminate weapons and then indiscriminate *use* of weapons.

The prohibition on the use of weapons that are indiscriminate by nature

The prohibition on the use of weapons that are indiscriminate by nature refers in part to which characteristics of a weapon would make its use inherently unlawful.[35] The prohibition encompasses attacks with the use of at least two types of indiscriminate weapons: those that cannot be directed at a specific military objective and those whose effects cannot be limited as required by IHL.[36] The legal determination of whether a weapon is indiscriminate by nature is typically made during the legal review process.[37] In the legal review, several factors are taken into account, including (*a*) the nature of the payload (i.e., what are the expected effects of the weapon and the scale of effect

[32] For the prohibition of weapons that are of a nature to cause superfluous injury or unnecessary suffering, see AP I (note 7), Art. 35(2); and ICRC, Rules, Customary IHL Database (note 7), Rule 70. For the prohibition of weapons that are intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment, see AP I, Arts 35(3) and 55; and ICRC, Rules, Customary IHL Database, Rule 45.

[33] Boothby, W. H., 'Highly automated and autonomous technologies', ed. W. H. Boothby, *New Technologies and the Law in War and Peace* (Cambridge University Press: Cambridge, 2018), p. 146; and Thurner, J. S, 'Means and methods of the future: Autonomous systems', eds P. Ducheine, M. Schmitt and F. Osinga, *Targeting: The Challenges of Modern Warfare* (T. M. C. Asser Press: The Hague, 2016), pp. 186 and 187.

[34] See Protocol Additional to the 1949 Geneva Conventions, and relating to the Protection of Victims of Non-International Armed Conflicts (AP II), opened for signature 12 Dec. 1977, entered into force 7 Dec. 1978, Art. 51(4)(a), (b) and (c); and ICRC, Rules, Customary IHL Database (note 7), Rules 11–13.

[35] However, 'there are differing views on whether the rule itself renders a weapon illegal or whether a weapon is illegal only if a specific treaty or customary rule prohibits its use'. ICRC, Rules, Customary IHL Database (note 7), 'Rule 71. Weapons that are by nature indiscriminate', Interpretation.

[36] AP I (note 7), Art. 51(4)(b) and (c).

[37] Article 36 of AP I (note 7) lays down an obligation to conduct legal reviews of new weapons, means and methods of warfare. While this obligation is not recognized as customary international law and states not party to AP I are, therefore, not bound by it, many states consider it good practice to implement legal reviews of new weapons, means and methods of warfare.

both in time and space); (*b*) how the payload will be delivered and whether it could be directed against a specific target; and (*c*) its intended use.[38] All these components and their interactions with each other need to be considered to generate a picture of the foreseeable performance, behaviour and effects of a weapon and assess whether it would be indiscriminate by nature.

While some argue that the lawfulness of almost all weapons (even, according to some, nuclear weapons) mainly depends on *how* they are used rather than on their intrinsic characteristics, there may be some elements or technical features of an AWS that could contribute to its characterization as indiscriminate by nature.[39] Determining what would make the use of an AWS inherently indiscriminate involves evaluating different combinations of payload, target selection parameters and context of use. This section focuses on one critical aspect of this determination: **which technical features of an AWS are prohibited on the basis of the prohibitions against weapons that are indiscriminate by nature?**

AWS that cannot be directed at a specific military objective are prohibited

The prohibition on indiscriminate attacks establishes that any means or method of warfare, including AWS, that cannot be directed at a specific military objective, and hence is of a nature to strike military objectives and civilians or civilian objects without distinction, is inherently indiscriminate.[40] An interpretation of that rule in the context of AWS, shared by many states in the GGE, is that an AWS has to be 'directable' at a specific military objective if it is not to be inherently indiscriminate.[41] While it remains unclear how this requirement should be interpreted and applied in the context of AWS, one crucial starting point pertains to the definition of a '*specific* military objective', which demands further clarification in a number of areas.

Workshop discussions revealed that the definition of a *specific* military objective is unclear among legal experts. While IHL contains a definition of 'military objectives', commentaries are silent on how 'specific' should be interpreted in the context of this norm.[42] However, the phrase 'a specific military objective' has traditionally been interpreted as referring to 'one person or object' identified by attackers as a lawful target.[43] But, as multiple workshop participants raised, this interpretation does not seem to fit AWS because, unlike traditional weapons, AWS can be designed to target generic classes of people and objects and in this sense differ from so-called signature strikes that target identified individuals. Workshop participants placed particular emphasis

[38] Boulanin, Bruun and Goussac (note 3), p. 29.

[39] See e.g. ICRC, Commentary of 1987, 'Article 51: Protection of the civilian population', p. 623, para. 1965: 'it is true that in most cases the indiscriminate character of an attack does not depend on the nature of the weapons concerned, but on the way in which they are used. However, . . . there are some weapons which by their very nature have an indiscriminate effect. The example of bacteriological means of warfare is an obvious illustration of this point. There are also other weapons which have similar indiscriminate effects, such as poisoning sources of drinking water.'

[40] AP I (note 7), Art. 51(4)(b).

[41] View expressed at Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by Argentina et al. (note 14); 'Possible answers to the Guiding Questions of the Chair of the Group of Governmental Experts (GGE) on Lethal Autonomous Weapon Systems (LAWS)', Submission by the Netherlands, Sep. 2021; and CCW Convention, 'Principles and good practices on emerging technologies in the area of lethal autonomous weapons systems' (note 11).

[42] Article 52(2) of AP I (note 7) defines 'military objectives as those 'which by their nature, location, purpose or use: (*a*) make an effective contribution to military action, and (*b*) whose total or partial destruction, capture or neutralization, in the circumstance ruling at the time, offers a definite military advantage'. For commentaries, see e.g. Bothe, M. et al., *New Rules for Victims of Armed Conflicts: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949*, Nijhoff Classics in International Law vol. 1, 2nd edn (Martinus Nijhoff: The Hague, 2013), pp. 347–48; and ICRC, Commentary of 1987, 'Article 51: Protection of the civilian population' (note 39), pp. 620–21, paras 1951–55.

[43] Views expressed at the Expert workshop (note 6).

not only on such a requirement but also on whether it should be understood differently in the context of AWS. And while experts agreed that a *specific* military objective is not as broad as a 'goal' or a 'mission', they were divided on the further specificities. According to some, a 'specific military objective' can be interpreted broadly enough to encompass a 'class of objectives' or a 'specific target group'.[44] Under this interpretation, IHL would not prohibit the use of an AWS to target an object, such as a tank, within the class of objects that it was programmed to attack. Others, however, cautioned that Article 51(5)(a) of Additional Protocol I and its corresponding customary rule prohibit aggregating separate and distinct military objectives into a *single* military objective.[45] Under this approach, IHL would potentially prohibit the use of AWS programmed to attack multiple and distinct military targets in the same mission. This debate highlights the importance for states to clarify several interpretative questions, notably around the requirements of 'specific military objective' and 'single military objective', as these are key to the determination of whether an AWS is indiscriminate by nature.

Flowing from discussions around the definition of a specific military objective are the issues of target profiles and accuracy in target recognition. Concerning the first issue, experts at the workshop particularly stressed the importance of taking the *type* of target (or target profile) into account when assessing whether attacks involving AWS can be directed at a specific military objective. A relevant distinction concerns objectives that are military by nature (such as a tank) and those that are military by location, purpose or use (such as a bridge or public building). A majority view in the workshop and expert debate appears to hold that an AWS would more reliably identify military objectives by nature because their military status never changes, whereas objectives that are military by location, purpose or use may be more challenging due to their more fluid status.[46] Moreover, a number of states, as well as the ICRC, find it relevant to distinguish AWS designed to target people (anti-personnel AWS) from those designed to target objects (anti-material AWS). Besides ethical considerations, specific concerns about the technical ability of an AWS to identify persons *hors de combat* have led a number of actors to argue that the development and use of anti-personnel AWS should be prohibited.[47] However, a number of states and experts acknowledge that these legal interpretations are based on the limitations of current recognition technology and that recognition ability in AWS might advance to the point of matching if not surpassing human recognition abilities.[48]

In relation to the second issue, accuracy in target recognition, some consider this a critical capability for determining whether an AWS can be *directed* against a specific military objective. However, not only is IHL silent as to quantifiable standards and metrics for assessing compliance with the principle of distinction, including acceptable thresholds of accuracy in discriminating between lawful military objectives and unlawful targets, but the determination of acceptable thresholds of accuracy in target identification is context-dependent and subject to different understandings.[49] In light

[44] Views expressed at the Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'US commentaries on the Guiding Principles' (note 24), p. 2, para. 1.

[45] ICRC, Rules, Customary IHL Database (note 7), Rule 13. Article 51(5)(a) of AP I (note 7) considers indiscriminate 'an attack by bombardment by any methods or means which treats as a single military objective a number of clearly separated and distinct military objectives located in a city, town, village or other area containing a similar concentration of civilians or civilian objects'.

[46] View expressed at the Expert workshop (note 6).

[47] ICRC, 'ICRC position on autonomous weapon systems' (note 14), pp. 8–9; CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by Argentina et al. (note 14), pp. 1 and 5; and Views expressed at the Expert workshop (note 6).

[48] View expressed at the Expert workshop (note 6). See also Heller, K. J., 'The concept of "the human" in the critique of autonomous weapons', *Harvard National Security Journal*, vol. 14 (Forthcoming, 2023), pp. 19–22.

[49] Bo, Bruun and Boulanin (note 3), p. 13.

of these uncertainties, the issue of standards of performance and accuracy in target recognition needs further clarification.

Furthering the understanding of these questions among states and experts—especially the standards against which to assess whether an AWS can be directed at a specific military objective, the level of generality of target profiles that is acceptable, and which types of targets are off-limits—is important for identifying more concretely the limits or requirements that the prohibition on inherently indiscriminate weapons places on the development and use of AWS.

AWS whose effects cannot be limited as required by IHL are prohibited

The prohibition on indiscriminate attacks establishes that any means or method of warfare whose effects cannot be limited as required by IHL, such that it strikes military objectives and civilians or civilian objects without distinction, is deemed indiscriminate by nature.[50] States share the view that compliance with this rule requires AWS users to be able to foresee the operation, performance and effects of the system to ensure that they can administer the weapon's operation and limit its effects as required by IHL.[51] This means that an AWS that by design prevents its users from foreseeing and administering its operation, performance and effects is prohibited. Accordingly, the UK has argued that AWS that do not have regard to 'the ability to observe, recognise and exercise situational judgement' should be prohibited.[52] However, which technical design features would be prohibited remains the subject of debate. This links to questions of foreseeability and knowledge that IHL requires of AWS users, which are addressed in detail in chapter 4.

AWS that contain inherently unpredictable characteristics are prohibited

Contributions to the GGE and workshop interventions show that the issue of predictability is a key variable when assessing whether an AWS is indiscriminate by nature.[53] It seems that an implicit requirement of ensuring that a weapon can be directed at a specific military objective and that its effects can be limited as required by IHL pertains to the user's ability to reasonably predict the behaviour and effects of the system. In the context of AWS, unpredictability concerns both technical unpredictability—that is, a system's ability to execute a task with the same system performance that it exhibited in testing, in previous applications or on its training data—and user unpredictability—that is, the user's ability to know and foresee the likely effect of an AWS (which depends on the system's features and performances as well as on a wider array of factors, such as the environment of use and mission).[54] In the context of determining whether an AWS contains technical features that make it inherently unpredictable, these two aspects of predictability mean that the issue is whether the AWS contains technical features that

[50] AP I (note 7), Art. 51(4)(b).

[51] Boulanin, Bruun and Goussac (note 3).3

[52] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'United Kingdom proposal for a GGE document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)' (note 2), Annex A, p. 1 .

[53] Views expressed at the Expert workshop (note 6); CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by Argentina et al. (note 14), p. 1; Submission by the Netherlands (note 41), p. 2; 'Elements for a future normative framework conducive to a legally binding instrument to address the ethical humanitarian and legal concerns posed by emerging technologies in the area of (lethal) autonomous weapons (LAWS)', Submission by Brazil, Chile and Mexico, 2021; and Commentary by Switzerland on operationalizing the guiding principles at a national level, 2020. See also ICRC, 'ICRC position on autonomous weapon systems' (note 14), pp. 7–8.

[54] Holland Michel, A., *The Black Box, Unlocked: Predictability and Understandability in Military AI* (United Nations Institute for Disarmament Research (UNIDIR): Geneva, 2020), p. 5.

are designed to be unpredictable or that prevent users from predicting the likely effects of its use.

Regarding technical unpredictability, a common ground emerged at the workshop that the use of AWS that are 'predictably unpredictable'—either because the AWS cannot be directed at a specific military objective or because the effects of the AWS cannot be limited as required by IHL—would violate the prohibition against indiscriminate weapons by nature.[55] This would be the case with an AWS that, for example, 'targets everything that moves' (as several experts put it).

Concerning user unpredictability, one recurring theme addressed both in policy discussions and at the workshop pertains to the issue of machine learning in general, and online learning in particular. At least a handful of workshop participants expressed the view that AWS powered by continuous online or self-learning algorithms, or that can change parameters 'on the job', would make it impossible for human users to make assessments and evaluations after testing and deployment, and are thus to be considered indiscriminate by nature, 'regardless of how the AWS is used'.[56] However, whether such technological characteristics associated with unpredictability would qualify an AWS as inherently indiscriminate, and therefore prohibited under IHL, remains a contested issue. A few experts pointed out that self-learning algorithms could make a weapon system more reliable as long as such capability has gone through testing and review prior to deployment.

As most workshop participants recognized, IHL is silent on standards of both technical predictability and user predictability. Linking back to the issue around a 'specific military objective', participants were divided on whether directing attacks against a 'specific military objective' requires the user to foresee only *what* target the AWS will engage or also *when and where* a target is engaged—that is, whether it is legally required to know the exact location of target engagement or whether it is legally sufficient to know the type of target and the geographical area in which the target can be engaged. The USA has argued that 'a weapon system's autonomous function could be used by a commander or operator *to select and engage specific targets that the commander or operator did not know of when he or she activated the weapon system*'.[57] The US view is that some levels of unpredictability about when and where force will be applied are acceptable as long as the AWS only applies forces within the preprogrammed target profile and within a pre-determined geographical area. However, this lack of clarity in IHL around the standards for predictability perhaps constitutes the greatest barrier to using unpredictability as a criterion against which to assess whether an AWS is indiscriminate by nature. Moreover, at least a handful of workshop participants stressed that it is difficult to 'stamp' certain AWS as inherently indiscriminate solely on the basis of unpredictability, as the level of acceptable unpredictability will often depend on the context of use.

Generally, because the behaviour and effects of an AWS largely depend on its interactions with the environment, a majority view expressed at the workshop was that more attention should be paid to *how* the AWS is used. Several experts argued that the 'heavy lifting' is done by the prohibition against indiscriminate use of a weapon rather than by the prohibition against indiscriminate-by-nature weapons. However, this does

[55] Views expressed at the Expert workshop (note 6). For official statements of this view, see ICRC, 'ICRC position on autonomous weapon systems' (note 14), pp. 7–8; CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by the Netherlands (note 41), p. 2; and Submission by Argentina et al. (note 14), p. 1.

[56] Views expressed at the Expert workshop (note 6).

[57] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, CCW/GGE.1/2019/WP.5 (note 9), para. 5(c) (emphasis in original).

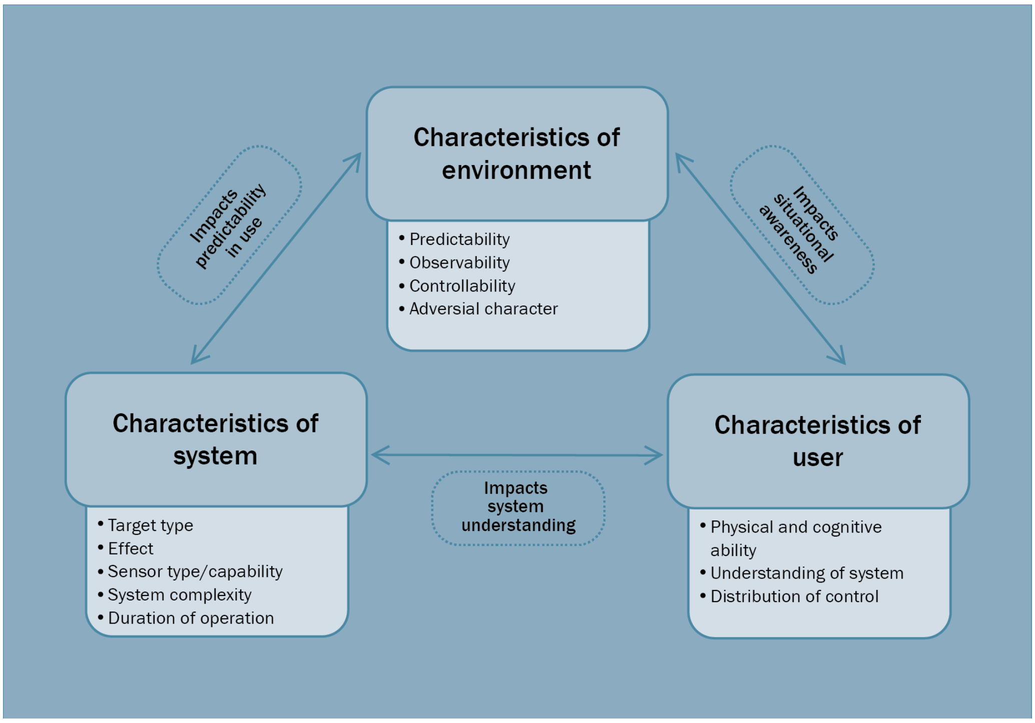
Characteristics of environment
• Predictability
• Observability
• Controllability
• Adversial character
Impacts predictability in use
Impacts situational awareness
Characteristics of system
• Target type
• Effect
• Sensor type/capability
• System complexity
• Duration of operation
Impacts system understanding
Characteristics of user
• Physical and cognitive ability
• Understanding of system
• Distribution of control

Figure 3.1. Conceptual approach to ascertaining circumstances and conditions in which an AWS could be used in compliance with IHL

Source: This figure formed part of an analysis of how to exercise human control in the context of AWS in Boulanin, V. et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (SIPRI and ICRC: Stockholm, June 2020), p. 27, fig. 3.2 (extract).

not mean that the policy debate should not engage in deeper discussions around which technical features (if any) could render an AWS indiscriminate by nature.

The prohibition on the indiscriminate use of weapons

The prohibition on indiscriminate attacks include attacks that strike military objectives and civilians or civilian objects without distinction.[58] This prohibition relates to attacks that violate the principle of distinction because the attacks are 'not aimed (for instance, an artillery shell blindly fired)'.[59] This provision has been interpreted as meaning that whether a means or method of warfare is indiscriminate comes down to its specific use: 'means or methods of combat which can be used perfectly legitimately in some situations could, in other circumstances, have effects that would be contrary to some limitations contained in the Protocol, in which event their use in those circumstances would involve an indiscriminate attack'.[60] While this is true for most means and methods of warfare, there seems to be a degree of convergence among experts that it is arguably even more relevant in the case of AWS, due to the common assumption that AWS compliance with IHL is highly context-dependent. This section explores **which circumstances and conditions would preclude parties to a conflict from using AWS in compliance with the prohibition against indiscriminate attacks**.

[58] AP I (note 7), Art. 51(4)(a); and ICRC, Rules, Customary IHL Database (note 7), Rules 1 and 7.

[59] Townley, S., 'Indiscriminate attacks and the past, present, and future of the rules/standards and objective/subjective debates in international humanitarian law', *Vanderbilt Law Review*, vol. 50, no. 5 (2021), p. 1226.

[60] ICRC, Commentary of 1987, 'Article 51: Protection of the civilian population' (note 39), p. 623, para. 1962.

Circumstances and conditions that preclude humans from reasonably foreseeing the effects and making evaluative decisions

Compliance with the prohibition against indiscriminate use of a weapon requires users to ensure that an attack is directed at a specific military objective. In general, and in the context of AWS in particular, this rule has been used by many to highlight the importance of users maintaining the ability to foresee and evaluate the performance and effects of the weapon in the intended environment of use in order to ensure the attack is limited to a specific military objective.[61] States generally seem to agree that compliance with this rule would, therefore, require limiting the geographical scope of the system's effects and target areas. A number of states and workshop participants have specifically suggested that the use of AWS must be limited to uncluttered, constrained or structured areas (such as maritime, aerial or rural environments) where AWS users are in a better position to ensure that the AWS will foresee the behaviour and effects of the systems and recognize target profiles with 'sufficient' accuracy.[62] Further proposed limits include the temporal scope of an AWS and its loitering period, to ensure sufficient control and foreseeability of the effects and behaviour of the system.[63]

With that said, states and experts are yet to agree on the exact circumstances and conditions that would preclude users from reasonably foreseeing the effects and making evaluative decisions. They do, however, seem to agree that, generally speaking, the circumstances and conditions are hard to list in the abstract as they depend on the context of use. Several workshop participants suggested that one conceptual approach that could be useful in ascertaining the circumstances and conditions in which use of an AWS would amount to an indiscriminate attack is to considering the characteristics of the environment, the user and the system as the points of a triangle, with their interactions having an impact on predictability in use, situational awareness and system understanding (figure 3.1).

The type of target profile matters

While states involved in the AWS debate remain divided in terms of which characteristics of the system and the environment (if any) should be prohibited in the abstract, there is a degree of consensus, at least observed at the workshop, that certain circumstances may be less problematic than others. As in the discussion around weapons that are indiscriminate by nature, a relevant distinction in this context concerns target profiles. A majority view appears to hold that the ability to foresee the effects of an attack using AWS (and so ensure that the attack is aimed and directed at a specific military objective) is less problematic when AWS are used in attacks against military objectives by nature.[64] In contrast, AWS used in attacks against target profiles defined by location, purpose or use means the ability to ensure that an attack is aimed and directed at a specific military objective may be more difficult, and should, according

[61] Views expressed at the Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Joint "commentary" on Guiding Principles A, B, C and D' (note 14); 'United Kingdom proposal for a GGE document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS) (note 2); and Working paper submitted by Finland et al. (note 11).

[62] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by Argentina et al. (note 14); ICRC, 'ICRC position on autonomous weapon systems' (note 14); Views expressed at the Expert workshop (note 6); and Boulanin and Verbruggen (note 4), pp. 13–14.

[63] Views expressed at the Expert workshop (note 6); and ICRC, 'ICRC position on autonomous weapon systems' (note 14), p. 10.

[64] ICRC, 'ICRC position on autonomous weapon systems' (note 14), pp. 9–10; and Views expressed at the Expert workshop (note 6).

to some (including the ICRC), be limited, if not prohibited.[65] The underlying contention is that because the status of military-by-nature objectives is static, humans can make legally required evaluative decisions regarding their status as legitimate targets a longer time in advance.[66] However, just because states seem to agree that attacks involving AWS against military objectives by nature are less problematic, that does not mean that states agree that attacks against objectives that are military by location or use are problematic per se and should be prohibited. As argued by one workshop participant, in the context of AWS what matters is which targets can be translated into technical indicators. In a similar vein, the obligation to recognize whether a combatant has been placed *hors de combat*—which according to some constitutes one of the biggest legal challenges when using AWS directed against people—has led some states, as well as the ICRC, to suggest prohibiting such use.[67]

Despite the apparent agreement about the requirement that users must retain the ability to reasonably foresee the performance and effects of an AWS to ensure that an attack complies with the rule on indiscriminate attacks, there is no consensus on specific circumstances and conditions that would prevent users from fulfilling this requirement and what limits should consequently be placed on the development and use of AWS. Moreover, states hold different views as to the standards of knowledge and foreseeability required of users. This related issue is addressed in chapter 4.

[65] Views expressed at the Expert workshop (note 6); and ICRC, 'ICRC position on autonomous weapon systems' (note 14), pp. 9–10.

[66] ICRC, 'ICRC position on autonomous weapon systems' (note 14), p. 9.

[67] ICRC, 'ICRC position on autonomous weapon systems' (note 14), p. 9; CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Submission by Argentina et al. (note 14); and Views expressed at the Expert workshop (note 6).

4. What does IHL require with respect to knowledge, foreseeability, care, and precautions in the development and use of AWS?

This chapter examines what key IHL rules applicable to the conduct of hostilities—that is, the rules of distinction, proportionality and precautions in attack—require of humans in the development and use of AWS. This is important for three reasons. First, states have expressed a desire for practical guidance to ensure that they and their armed forces respect IHL when planning or deciding to use AWS in a military operation. Second, it will help contribute to a better understanding of what types and degrees of human–machine interaction are needed to ensure compliance with IHL in the context of AWS. Third, deeper knowledge of these requirements may aid the ability of states to identify when IHL has been violated.

This chapter first maps state views concerning what AWS users are required to know and foresee to comply with the rules of distinction, proportionality and precautions in attack. It then examines state views about what IHL requires of humans when planning, deciding and launching attacks, focusing on the IHL requirement to take precautions *in* attack, as these rules include practical details that give effect to IHL rules relating to distinction and proportionality.[68]

Standards of knowledge and foreseeability

One of the unique characteristics of AWS is that they can select and apply force to targets based on preprogrammed target profiles. This, in turn, means that those who use an AWS will not necessarily know the exact location, timing and circumstances of any future application of force by the AWS (box 1.1). This unique characteristic raises questions about what AWS users should know and foresee to comply with IHL, specifically the rules of distinction, proportionality, and precautions in attack (see table 1.1). A common reading of these rules is that compliance impliedly requires the ability to *reasonably foresee* whether the effects of an AWS would in some or all circumstances contravene specific or general prohibitions and restrictions on weapons, means and methods of warfare. However, IHL does not expressly lay down the required levels of reliability or predictability of weapon systems (e.g., a maximum 'fail rate').[69] To ensure that AWS users fulfil IHL obligations, and to strengthen the ability to distinguish IHL violations from accidents, a deeper and more precise understanding is needed of **what knowledge and foreseeability may be required of AWS users**.

AWS users are required to have some level of knowledge about the characteristics of the system, its environment of use and the anticipated effects

Contributions to the AWS debate indicate that, as with other types of weapons, the type of knowledge that must be held by users of AWS will depend on what is considered reasonable under the circumstances ruling at the time of the attack.[70] With that said, it is a common view that, regardless of the circumstances, users must have at least

[68] This focus is warranted given the balance of duties resting on attacking parties and also reflects the focus of discussions amongst legal and military experts in the GGE and elsewhere. However, this focus does not downplay the practical implications of the duty to take precautions *against the effects* of an attack involving an AWS, which is touched on here but merits deeper discussion elsewhere.

[69] Boulanin, Bruun and Goussac (note 3).

[70] Views expressed at the Expert workshop (note 6). See also e.g. CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'US commentaries on the Guiding Principles' (note 24), p. 2, para. 2(a), (c); and Working paper submitted by Finland et al. (note 11).

some level of knowledge about the characteristics of the system, its environment of use and the anticipated effects.[71] Workshop participants considered this to mean that a user of an AWS should be aware of at least six types of information: (*a*) the categories of persons or objects to which the AWS could apply force in its period and area of operations; (*b*) factors that will influence whether and when an AWS will apply force to a target; (*c*) any environmental factors that could have an effect on the functioning of the AWS; (*d*) how and when the user can interact with the AWS; (*e*) factors relevant to the assessment of the lawfulness of the proposed attack; and (*f*) risk of error or malfunctions and of target misidentification.

A key question is whether AWS users are legally required to have *technical* knowledge about the design and functioning of the AWS. Several workshop participants argued, like many have done before them, that it would be neither legally nor practically feasible to require a user of an AWS to have a deep technical knowledge of how the AWS has been programmed.[72] Yet it is arguable that users need *some* technical knowledge about the design and functioning of an AWS to be able to make decisions about the lawfulness of an attack. To this end, states could usefully consider the different knowledge requirements associated with particular types of AWS. For example, because some algorithms are descriptive (they describe the past) and others are prescriptive (they recommend actions), the user of an AWS using prescriptive algorithms needs to understand how they work to assess the likely effects and behaviour of the system, whereas the user of an AWS using descriptive algorithms does not necessarily need to know how they work to make the same assessment.[73]

IHL requires a high level of knowledge and foreseeability, but the level of uncertainty tolerated by IHL is not clear

A common interpretation of the rules of distinction, proportionality and precautions in attack suggests that these rules do not demand absolute certainty from users about the effects of an attack, which is likely considered an unrealistic standard.[74] Decisions by military commanders or other persons responsible for planning or deciding on an attack are to be made in good faith and based on the information reasonably available to them at the time. This view is reflected in the AWS debate where states seemingly agree that IHL does not require absolute certainty about either the characteristics of an AWS, the environment of use or anticipated effects. However, this begs the question of how much uncertainty is tolerated under IHL, both generally and in the case of AWS. A useful starting point might be to consider the ICRC's Commentary of 1987 on Additional Protocol I, which states that commanders in the phase of target identification, 'in case of doubt, even if there is only slight doubt', are required to acquire additional information, and how that requirement should be applied and understood in the context of AWS.[75]

Some workshop participants and states have proposed that a user of an AWS must be 'reasonably certain' or 'sufficiently certain' or have 'sufficient assurance' of its effects in a given attack.[76] Others have reiterated the general view that to comply with IHL

[71] Views expressed at the Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Working paper submitted by Finland et al. (note 11); and CCW Convention, 'Principles and good practices on emerging technologies in the area of lethal autonomous weapons systems' (note 11), paras 19 and 20(f).

[72] Views expressed at the Expert workshop (note 6).

[73] View expressed at the Expert workshop (note 6).

[74] See ICRC, Commentary of 1987, 'Article 57: Precautions in attack'.

[75] See ICRC, Commentary of 1987, 'Article 57: Precautions in attack' (note 74), p. 6180, para. 2195.

[76] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Working paper submitted by Finland et al. (note 11), p. 2.

rules on the conduct of hostilities, 'commanders and other decision-makers must make a good faith assessment of the information that is available to them at the time'.[77] These contributions reflect existing debates about the interpretation and application of IHL rules on the conduct of hostilities regarding non-autonomous weapons, for example in the context of the reverberating effects from the use of explosive weapons with wide-area effects in populated areas.[78] What distinguishes the discussion regarding AWS is that the question of foreseeability relates to the foreseeability and knowledge of 'first order' effects of an attack (i.e., the application of force to a target), not only the potential secondary effects.

Debate on this topic in the context of AWS is focused on two issues. The first issue is the relationship between IHL and international criminal law (ICL). Some experts have advised against importing notions of 'mental elements' (e.g., knowledge, intention) from ICL to IHL.[79] These experts emphasize that mistakes or errors made by commanders during the conduct of hostilities will not necessarily amount to violations of IHL, let alone war crimes. Others point out, however, that clarifying the level of knowledge or intent required by IHL rules relating to the conduct of hostilities, including in comparison with the relevant standard under ICL, would assist users of AWS to ensure that they exercise the requisite level of control in given circumstances.

The second issue is how to balance the need for clarity about what is expected of human users of AWS against those users' need for flexibility on the battlefield. Not everyone sees it as helpful to articulate an explicit standard of knowledge and foreseeability with regard to AWS use. However, while it is important not to conflate IHL and ICL, such articulation may nonetheless be important in the context of AWS use, not least to distinguish unintended effects (that are not the result of an IHL violation) from unintended effects that may be indicative of an IHL violation.

Not all unintended incidents will be IHL violations

While it is undisputed that attacks *intentionally* directed against protected persons or objects or *intentionally* causing excessive harm constitute violations of IHL, the legality of similar but *unintentional* effects is less clear. This lack of clarity is arguably exacerbated by AWS, at least according to those who associate the use of AWS with higher degrees of unpredictability and are thus concerned that the risk of unintended incidents will be higher or that the concept of unintended incidents will be used to deflect responsibility for IHL violations involving AWS.[80]

Nevertheless, states and experts appear to concur that unintended incidents are not violations of IHL as long as at least two conditions are met. The first condition is that the incident was the consequence of an 'unwilful' malfunction, error or mistake (i.e., something went wrong that no one intended or foresaw).[81] The second condition is that the unintended consequence arose *despite* the party taking sufficient/due/reasonable

[77] CCW Convention, GGE on Emerging Technologies in the Area LAWS, 'US commentaries on the Guiding Principles' (note 24), p. 2, para. 2(a); see also CCW Convention, CCW/GGE.1/2019/3 (note 5), para. 17(f).

[78] See e.g. ICRC, *Explosive Weapons with Wide Area Effects: A Deadly Choice in Populated Areas* (ICRC: Geneva, Jan. 2022), p. 97: 'While there must be a causal link between the attack and the reverberating effects, there are no temporal or geographic requirements *other than being reasonably foreseeable* for the determination of the effects to be considered.' (emphasis added). See also Giorgou, E., 'Explosive weapons with wide area effects: A deadly choice in populated areas', ICRC Humanitarian Law & Policy Blog, 25. Jan 2022.

[79] See e.g. Rome Statute of the International Criminal Court, opened for signature 17 July 1998 (Rome) and 18 Oct. 1998 (New York), entered into force 1 July 2002, Art. 28, which holds that commanders are criminally responsible if they knew or should have known that subordinates committed, were committing or were going to commit a war crime.

[80] Views expressed at the Expert workshop (note 6).

[81] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'US commentaries on the Guiding Principles' (note 24), p. 5 para. 7; and Author's note on ICRC statement made at the Second Session of the CCW Convention, GGE on Emerging Technologies in the Area of LAWS, Geneva, 27 July 2022, UN Web TV, 00:28:00.

care (the exact standard is unclear) in the military operation to comply with IHL rules of distinction, proportionality and precautions in attack.[82] These conditions underline the importance of understanding what IHL requires of AWS users.

Despite the emerging common view that unintended incidents involving AWS are not IHL violations, more clarity is needed regarding what situations these terms cover in attacks involving AWS. States have used a range of terms in the GGE to describe incidents that would not amount to IHL violations, for example, 'unintended engagements', 'accidents', 'malfunctions' and 'technical errors'. It is not clear whether these terms, and the kinds of incidents to which they refer, are congruent.

In sum, a deeper and more structured discussion is needed around the characteristics that define incidents involving AWS as the consequence of 'accidents', technical or user 'errors', 'malfunctions' and 'mistakes', and which types are (or should) be considered violations of IHL.[83]

The obligation to take feasible precautions in attack

The obligation to take feasible precautions in attack is an important protective rule, which also helps parties to conflict ensure they respect the principles of distinction and proportionality. When it comes to AWS, the obligation to take feasible precautions is touted as a critical safety net. Despite the importance of this fundamental principle, it remains unclear in certain key respects what compliance with this rule requires, in concrete terms, of the people involved in the development and use of AWS. While AWS may offer the possibility of 'technical' precautions (i.e., built-in or preprogrammed safeguards) not available in non-autonomous weapons, they also place a significant demand on individuals involved in planning or deciding on their use to ensure that all feasible precautions are taken to avoid, or at least minimize, harm to civilians and civilian objects. Most, if not all, workshop participants stressed that the principle of precautions in attack deserves particular attention in relation to AWS. A deeper understanding will not only strengthen compliance with the rules guiding the conduct of hostilities, but also support policy efforts to identify the requisite types and degrees of human–machine interaction that will ensure IHL compliance in the context of AWS. This section examines **the spectrum of views concerning what the precautions obligation entails across the personal (who), material (what) and temporal (when) dimensions**.

Every human involved in planning or deciding on an attack involving AWS is required to take feasible precautions

Additional Protocol I states that it is 'those who plan or decide upon an attack' who must take precautions, framed in terms of a number of requirements.[84] It is uncontested that this also applies in the context of AWS.[85]

Moreover, states in the GGE and workshop participants appear to have found common ground that the obligation to assess and adopt feasible precautions in any potential attack may rest with several individuals. This flows from common military practice where planning and deciding on an attack typically constitutes a shared set of

[82] Views expressed at the Expert workshop (note 6).

[83] A. Seixas-Nunes, *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective* (Cambridge University Press: Cambridge, 2022), pp. 209–15.

[84] AP I (note 7), Art. 57(2). Compare CIHL Rule 15, which is phrased in the passive voice; see ICRC, Rules, Customary IHL Database (note 7), Rule 15.

[85] See e.g. CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'United Kingdom proposal for a GGE Document on the application of international humanitarian law to emerging technologies in the area of lethal autonomous weapon systems (LAWS)' (note 2), Annex A, p. 2; and Views expressed at the Expert workshop (note 6).

tasks.[86] For example, the USA has argued that the requirements to take precautions are implemented in military operations through 'responsible commands', where different individuals within the command will implement different duties—for example, 'the decision about whether a particular precaution is feasible might be made by a commander at a particular level of command with the authority to direct the resources necessary to take that precaution, or individual units within the command might be tasked with carrying out a precaution, such as delivering a warning'.[87] However, this broad personal scope for implementation of the obligation to take feasible precautions (and indeed other IHL rules on the conduct of hostilities) warrants further elaboration. In the context of AWS, states could focus discussions on the implications for this obligation of distributed responsibilities for AWS use within modern command-and-control structures.[88] Such discussions should aim to clarify the roles and responsibilities of the people involved in the development and use of AWS and explore the specific ways, if any, that the multiple people involved can enhance the implementation of the obligation in terms of, for example, catching potential mistakes in targeting, verifying targeting information and identifying risks of wrongful engagement.

The specific feasible precautions that are required under IHL will vary according to context

IHL does not explicitly state *what* feasible precautions entail nor the means by which they should be achieved. But as with other IHL rules on the conduct of hostilities, most experts accept that feasible precautionary measures in attack are highly context-dependent and cannot be determined in the abstract.[89] That is, whether a certain precaution is feasible depends on the circumstances of the attack, including the military advantage sought, the risks faced by the party's forces, the nature of the target, the technical characteristics of the weapon system in question, and the availability of other means or methods to achieve the same advantage. This general interpretation is also widely shared by states in the AWS debate.

However, besides agreeing that which measures are feasible depends on the circumstances, states have offered different interpretations of what this rule requires of humans in the context of AWS. A number of states have interpreted the obligation to take feasible precautions in attack as placing a requirement for real-time human involvement and judgement during the application of force. This interpretation flows especially from the obligation in Additional Protocol I to cancel or suspend attacks under certain circumstances.[90] However, the need for situational human involvement and judgement presents a challenge when using AWS, because the user may not necessarily know the specific circumstances of the resulting application of force. This interpretation finds support in the ICRC's Commentary of 1987 on Additional Protocol I which observes that 'with the increased range of weapons, particularly in military operations on land, it may happen that the attacker has no direct view of the objective, either because it is very far away, or because the attack takes place at night. In this case, even greater caution is required.'[91] The ICRC, among others, has therefore

[86] See e.g. Ekelhof, M. and Persi Paoli, G., 'The human element in decisions about the use of force', UNIDIR, 2020; Schulzke, M., 'Autonomous weapons and distributed responsibility', *Philosophy and Technology*, vol. 26 (June 2013); and Bo, Bruun and Boulanin (note 3), p. 19.

[87] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, CCW/GGE.1/2019/WP.5 (note 9), para. 4(c).

[88] See e.g. Boulanin, V. and Lewis, D., 'Responsible reliance concerning development and use of AI in the military domain', *Journal of Ethics and Information Technology*, vol. 25, art. 8 (2023).

[89] Views expressed at the Expert workshop (note 6); CCW Convention, GGE on Emerging Technologies in the Area of LAWS, CCW/GGE.1/2019/3 (note 5), para. 17(f); and Boulanin, Bruun and Goussac (note 3), p. 25.

[90] AP I (note 7), Art. 57(2)(b).

[91] ICRC, Commentary of 1987, 'Article 57: Precautions in attack' (note 74), p. 686, para. 2221.

Box 4.1. The concept of an 'attack' and implications for compliance with international humanitarian law

The beginning of an 'attack' in armed conflict does not signify the beginning of international humanitarian law (IHL) obligations. However, many of the IHL rules relating to the conduct of hostilities apply to 'an attack'. The temporal and geographic scope of an attack is therefore critical to determining whether there has been a breach of an obligation under IHL. Article 49(1) of Additional Protocol I defines 'attacks' as 'acts of violence against the adversary, whether in offence or in defence'.[a] How this definition should be interpreted remains subject to debate. In the case of autonomous weapon systems (AWS), whose functions may involve a time delay between activation and application of force, a key issue is when an attack begins and ends. For example, the United States considers that 'the single firing of a weapon system might only be one part of an "attack," and the mere activation of a weapon system might not constitute an "attack" at all'.[b] In contrast, some workshop participants argued that 'every shot is an attack'.[c] These distinctions have significant legal implications, as each new attack demands a new legal assessment and the taking of feasible precautions.[d] And, as Article 36 has pointed out, an attack must have some boundaries so as not to undermine the structure of the law.[e]

[a] Protocol Additional to the 1949 Geneva Conventions, and relating to the Protection of Victims of International Armed Conflicts, opened for signature 12 Dec. 1977, entered into force 7 Dec. 1978.

[b] CCW Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, 'Implementing international humanitarian law in the use of autonomy in weapon systems', Submission by the USA, CCW/GGE.1/2019/WP.5, 28 Mar. 2019, para. 4(b).

[c] Views expressed at the Expert workshop, Stockholm, 10–11 Nov. 2022.

[d] Bothe, M. et al, *New Rules for Victims of Armed Conflicts: Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949*, Nijhoff Classics in International Law vol. 1, 2nd edn (Martinus Nijhoff: The Hague, 2013), p. 329; and Dinstein, Y., 'Legitimate military objectives under the current jus in bello', *International Law Studies*, vol. 78 (2001), pp. 139 and 141.

[e] Article 36, 'Target profiles', Discussion paper, Aug. 2019, p. 9.

proposed that specific precautions must be taken in all instances of AWS use, such as limits on the types of targets, the duration, geographical scope and scale of use, and limits on the design of AWS that would permit intervention and deactivation.[92]

A competing interpretation argues that the principle of precautions in attack does not 'necessarily' require human intervention, nor that humans must 'exercise physical control at all times'.[93] An example scenario is where human users of an AWS have 'sufficient assurance' that the AWS, 'once activated, acts in a foreseeable manner in order to determine that its actions are entirely in conformity with applicable law, rules of engagement and the [users'] intentions'.[94] Germany has also argued that humans

[92] ICRC, 'Contribution by the International Committee of the Red Cross submitted to the Chair of the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems as a proposal for consensus recommendations in relation to the clarification, consideration and development of aspects of the normative and operational framework', 11 June 2021, p. 2, recommendation 3; CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'Joint "commentary" on Guiding Principles A, B, C and D' (note 14), pp. 3–4; and 'Roadmap towards new protocol on autonomous weapons systems', (note 2), section II, paras 11, 18 and 28.

[93] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'German commentary on "operationalizing" all eleven guiding principles at a national level', 2020, p. 2; and Views expressed at the Expert workshop (note 6). See also CCW Convention, GGE on Emerging Technologies in the Area of LAWS, CCW/GGE.1/2019/WP.5 (note 9).

[94] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'German commentary on "operationalizing" all eleven guiding principles at a national level' (note 93), p. 2.

are not necessarily required to be the ones who deactivate a system, but if 'necessary' a weapon system can also (lawfully) deactivate itself.[95]

Such different interpretations of what constitutes feasible precautions in the context of AWS, and what the obligation requires of humans (as opposed to autonomous functions), highlight the need for states to address this question in more detail. Further discussion of this point could usefully focus on specific types of precautionary measures and when and how they might be feasible in relation to specific uses and categories of weapons.

Feasible precautions must be taken at all stages of an attack but not necessarily in advance

Discussions on the practical aspects of the duty to take feasible precautions in attack have naturally focused on the period that immediately precedes an application of force—when the decision is made. But when does the obligation to take feasible precautions commence? This question is particularly relevant to the use of AWS because one of their defining characteristics is that they are *preprogrammed* to apply force based partly on inputs from sensors.

Several aspects of the precautions obligation are linked to the notion of an attack under IHL, the scope of which is a matter of debate (see box 4.1). When it comes to an attack using an AWS, the earliest time at which the attack has arguably commenced would be at the point when the AWS, once activated, begins attempting to match input data to a target profile. The latest point at which the attack has arguably commenced is when the AWS engages its target (person or object) by applying force. Against this background, workshop participants generally accepted that the obligation to take feasible precautions in attack is not, on its face, concerned with the development of an AWS, including its design. However, for some this did not mean that feasible precautions were irrelevant during the AWS design and development phases, especially in light of the obligation to take constant care in the conduct of military operations (discussed below). These participants posited that compliance with the obligation to take feasible precautions, in particular the obligation to cancel or suspend an attack in certain circumstances, demands that states developing AWS select means and methods that preserve the ability of AWS users to take such precautionary measures. This view is not universally held. A contrasting view, expressed by some at the workshop, is that measures aimed at minimizing harm to civilians and civilian objects that can be taken only at the design and development phase, such as 'fail-safes', should not be considered feasible precautions, but are rather simply characteristics of an AWS that are relevant to a commander's decision on the choice of means of methods of warfare. This view implies that the use of an AWS that does not include technical fail-safes is not prohibited by IHL per se, by virtue of the rule on precautions. Rather, under this view, the use of an AWS that precludes intervention or deactivation may still be lawful in certain (albeit limited) circumstances (see chapter 2).

The obligation to take constant care to spare the civilian population, civilians and civilian objects, which forms part of the principle of precautions, is not limited to the duration of a specific attack, but applies 'in the conduct of military operations'. In light of such contrasting views and the potential implications of the broader temporal scope of the 'constant care obligation', the issue of whether and how the obligation to take feasible precautions is relevant for the design of AWS and operational planning involving AWS, warrants further discussion.[96]

[95] CCW Convention, GGE on Emerging Technologies in the Area of LAWS, 'German commentary on "operationalizing" all eleven guiding principles at a national level' (note 93), p. 2.

[96] See e.g. Jenks, C. and Liivoja, R., 'Machine autonomy and the constant care obligation', ICRC Humanitarian Law and Policy Blog, 11 Dec. 2018.

5. Key findings and recommendations

This report aims to help states achieve a more precise and common understanding of how IHL applies in the development and use of AWS. Such clarification is critical to identifying what types and uses of AWS are prohibited or otherwise regulated under existing norms and which, if any, should be subject to new regulations. To this end, the report identified areas of (emerging) common ground regarding the interpretation and application of IHL in the context of AWS, as well as specific areas that remain unclear or debated and warrant further clarification, by addressing three overarching questions:

1. What does IHL permit with respect to reliance on autonomous functions in targeting-related decisions and judgements?
2. What does the IHL rule on indiscriminate attacks prohibit in the development and use of AWS?
3. What does IHL require with respect to knowledge, foreseeability, care, and precautions in the development and use of AWS?

The first section of this chapter summarizes the key areas of convergence that the report has identified in response to these questions, while the second section provides a series of recommendations to the governmental and non-governmental experts that contribute to the international debate on AWS at the GGE and in other forums. These findings and recommendations aim to help determine which aspects of the normative and operational framework applicable to AWS may need further clarification or development.

Key findings

IHL permits humans to rely on autonomous functions to inform, support and implement targeting decisions, but not without limits

There appears to be nothing in IHL treaties or customary IHL that prevents humans from relying at least in part on AWS or autonomous functions in performing obligations under IHL. It thus appears largely uncontested among states that such systems can be lawfully used to support and inform those humans in performing obligations under IHL. With that said, states also seem to agree that the extent to which IHL permits reliance on autonomous systems and functions is not unlimited. Experts reached this conclusion for a variety of reasons, including by considering accountability frameworks, the nature and structure of legal obligations (notably those requiring context-based decisions and judgements), and ethical norms. However, while it appears uncontested that IHL does not permit unlimited reliance, it is debated *where* these limits lie. A key source of divergence stems from whether the issue is considered technical or legal. To some, nothing in IHL limits humans from relying on autonomous functions (or any other means or method of warfare for that matter) when performing IHL obligations as long as human accountability is ensured. Others adopt a more principle-based approach, arguing that, regardless of how sophisticated technology becomes, IHL generally and, in particular, the context-based decisions and judgements associated with the rules on distinction, proportionality and precautions in attack, require human involvement to be reasonably temporally proximate to the application to force to remain valid, and that this therefore places significant limits on reliance on AWS or autonomous functions.

IHL prohibits the use of AWS that cannot be directed at a specific military objective or whose effects cannot be limited as required by IHL

It appears to be a shared view that the prohibition against indiscriminate attacks prohibits the use of an AWS that cannot be directed at a specific military objective or whose effects cannot be limited as required by IHL. However, how that prohibition translates to specific limits on the development and use of AWS raises important questions around, for example, how specific 'a specific military objective' should be understood in the case of AWS, the specificity and type of target profiles and acceptable standards of accuracy in target recognition. The answers are contested; according to many states and experts, the difficulty stems from compliance with this rule being context-dependent.

With that said, a number of states and experts have stressed that characteristics and uses that undermine the user's ability to 'foresee and predict' the effects of an AWS and to make evaluative decisions and judgements, as required by certain IHL rules and standards, may contravene this prohibition. All workshop participants, for example, seemed to agree that AWS that are 'predictably unpredictable' are prohibited under IHL. Generally, the issue of 'unpredictability' appears for some (but not all) states to constitute a relevant criterion for assessing whether an AWS would violate the prohibition against indiscriminate attacks. However, IHL does not lay down standards concerning predictability and views differ as to which (if any) design characteristics and scenarios of use should be prohibited on the basis of the unpredictability of an AWS and its effects. With that said, it appears that certain circumstances of use are less likely to contravene the prohibition against indiscriminate attacks on the basis of unpredictability: AWS used to target objects (as opposed to people), and only objects that are military by nature, or AWS that are used in situations where civilians or other people subject to special protection are not present. However, all circumstances of AWS use, including technical characteristics, types of military objectives and operational environments, need more detailed elaboration to enable states to reach a common understanding of the design characteristics and uses that would contravene the prohibition against indiscriminate attacks.

IHL requires AWS users to reasonably foresee and limit the effects of the use of force

The rules guiding the conduct of hostilities, and especially the principles of distinction, proportionality and precautions in attack, are commonly interpreted as requiring AWS users to reasonably foresee and limit the effects of the use of force. However, the standards of behaviour demanded by these rules—such as standards of knowledge and foreseeability, as well as the types of feasible precautions that should be taken and when—are not clearly set out in IHL. While states agree that answers to these questions are highly context-dependent, they remain divided in terms of *what* IHL requires of *whom* and *when* in the development and use of AWS. Whether this constitutes a legal gap is also subject to debate. Some consider the open-textured nature of the rules guiding the conduct of hostilities as offering the necessary flexibility, while others argue that they call for clarification—if not the introduction of new rules that would explicate what users are required to do or foresee.

Recommendations

As reiterated by most actors in the policy debate on AWS, compliance with IHL in the development and use of AWS is often context-dependent. Thus, discussing legal limits and requirements in the abstract is difficult as concerns (and opportunities) might vary depending on the characteristics of the AWS and the circumstances of use. To advance the discussion on what types and uses of AWS are, or should be, prohibited or subject to regulation, states would need to unpack the 'context dependency problem' and use concrete scenarios as a useful next step. Thus the key recommendation of this report is for the international policy process on AWS to:

Use scenario exercises to generate more focused and constructive discussions on the types and uses of AWS that are prohibited or subject to regulation.

Using scenarios exercises could help states compare their interpretations of IHL treaties and customary IHL, and how they would apply them in different contexts involving AWS. Concrete scenarios should take into account different combinations of technical capabilities and operational environments. Importantly, they should consider different target profiles, including both military objectives by nature and those that are military by use, location or purpose; and explore the limits that the principle of distinction, including the obligation to differentiate between combatants and individuals *hors de combat*, places on the use of anti-personnel AWS and the circumstances under which AWS can (if at all) be used in unpopulated versus populated areas and in defensive versus offensive operations.

Such scenario exercises could also support regulatory efforts to identify how existing IHL may be strengthened by, for example, explicitly prohibiting or regulating certain types and uses of AWS. To this end, the report recommends states to use scenario exercises in three specific ways.

Use scenario exercises to identify design characteristics and specific uses of AWS that are prohibited under IHL. Using concrete scenarios could generate a deeper understanding of what specific AWS design characteristics and uses are prohibited under IHL. For example, an exercise that particularly focuses on which design characteristics and uses of a specific type of AWS would prevent users from reasonably foreseeing the effects of an attack using that AWS and from making evaluative decisions and judgements that comply with the prohibition against indiscriminate attacks.

Use scenario exercises to specify standards and behavioural requirements required of humans in the development and use of AWS. Scenario exercises could help states present in more concrete terms what IHL compliance *requires* of the different people involved in the development and use of AWS, including the extent to which humans can (legally) rely on autonomous systems. Although applied to AWS, such exercises could also help states identify what IHL requires of humans, regardless of technological advancements. In addition, exploring what is required of different individuals in different contexts could help clarify what states should do as a matter of both law and best practices to ensure compliance with IHL.

Use scenario exercises to identify limits on AWS flowing from legal as well as ethical, policy, security and operational considerations. States should also use scenarios to explore the limits and requirements they see flowing not only from obligations under IHL but also from ethical, policy, security and operational considerations. For example, scenarios could consider which ethical standards should guide the use of AWS when used in attacks against people, or which spatial and temporal limits on the use of AWS can be derived from operational or strategic considerations.

About the authors

Laura Bruun (Denmark) is a Researcher in SIPRI's Governance of Artificial Intelligence Programme. Laura's focus is on how emerging military technologies, notably autonomous weapon systems (AWS) and military applications of artificial intelligence, affect compliance with—and interpretation of—international humanitarian law (IHL). Laura has a background in international law and security and before joining SIPRI in 2020, she worked for the London-based NGO Airwars, where she monitored and assessed civilian casualty reports from US and Russian airstrikes in Syria and Iraq.

Laura's recent publications include 'Autonomous Weapon Systems and International Law: Identifying Limits and the Required Type and Degree of Human-Machine Interaction', SIPRI Report (2021, co-author) and 'Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS', SIPRI Report (2022, co-author).

Dr Marta Bo (Italy) is an Associate Senior Researcher within SIPRI's Armament and Disarmament research area. Marta is also a Researcher at the University of Amsterdam-Asser Institute in The Hague and a Research Fellow at the Graduate Institute Geneva where in July 2022 she completed a four-year research project on autonomous weapon systems and war crimes. Moreover, Marta leads, designs and implements capacity-building training programmes for judges and prosecutors in international and transnational criminal law, international humanitarian law, and human rights law. Marta is a Member of the Steering Committee of the Antonio Cassese Initiative for Justice, Peace and Humanity and editor of the international criminal law section of the Leiden Journal of International Law.

Marta was the lead author of the SIPRI publication 'Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS', SIPRI Report (2022).

Netta Goussac (Australia) is a Special Counsel with Lexbridge Lawyers. Netta has worked as an international lawyer for over a decade, including for the International Committee of the Red Cross (ICRC, 2014–20) and the Australian Government's Office of International Law (2007–14), and as a lecturer at the Australian National University. Netta has provided legal and policy advice related to new technologies of warfare, including autonomous weapons, military applications of artificial intelligence, and cyber and space security.

Netta's recent publications include 'Autonomous Weapon Systems and and International Law: Identifying Practical Elements of Human Control', SIPRI Report, (2021, co-author) and 'Autonomous Weapon Systems and International Law: Identifying Limits and the Required Type and Degree of Human-Machine Interaction', SIPRI Report (2021, co-author).

**STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE**

Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org