sipri

AUTONOMOUS WEAPON SYSTEMS AND INTERNATIONAL HUMANITARIAN LAW

Identifying Limits and the Required Type and Degree of Human–Machine Interaction

VINCENT BOULANIN, LAURA BRUUN 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.

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

AUTONOMOUS WEAPON SYSTEMS AND INTERNATIONAL HUMANITARIAN LAW

Identifying Limits and the Required Type and Degree of Human–Machine Interaction

VINCENT BOULANIN, LAURA BRUUN AND NETTA GOUSSAC

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

June 2021

Contents

Preface

The Stockholm International Peace Research Institute (SIPRI) has been involved in the international debate on autonomous weapon systems (AWS) since the issue was tabled on the agenda of the United Nations Certain Conventional Weapons Convention in 2013. SIPRI's contributions have consistently aimed at supporting a more informed and structured discussion. In 2019 and 2020, several governments stressed the need for further focused work on the applicability of international humanitarian law (IHL), and in early 2020 SIPRI initiated a one-year research project to explore the issue.

This report outlines the key findings of that project, with the aim of helping states elaborate their views on the legal provisions for the development and use of AWS, particularly with respect to the required type and degree of human–machine interaction. Rather than taking the form of legal advice, the report is a mapping exercise. It seeks to help states identify key issues and challenges related to the interpretation and application of IHL in this field, with a particular focus on where 'legal gaps' might exist or emerge. The report does not pre-judge the policy response that should regulate AWS. Instead, it aims to provide an analytical framework for states and experts to assess how the normative and operational framework regulating the development and use of AWS may need to be clarified and developed further.

The report concludes with recommendations for the need to sharpen and deepen the discussions on IHL by governments, the need for practical measures concerning IHL compliance, and the need for ethical clarity among governments on the issue of IHL and AWS as a whole. SIPRI commends this report primarily to government decision makers and experts who are engaged in the debate on the governance of AWS and, therefore, on the interpretation and application of IHL for emerging military technologies, but also to members of the interested general public.

Dan Smith
Director, SIPRI
Stockholm, June 2021

Acknowledgements

The authors would like to express their sincere gratitude to the Dutch Ministry of Foreign Affairs, the Swedish Ministry for Foreign Affairs and the Swiss Federal Department of Foreign Affairs for their generous financial support of this project.

The authors are indebted to all the experts who shared their knowledge and experience in the virtual discussion series that SIPRI conducted in June 2020 on 'Autonomous Weapon Systems and International Humanitarian Law': Mirco Anderegg, Marta Bo, Tabitha Bonney, Maya Brehm, Karl Chang, Rebecca Crootof, Mike Culver, Neil Davison, Merel Ekelhof, Ola Engdahl, Paola Gaeta, Robin Geiss, Anthony Gillespie, Martin Hagström, Erin Hahn, Natalia Jevglevskaja, Dustin Lewis, Rain Liivoja, Ian MacLeod, Nadia Marsan, Eve Massingham, Simon McKenzie, Mitch Mitchell, Charlie Savourel, Joy Sellam, Michael Siegrist, Nicoline Swinkels, Aaron Waldo, Pauline Warnotte, George Wilson, Jun Yamada, Binxin Zhang and Wen Zhou.

In particular, they wish to acknowledge the substantial intellectual contribution of Dustin Lewis, who kindly provided extensive and constructive feedback throughout the production of the report.

The authors are also grateful for the comments provided by Neil Davison and Maya Brehm, as well as a number of SIPRI colleagues: Sibylle Bauer, Lucie Béraud-Sudreau, Elena Haapaniemi, Luke Richards, Nivedita Raju and Dan Smith. Finally, they would like to acknowledge the invaluable work of the SIPRI Editorial Department.

Responsibility for the information set out in this report lies entirely with the authors. The views and opinions expressed are those of the authors and do not represent the official views of SIPRI or the project funders.

Abbreviations

AI	Artificial intelligence
AWS	Autonomous weapon systems
CCW	Certain Conventional Weapons (Convention)
GGE	Group of Governmental Experts
ICL	International criminal law
IHL	International humanitarian law
IHRL	International human rights law
LAWS	Lethal autonomous weapon systems
NGO	Non-governmental organization

Executive Summary

Compliance with international humanitarian law (IHL), along with ethical and security considerations, is a critical benchmark for assessing the acceptability of autonomous weapon systems (AWS). However, in certain key respects, how and to what extent existing IHL rules provide limits on the development and use of AWS remains disputed among states and non-governmental experts contributing to the debate on AWS at the United Nations Certain Conventional Weapons (CCW) Convention and in other relevant forums. This report explores a central element of that dispute: the question of what type and degree of human–machine interaction is required to comply with IHL.

The report is the result of a one-year SIPRI project on 'Autonomous Weapon Systems and International Humanitarian Law', which aims to provide states and experts with: (*a*) an overview of the relevant rules on IHL and the limits they place on the development and use of AWS (chapter 2); (*b*) an overview of key issues associated with the interpretation and application of the rules on weapons, means and methods of warfare (chapter 3), the rules on legal reviews and legal advice (chapter 4), and the legal frameworks governing state responsibility and individual criminal responsibility (chapter 5); and (*c*) recommendations for clarification and development of the normative and operational framework (chapter 6). The key takeaways can be summarized as follows.

IHL already places limits on the development and use of AWS, notably through its specific and general rules on weapons, means and methods of warfare. Based on an overview of the relevant IHL rules, the report concludes that securing respect for IHL in the development and use of AWS presupposes the fulfilment of three conditions:

1. The ability to reliably *foresee* whether the effects of an AWS would in some or all circumstances contravene specific and/or general prohibitions and restrictions on weapons, means and methods of warfare. The employment of an AWS whose operation, behaviour and effects cannot be sufficiently foreseen would most likely be unlawful.

2. The ability to *administer* the operation of an AWS in a manner that is consistent with the rules governing the conduct of hostilities. The use of an AWS whose operation, behaviour and effects cannot be limited according to IHL, notably the principles of distinction, proportionality and precautions, would be unlawful.

3. The ability to *trace* the operation, performance and effects of AWS back to the relevant human agent(s). The employment of an AWS that cannot be satisfactorily attributed, discerned or scrutinized would preclude assessing and imposing state responsibility and/or individual criminal responsibility for violations.

How these three conditions ought to be secured through human–machine interaction is a critical interpretative question that states need to further articulate views on. The report finds that IHL does not provide a general answer on what type and degree of human–machine interaction is required for IHL compliance. It also finds that states and experts may approach this question differently, depending on whether they see IHL as: (*a*) solely an effects-based regime, permitting militaries to use any combination of humans and machines to undertake military action as long as the anticipated or actual effects are not unlawful; or (*b*) a regime that also mandates evaluations and judgements by human beings in the conduct of military operations. Consequently, the report identifies a number of threshold questions that states and experts could use

to clarify and elaborate their interpretations of what IHL compliance demands (listed in appendix A). These pertain to respect for IHL provisions across four dimensions: personal, material, temporal and geographical.

The first dimension relates to *who* may or must respect IHL provisions. For example, while the GGE has acknowledged the need to retain 'human responsibility', it remains unclear whether responsibility for the decision to employ an AWS, and the resulting effects, reside with one single person or multiple people. If the latter, how do individual contributions to the ultimate decision to employ an AWS, and to administer its operation, interact in the systemic multi-agent model? The second dimension relates to *what* type and degree of human–machine interaction the substantive and procedural rules of IHL require, permit, or prohibit. For example, to what extent does IHL mandate, allow or bar legally required value judgements to be entrusted partly or fully to machine processes? The third and fourth dimensions relate to *when* and *where* (in relation to what locations) IHL provisions need to be respected. For example, how far back in time and how far in space may IHL-mandated evaluations be made?

A final and cross-cutting question is *how* the lack of foreseeability introduced by the use of autonomy should be addressed and managed under IHL. For example, how should unpredictability be evaluated in advance (as part of the legal review and legal advice), controlled during employment, and assessed afterwards (to trace individual and state responsibility for potential IHL violations)? The combination of these different sets of questions is critical in order to determine what would make an AWS unlawful per se, but also to evaluate how IHL provisions should be respected in the development and use of AWS: who must do what, when, where and how.

The report concludes with three recommendations. First, *states should deepen and sharpen their discussions on what respecting and ensuring respect for IHL means for AWS*. While the adoption of 11 guiding principles by the CCW Convention's Group of Governmental Experts contributed to clarifying some of the legal issues, many questions remain—both old and new—with regard to what IHL requires, permits or prohibits in the development and use of AWS. This report provides a framework for states to elaborate their views on a range of central matters, particularly the question of what is and should be expected from humans, AWS and their interactions in order to secure respect for IHL.

Second, *states should share views and experiences about practical measures that could enhance respect for IHL in the development and use of AWS*. States could further elaborate what standards of knowledge and behaviour are expected to: (*a*) allow the user to foresee whether the operation, performance and effects of the system would be lawful; (*b*) ensure that the user can administer the weapon's operation and limit its effects as required by IHL; and (*c*) ensure that the consequences of employing an AWS can be satisfactorily traced back to an individual and/or a state.

Third, *states should further elaborate the legal and ethical bases for human–machine interaction in relation to IHL compliance*. States need to clarify their ethical presumptions about why particular forms of human–machine interaction may be warranted in relation to IHL compliance: whether the choice of human–machine interaction only needs to be guided by the need to limit the risk of producing unlawful effects or whether it also needs to ensure human agency and responsibility in the exercise of IHL obligations. Answering that question is essential for clarifying whether some type and degree of human–machine interaction would always be needed regardless of the characteristics of the weapon system and environment of use. The question is also relevant for discussion of whether the CCW process should look beyond the sole case of AWS and aim more broadly to develop norms to preserve human agency in exercising IHL obligations.

1. Introduction

Since 2013, government officials, scholars and civil society representatives have been discussing the range of legal, ethical and security challenges that could be associated with emerging technologies in the area of lethal autonomous weapon systems (LAWS) (see the definition in box 1.1). The debate, which is now primarily carried out through the work of a Group of Governmental Experts (GGE) under the auspices of the United Nations Certain Conventional Weapons (CCW) Convention, was ignited by a coalition of non-governmental organizations (NGOs) inviting states to adopt a new CCW protocol to ban the development and use of LAWS.[1]

Whether a new CCW protocol is needed or otherwise warranted to regulate or prohibit the development, use or transfer of (lethal) autonomous weapon systems (AWS) and related technologies has remained a vexed question since the CCW debate started. In 2019, states were able to agree on the principle that the development and use of AWS do not take place in a legal vacuum, as 'international humanitarian law (IHL) continues to apply fully to all weapon systems, including the potential development and use of lethal autonomous weapon systems'.[2] Compliance with IHL—already important on its own—is also seen as a critical benchmark for assessing the acceptability of such weapon systems, as well as for addressing the potential humanitarian consequences of their use. However, in certain key respects, how and to what extent existing IHL rules provide limits on the development and use of AWS remains disputed.[3] A central element of that dispute concerns, in the language of the GGE process, what 'type and degree of human–machine interaction' is required to comply with IHL.[4]

This question touches on the nature of IHL rules. With respect to the conduct of hostilities, is IHL solely or primarily an effects-based regime—one that regulates military action which is expected to result, or does result, in unlawful consequences? If so, can militaries use any combination of humans and machines to undertake military action as long as the anticipated or actual effects are lawful? Alternatively, does IHL demand that humans must be involved in some, all or none of the tasks involved in military operations? Does IHL mandate that all attempts to kill combatants or destroy enemy installations must be volitional—for instance, must attacks involve evaluations and judgements by humans based on their personal knowledge and reflecting their intent? If so, what does that mean for the practical regulation of military action where

[1] Certain Conventional Weapons (CCW) Convention, Fifth Review Conference of the High Contracting Parties, 'Report of the 2016 informal meeting of experts on lethal autonomous weapons systems (LAWS)', CCW/CONF.V/2, 10 June 2016, Annex, para. 3.

[2] CCW Convention, Group of Governmental Experts (GGE) on Emerging Technologies in the Area of LAWS, 'Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', CCW/GGE.1/2019/3, 25 Sep. 2019, Annex IV, 'Guiding principles', para. (a).

[3] CCW Convention, GGE LAWS, 'Lethal Autonomous Weapons Systems: National Commentary—Australia', Sep. 2020; CCW Convention, GGE LAWS, 'UK commentary on the operationalisation of the laws guiding principles', Sep. 2020; CCW Convention, GGE LAWS, 'National commentary by the Kingdom of the Netherlands regarding the national interpretation and implementation of the Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System', Sep. 2020; CCW Convention, GGE LAWS, 'Commentary on the operationalization of the Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems at national level: Submitted by Japan', Sep. 2020; and CCW Convention, GGE LAWS, 'Working paper by the Bolivarian Republic of Venezuela on behalf of the Non-Aligned Movement (NAM) and other states parties to the Convention on Certain Conventional Weapons (CCW)', CCW/GGE.1/2020/WP.5, 14 Sep. 2020.

[4] The 'type and degree of human–machine interaction required, including elements of control and judgement' is part of the terminology that the GGE adopted in relation to the 'human element in the use of lethal force' involving AWS. Potentially related terms that states and experts have used in the debate include 'human control', 'human judgment', 'human involvement' and 'human supervision'. To sidestep the unresolved terminological and corresponding conceptual debate about the meaning of and interplay between these notions, this report draws on the GGE terminology. It uses 'type and degree of human–machine interaction' in a broad sense, as a catch-all notion to discuss elements of control and judgement throughout the development and use of AWS. See CCW Convention, GGE LAWS, 'Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', CCW/GGE.1/2019/3, 25 Sep. 2019, p. 5.

Box 1.1. A working definition of autonomous weapon systems

There is no internationally agreed definition of 'autonomous weapon systems'. This report defines them as weapons that, once activated, can identify and select targets and apply force to them without human intervention. The term autonomous weapon systems (AWS) is preferred to that of lethal autonomous weapon systems (LAWS), although the latter is used in the mandate of the GGE. AWS is preferred due to the interpretation, shared by the International Committee of the Red Cross and a number of states, that 'lethality' is a superfluous and misleading qualifier.[a] Lethality flows from how the weapon system is used rather than the way it is designed. Moreover, harm may result irrespective of whether or not death results.

AWS can come in many shapes and forms, but at the core they share several distinctive socio-technical features that are essential for the legal analysis. First, AWS function based on preprogrammed target profiles and technical indicators that can be recognized through the weapon's sensors and software. Second, since AWS are triggered to apply force partly by their environment of use (rather than a user's input), a decision to apply force can be made further in advance than with traditional weapons, based on assumptions about the circumstances that will prevail at the time of the attack. These features mean that AWS can be operated in 'communications denied' environments and permit faster reaction time in decisions to use force. However, these features also mean that those who configure and employ an AWS will not necessarily know the exact targets, location, timing and circumstances of the resulting use of force.

[a] Certain Conventional Weapons (CCW) Convention, Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS), Statements by Ireland and Germany, supported by South Africa, Austria, Pakistan and the United States, 4th Meeting, 1st Session, 22 Sep. 2020, UN Web TV; and CCW Convention, GGE LAWS, 'Agenda item 5(b): Characterization of the systems under consideration in order to promote a common understanding on concepts and characteristics relevant to the objectives and purposes of the Convention', Statement by the USA, 22. Sep. 2020.

Sources: Moyles, 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).

humans rely on machines to help decide aspects of who to attack, when, where and how? Does IHL permit some, all or none of the mandated processes and judgements to be performed by machines, including AWS? And if so, what are the implications for attributing military action that involves combinations of humans and machines, and for ensuring responsibility for IHL violations?

I. Human–machine interaction and IHL: An analytical framework

One way to answer the above questions is to explore key assumptions underlying states' positions on what they must do, what they can do and what they must not do in order to comply with IHL. To help illuminate the necessary or otherwise prudent type and degree of human–machine interaction warranted under IHL, these assumptions can be considered across four dimensions: personal, material, temporal and geographical.

The personal dimension

Of the *various people potentially involved or implicated in aspects of military action*, what type and degree of human–machine interaction does IHL require, permit or prohibit? For instance, as *subjects* of IHL, who is responsible for performing the legal obligations? Further, what can and should be demanded from IHL by the people who are the potential *objects* of military action, in terms of the type and degree of human–machine interaction?

The material dimension

Regarding *various forms and means of military action*, what type and degree of human–machine interaction do the substantive and procedural rules of IHL require, permit or prohibit? For instance, does IHL mandate, allow or bar legally required value judgements to be entrusted partly or fully to machine processes?

The temporal dimension

At the *various points in time in which aspects of military action may occur*, what type and degree of human–machine interaction does IHL require, permit or prohibit? For instance, what limits does IHL impose on how much time may lapse between the activation of an AWS and when its operation must be suspended or terminated?

The geographical dimension

Regarding the *various locations and types of environments in which aspects of military action may occur*, what type and degree of human–machine interaction does IHL require, permit or prohibit? For instance, what spatial limits (if any) does IHL impose on where AWS may travel and be used? And how do the characteristics of the environment of use impact the legal parameters of use of an AWS?

Each individual dimension merits attention, but it may be in their combination that the most central questions can actually be found. The combination of these different dimensions is essential to the determination of what would make an AWS unlawful per se, but also to the evaluation of how humans should be exercising their legal obligations in the development and use of an AWS: who must do what, when, where and how. These questions are also of critical importance with regard to ensuring state and individual responsibility for performing IHL obligations and imposing consequences in case of violations.

States have begun to address parts of these questions through statements and written commentaries on IHL, including with respect to the 'human element' concerning AWS. However, these questions deserve to be more systematically answered, as the extent to which they may be subject to different interpretations is critical to the determination of whether IHL rules need to be further clarified or developed to more concretely identify and respond to the legal issues posed by AWS.

II. Identifying the key legal questions for the development of the normative and operational framework

The idea that more substantial discussions on IHL compliance would be beneficial to the CCW process is already widely accepted. In 2019, Portugal notably proposed that the GGE produces a 'reference document compiling existing norms and principles of international law applicable to [lethal autonomous weapon systems] and identifying related good practices for producers, commanders and operators', receiving support from other states.[5] In 2020, the chair of the GGE also noted that such an exercise would be useful to identify potential 'legal gaps'.[6]

This report is designed to support that exercise. It aims to help states form and express their views on the legal provisions that already do, or should, govern the development and use of AWS, particularly with respect to the required type and degree of human–machine interaction. It does not take the form of legal advice. Rather, it is a mapping study that explores (*a*) what limits IHL already places on the development

[5] CCW Convention, GGE LAWS, 'CCW Group of Governmental Experts on Lethal Autonomous Weapons Systems: Statement by Portugal', 25 Mar. 2019; and CCW Convention, GGE LAWS, 'Operationalizing the guiding principles: A roadmap for the GGE on LAWS', Working paper submitted by Brazil, CCW/GGE.1/2020/WP.3, 6 Aug. 2020.

[6] In 2020 the chair of the GGE, Ambassador Karklins, and numerous states expressed the continued need to 'identify applicable law and possible gaps in the normative framework'. See CCW Convention, GGE LAWS, 'Commonalities in national commentaries on guiding principles', Working paper by the chair, Sep. 2020. This view is also expressed in the expert literature; see e.g. Lewis, D. A., 'Three pathways to secure greater respect for international law concerning war algorithms', Legal commentary, Harvard Law School Program on International Law and Armed Conflict, 2020, p. 11.

and use of AWS, and (*b*) what IHL demands from users of AWS to perform and satisfy IHL obligations, whether the obligations are of a state, an individual or both.[7] It aims to get to the heart of the nature of legal norms and frameworks of responsibility, and to reflect the ethical presumptions underlying states' interpretation and application of the law. The ultimate aim is to provide legal advisers, policymakers and diplomats with an analytical framework to help unpack a key question that is before the GGE: is existing IHL sufficient to govern AWS?

This report is the result of a one-year project at SIPRI on 'Autonomous Weapon Systems and International Humanitarian Law', which involved both desk research and a series of virtual expert discussions held in June 2020 under the Chatham House Rule. The report focuses on IHL, which is the primary legal framework agreed by states to regulate conduct in armed conflicts, although the use of AWS may also implicate other fields of law.

The report starts by identifying and listing relevant rules of IHL in relation to AWS and the limits they place on the development and use of weapons, means and methods of warfare, including AWS (chapter 2). The substantial discussion on the interpretation and application of IHL in the case of AWS is then organized around three categories of IHL provisions: the rules that prohibit or limit certain weapons, means and methods of warfare in armed conflicts (chapter 3); the rules on reviewing the legality of new weapons, means and methods of warfare before their use in armed conflict, and on providing legal advice when necessary to comply with IHL (chapter 4); and the legal frameworks governing state responsibility and individual criminal responsibility for IHL violations (chapter 5). These three chapters systematically review the legal questions posed by AWS across the four dimensions outlined above: personal, material, temporal and geographical. For each issue, the report seeks to map the extent to which states agree in their interpretations and application of the rule. The report concludes by summarizing the key findings of SIPRI's one-year project and presenting recommendations for states and non-governmental experts, particularly in relation to the discussions taking place in the GGE (chapter 6). Finally, the accompanying appendix features a list of legal questions that states and experts may use to guide and support future discussions and elaboration on IHL, AWS and human–machine interaction (appendix A).

[7] The term 'user' in this report refers to a person or group of persons who plan, decide on or carry out military action involving an AWS. Other literature on the topic of AWS and IHL variously refers to these persons as 'operators' or 'commanders'. 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. See Ekelhof, M. and Persi Paoli, G., 'The human element in decisions about the use of force', United Nations Institute for Disarmament Research (UNIDIR), 2020.

2. An overview of the limits on autonomous weapon systems under international humanitarian law

International humanitarian law (IHL), also known as the law of armed conflict, is a body of rules that sets out restrictions and prohibitions that must be complied with in armed conflicts, international as well as non-international. Many of the provisions laid down in IHL treaties are also reflected as general principles of international law and in rules of customary IHL, and are thereby binding for all parties to armed conflicts.[8] These fundamental and internationally recognized rules form the basis for the CCW discussions on regulating AWS.[9] While the wider IHL framework is broader than the scope of this report, this chapter identifies and outlines key rules of relevance with respect to AWS. By outlining the key provisions, it aims to provide an overview of the substantive and procedural limits that IHL already places on the design and employment of AWS. These provisions can be divided into three: the rules on weapons, means and methods of warfare (section I); the rules on legal reviews and advice (section II); and the frameworks governing state responsibility and individual criminal responsibility for IHL violations (section III).

I. Rules on weapons, means and methods of warfare

The rules of IHL that limit the development and use of AWS can be sorted into three categories: (*a*) rules prohibiting or restricting specific weapons, means and methods of warfare; (*b*) general prohibitions and restrictions on weapons, means and methods of warfare; and (*c*) general prohibitions and restrictions on the conduct of hostilities. While the first and second categories can be said to relate to whether a weapon, means or method of warfare is unlawful per se (also known as weapons law), the third category of IHL rules, in contrast, regulates how weapons, means and methods can be lawfully used (also known as targeting law). However, it would be incomplete to assess the legality of AWS without considering the Martens Clause, which stipulates that in cases not covered by IHL conventions, neither combatants nor civilians find themselves completely deprived of protection (see below).[10]

Specific and general rules prohibiting or restricting specific weapons, means and methods of warfare

Under IHL, any new weapon, means or method of warfare, including AWS, would be deemed inherently unlawful if it has one or more of the following characteristics:

1. The weapon (or its injury mechanism) is already prohibited by a specific treaty, such as the prohibition on the use of biological weapons, chemical weapons, poison or blinding lasers (see box 2.1).
2. The weapon is of a nature to cause superfluous injury or unnecessary suffering.[11]

[8] In 2005 the International Committee of the Red Cross (ICRC) compiled and published these rules in a widely accepted and referenced study on customary IHL: Henckaerts, J. and Doswald-Beck, L., *Customary International Humanitarian Law*, ICRC, vol. 1 (Cambridge University Press: Cambridge, 2005). This study is now available online as the ICRC, Customary IHL Database, <https://ihl-databases.icrc.org/customary-ihl/eng/docs/home>.

[9] As pointed out by Portugal, the CCW Convention does not distinguish between treaty and customary IHL. See CCW Convention, GGE LAWS, 'Commentaries by Portugal on "Operationalising all eleven guiding principles at a national level"', Aug. 2020.

[10] Koutroulis, V., 'Martens Clause', Oxford Bibliographies, 24 July 2013.

[11] Protocol I 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, Article 35(2); and ICRC, Customary

3. The weapon is by nature indiscriminate, that is, the weapon cannot be directed at a specific military objective or its effects cannot be limited as required by IHL, hence it is of a nature to strike military objectives and civilians or civilian objects without distinction.[12]

4. The weapon is intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment.[13]

General prohibitions and restrictions on the conduct of hostilities

In cases where weapons, means and methods of warfare are not deemed inherently unlawful, their use is still limited by general prohibitions and rules governing the conduct of hostilities. These prohibitions and restrictions have been established in several IHL treaties, notably the 1899 and 1907 Hague Conventions and the 1949 Geneva Conventions and their 1977 Additional Protocols, and are generally considered to constitute rules of customary IHL.[14] Under IHL, the use of a weapon, including AWS, would be unlawful in any of the following circumstances:

1. An attack by bombardment by any method 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.[15]

2. An 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 the effects of which cannot be limited as required by IHL.[16]

3. An attack that 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.[17]

IHL treaties also lay down fundamental rules that aim to protect the civilian population against the effects of hostilities. These rules oblige parties to armed conflict to comply with the principles of distinction, proportionality and precautions in attack. 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*, and between civilian objects and military objectives.[18] This overarching principle of distinction is operationalized in part by the rule that a party may direct an attack only against militarily active combatants and military

IHL Database, 'Rule 70. Weapons of a nature to cause superfluous injury or unnecessary suffering'.

[12] The use of indiscriminate weapons is prohibited in all circumstances under customary IHL, c.f. the IHL prohibition of indiscriminate attacks, see Additional Protocol I (note 11), Article 51(4)(b) and (c); and ICRC, Customary IHL Database, 'Rule 71. Weapons that are by nature indiscriminate'.

[13] Additional Protocol I (note 11), articles 35(3) and 55; and ICRC, Customary IHL Database, 'Rule 45. Causing serious damage to the natural environment'.

[14] ICRC, Customary IHL Database (note 8), rules 1–86.

[15] Additional Protocol I (note 11), Article 51(5)(a); and ICRC, Customary IHL Database, 'Rule 13. Area bombardment'.

[16] Additional Protocol I (note 11), Article 51(4)(a); and ICRC, Customary IHL Database, 'Rule 13. Indiscriminate attacks'.

[17] Additional Protocol I (note 11), Article 51(5)(b); and ICRC, Customary IHL Database, 'Rule 14. Proportionality in attack'.

[18] Additional Protocol I (note 11), Article 48; and ICRC, Customary IHL Database (note 8), rules 1, 7.

objectives, not against civilians and civilian objects (unless, and for such time as, civilians take a direct part in hostilities).[19]

The principle of proportionality implicitly recognizes that civilians and civilian objects may be affected incidentally by an attack that is directed against a lawful military objective. Under this rule, it is unlawful to conduct an attack that may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be expected to be excessive in relation to the concrete and direct military advantage anticipated.[20]

The principle of precautions includes two interrelated components, one concerning military operations and the other concerning attacks. First, in the conduct of military operations, IHL obliges parties to take constant care to spare the civilian population, civilians and civilian objects.[21] Second, IHL obliges parties to take several sets of precautions regarding specific attacks. In particular, the obligation to take precautions in attacks entails a requirement to: (*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 to avoid, and in any event minimize, incidental loss of civilian life, injury to civilians and damage to civilian objects; (*c*) refrain from deciding to launch an attack if it may be expected to violate the principle of proportionality; and (*d*) cancel or suspend an attack if it becomes apparent that the objective is not a military one, that the objective is subject to special protection, or that the attack may be expected to violate the principle of proportionality.[22]

IHL also places restrictions on how AWS may or may not be used in other contexts than the conduct of hostilities. Those restrictions include rules that limit whether and how an AWS may be used for guarding and transporting detainees, crowd control and ensuring public security in occupied territory.[23]

The Martens Clause

The Martens Clause, in its 1977 formulation in Additional Protocol I to the 1949 Geneva Conventions, states that: 'In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and the dictates of public conscience.'[24]

The nature, status and content of the Martens Clause is subject to 'enormous variations' in interpretations.[25] Disagreement has arisen between states and experts about whether it constitutes customary law, amounts to an independent source of law, or merely provides moral guidelines.[26] However, the position has been put forward that, at a minimum, 'not everything that is not explicitly prohibited can be said to be

[19] Additional Protocol I (note 11), articles 51(2) and 52(1); and ICRC, Customary IHL Database (note 8), rules 1, 7.

[20] Additional Protocol I (note 11), Article 51(5)(b); and ICRC, Rule 14 (note 17).

[21] Additional Protocol I (note 11), Article 57(1).

[22] Additional Protocol I (note 11), Article 57(2)(a) and (b); and ICRC, Customary IHL Database (note 8), rules 15–19.

[23] See e.g. ICRC, Customary IHL Database (note 8), rules 51, 75, 87, 123, 130.

[24] The clause first appeared in the 1899 Hague Conventions and is repeated in all four Geneva Conventions, latest in Additional Protocol I of 1977. Additional Protocol I (note 11), Article 1(2).

[25] Cassese, A., 'The Martens Clause: Half a loaf or simply pie in the sky?', *European Journal of International Law*, vol. 11, no. 1 (2000); Doswald-Beck, L., 'International humanitarian law and the Advisory Opinion of the International Court of Justice on the legality of the threat or use of nuclear weapons', *International Review of the Red Cross*, no. 316 (Feb. 1997); Ticehurst, R., 'The Martens Clause and the Laws of Armed Conflict', *International Review of the Red Cross*, no. 317 (Apr. 1997); and CCW Convention, GGE LAWS, 'Joint "commentary" on guiding principles A, B, C and D: Submitted by Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico and New Zealand', Sep. 2020.

[26] As expressed by Doswald-Beck (note 25). See also Ticehurst (note 25); ICRC, 'Ethics and autonomous weapon systems: An ethical basis for human control?', 3 Apr. 2018, p. 6; and Evans, T. D., 'At war with the robots: Autonomous weapon systems and the Martens Clause', *Hofstra Law Review*, vol. 4, no. 3 (2013), p. 716.

Box 2.1. Existing weapon prohibitions

The prohibition on:

- the use of poison or poisoned weapons.[a]
- the use of biological weapons.[b]
- the use of chemical weapons[c] and the use of riot control agents as a method of warfare.[d]
- the use, under certain circumstances, of automatic submarine contact mines.[e]
- under certain conditions, the use of herbicides as a method of warfare.[f]
- the use of bullets that expand or flatten easily in the human body.[g]
- the anti-personnel use of bullets that explode within the human body.[h]
- the use of weapons the primary effect of which is to injure by fragments that are not detectable by x-ray in the human body.[i]
- the use of booby traps that are in any way attached to or associated with objects or persons entitled to special protection under international humanitarian law or with objects that are likely to attract civilians.[j]
- the use of anti-personnel mines.[k]
- the use of cluster munitions.[l]
- under certain circumstances, the anti-personnel use of incendiary weapons.[m]
- the use of laser weapons that are specifically designed, as their sole combat function or as one of their combat functions, to cause permanent blindness to unenhanced vision.[n]
- the use of nuclear weapons.[o]

[a] Protocol for the Prohibition of the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare, Geneva, signed 17 June 1925, entered into force 8 Feb. 1928; and International Committee of the Red Cross (ICRC), Customary IHL Database, 'Rule 72. Poison and poisoned weapons'.

[b] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (Biological and Toxin Weapons Convention, BWC), opened for signature 10 Apr. 1972, entered into force 26 Mar. 1975; and ICRC, Customary IHL Database, 'Rule 73. Biological weapons'.

[c] Protocol for the Prohibition of the Use of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare (note a); Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction, opened for signature 13 Jan. 1993, entered into force 29 Apr. 1997; and ICRC, Customary IHL Database, 'Rule 74. Chemical weapons'.

[d] ICRC, Customary IHL Database, 'Rule 75. Riot Control Agents'.

[e] Convention (VIII) relative to the Laying of Automatic Submarine Contact Mines, The Hague, signed 18 Oct. 1907, entered into force 26 Jan. 1910.

[f] ICRC, Customary IHL Database, 'Rule 76. Herbicides'.

[g] Declaration (3) concerning the Prohibition of Using Bullets which Expand or Flatten Easily in the Human Body, The Hague, 29 July 1899; and ICRC, Customary IHL Database, 'Rule 77. Expanding bullets'.

[h] Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight, Saint Petersburg, 29 Nov./11 Dec. 1868; and ICRC, Customary IHL Database, 'Rule 78. Exploding bullets'.

[i] Protocol on Non-Detectable Fragments (Protocol I) to the CCW Convention, opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983; and ICRC, Customary IHL Database, 'Rule 79. Weapons primarily injuring by non-detectable fragments'.

[j] Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (Protocol II) to the CCW Convention, opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983, as amended 3 May 1996; and ICRC, Customary IHL Database, 'Rule 80. Booby-traps'.

[k] Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction, opened for signature 3–4 Dec. 1997, entered into force 1 Mar. 1999.

[l] Convention on Cluster Munitions, opened for signature 3 Dec. 2008, entered into force 1 Aug. 2010.

[m] Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons (Protocol III) to the CCW Convention, opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983; and ICRC, Customary IHL Database, rules 84–85.

[n] Protocol on Blinding Laser Weapons (Protocol IV) to the CCW Convention, issued 13 Oct. 1995, entered into force 30 July 1998; and ICRC, Customary IHL Database, 'Rule 86. Blinding laser weapons'.

[o] Treaty on the Prohibition of Nuclear Weapons (TPNW), opened for signature 20 Sep. 2017, entered into force 22 Jan. 2021.

legal if it would run counter the principles' in the Martens Clause.[27] At least from that standpoint, the clause 'may be said to imply positive obligations where contemplated military action would result in untenable humanitarian consequences'.[28]

II. Rules requiring legal reviews and advice

IHL obliges states parties both to conduct legal reviews of new weapons, means and methods of warfare and to make legal advisers available, when necessary, to advise certain military commanders on the application of IHL and on the appropriate instruction to be given to the armed forces. These obligations may be reflected in customary IHL as well.

The obligation to conduct legal reviews of new weapons, means and methods of warfare

Article 36 of Additional Protocol I to the 1949 Geneva Conventions places High Contracting Parties under the following obligation: 'In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.'[29] This obligation derives from the basic rule, set out in Article 35, that the right of states to choose means and methods of warfare is not unlimited and from the general legal principle mandating performance of that and other treaty obligations in good faith.[30]

The question of whether the obligation to conduct legal reviews is part of customary international law and therefore is applicable to states that are not parties to Additional Protocol I remains debated.[31] However, it is regarded as good practice by most states, serving as an important measure to help ensure that a state's armed forces can and will conduct hostilities in accordance with its international obligations.[32]

Article 36 does not provide concrete guidance on how legal reviews should be conducted. What falls within the material scope of a review and what methodologies should be used have been discussed extensively.[33] Nevertheless, it is commonly accepted that the scope of Article 36 is broad and could cover weapons of all types, regardless of lethality.[34] It also covers all new weapons, regardless of whether they are developed for further research and experimentation or procured 'off the shelf' from other states.[35] In terms of methodology, it is also accepted that although there cannot be a one-size-fits-all approach to legal reviews—since states have different needs, as well as different human and financial resources to conduct reviews—there can be elements of best practice.[36] These include conducting reviews as early

[27] CCW Convention, GGE LAWS, 'A "compliance-based" approach to autonomous weapon systems', Working paper submitted by Switzerland', 10 Nov. 2017, p. 4.

[28] CCW Convention (note 27).

[29] Additional Protocol I (note 11), Article 36.

[30] Commentary of 1987 to Additional Protocol I, 'New weapons', Commentary on Article 36, para. 1466, p. 423.

[31] Jevglevskaja, N., 'Weapons review obligation under customary international law', *International Law Studies*, vol. 94, no. 186 (2018).

[32] E.g. the USA first established a legal review mechanism in 1974, three years before the adoption of Additional Protocol I; see US Department of Defense, Directive 5000.01, 'The Defense Acquisition System', 9 Sep. 2020. As of 2021, however, relatively few states are known to regularly conduct legal reviews; see Boulanin, V. and Verbruggen, M., 'SIPRI compendium on Article 36 reviews', SIPRI Background Paper, Dec. 2017.

[33] ICRC, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977* (ICRC: Geneva, Jan. 2006); and Boulanin and Verbruggen (note 32), p. 9.

[34] Boulanin and Verbruggen (note 32).

[35] ICRC (note 33), pp. 9–10.

[36] ICRC (note 33); and McClelland, J., 'The review of weapons in accordance with Article 36 of Additional Protocol I', *International Review of the Red Cross*, vol. 85, no. 850 (June 2003), p. 414.

as possible, using a multidisciplinary approach, relying on empirical evidence and, if possible, conducting independent testing to assess a weapon's performance and the risks associated with its use.

The obligation to provide legal advice

Article 82 of Additional Protocol I to the 1949 Geneva Conventions requires the following in terms of legal advice: 'The High Contracting Parties at all times, and the Parties to the conflict in time of armed conflict, shall ensure that legal advisers are available, when necessary, to advise military commanders at the appropriate level on the application of the Conventions and this Protocol and on the appropriate instruction to be given to the armed forces on this subject.'[37] The adoption of this article was prompted by the increasingly complex nature of IHL and the fact that some legal assessments could be difficult to make.

Ensuring that legal advisers are available, when necessary, is a practical measure to help a state perform its obligation to respect and to ensure respect for IHL, which can be seen as flowing from the general principle mandating the performance of legal obligations in good faith.[38] The material scope of the obligation is generally understood as having two requirements: (*a*) to have legal advisers available to provide training and educational materials on IHL for members of the armed forces; and (*b*) to dispense direct legal advice to commanders during the conduct of military operations.

III. Frameworks for state responsibility and individual criminal responsibility

State responsibility and individual criminal responsibility are fundamental institutions of international law. They result from the legal personality of each state and individual under IHL and from the fact that states and individuals bear IHL obligations.[39] Depending on the author, nature and character of the breach, one or more states, one or more individuals, or a combination of state(s) and individual(s) may bear responsibility for a particular IHL violation. Thus, responsibility can be attributed, discerned and scrutinized through multiple frameworks simultaneously and the same conduct may constitute multiple violations, a single violation or no violation. The legal frameworks for individual and state responsibility help to structure compliance with IHL rules, to avoid impunity for IHL violations and to sustain confidence in the efficacy of the IHL regime.

State responsibility

Every state must respect and ensure respect for IHL, including by its armed forces and other persons or groups acting in fact on its instructions or under its direction or control.[40] Further, every state bears responsibility for internationally wrongful acts attributable to it.[41] The concept of state responsibility is fundamental for compliance and respect for international law, including IHL, and the legal doctrines

[37] Additional Protocol I (note 11), Article 82.

[38] ICRC, Customary IHL Database, 'Rule 141. Legal advisers for armed forces'.

[39] The principles of state responsibility and individual criminal responsibility are long-standing rules in international law. For state responsibility, see e.g. Geneva Convention (IV) Relative to the Protection of Civilian Persons in Time of War, opened for signature on 12 Aug. 1949, entered into force 21 Oct. 1950, Article 1; and ICRC, Customary IHL Database, 'Rule 139. Respect for international humanitarian law'. For individual criminal responsibility, see e.g. Additional Protocol I (note 11), Article; ICRC, Customary IHL Database, 'Rule 102. Individual criminal responsibility'; and ICRC, Customary IHL Database, 'Rule 151. Individual responsibility'.

[40] ICRC, Rule 139 (note 39).

[41] ICRC, Customary IHL Database, 'Rule 149. Responsibility for violations of international humanitarian law'.

underpinning the framework for state responsibility can be traced to the beginning of modern international law. IHL defines the scope and content of 'primary' rules. The legal institution of state responsibility sets out and structures the 'secondary' rules regarding breaches of those primary IHL rules.

The rules on state responsibility have been compiled and codified in the International Law Commission's widely recognized articles on 'Responsibility of States for Internationally Wrongful Acts' (draft articles).[42] According to the draft articles, four elements are required in order for state responsibility to arise. First, the conduct has to be attributable to the state. States are abstract entities and act through human agents. Human agents whose conduct is attributable to the state include members of the state's armed forces, persons or entities whom the state empowers to exercise elements of governmental authority, persons or groups acting in fact on the state's instructions or under its direction or control, and private persons or groups whose conduct a state acknowledges and adopts as its own.[43]

Second, there has to be a breach of one or more of the state's international obligations, whether general or specific, either through commission or omission. The rules on state responsibility do not define the content of the primary obligations—IHL does. As part of their general obligations, under Common Article 1 of the 1949 Geneva Conventions, states are obliged to respect and to ensure respect for IHL instruments.[44]

Third, for the act to be internationally wrongful, the wrongfulness of the conduct must not be precluded by any of the recognized excuses or justifications. The International Law Commission's draft articles enumerate six circumstances that preclude the wrongfulness of what would otherwise be considered a violation by a state of a primary rule of international law: consent, self-defence, countermeasures, *force majeure*, distress and necessity.[45] However, none of these circumstances authorizes or excuses any derogation from a peremptory norm of general international law.[46] Arguably, at least the cardinal principles of IHL—distinction, proportionality and precautions—constitute such peremptory norms.[47]

Fourth and finally, certain legal consequences flow from the commission of an internationally wrongful act. These consequences mainly include the obligations to cease the act (if it is continuing), to offer appropriate assurances and guarantees of non-repetition (if circumstances so require), and to make full reparation for the injury caused.[48]

Individual criminal responsibility

Individuals bear criminal responsibility for war crimes and other international crimes they commit. Besides grave breaches of IHL that amount to war crimes, other international crimes include crimes against humanity, the crime of genocide and the crime of aggression.[49] The international legal contours concerning what conduct and

[42] International Law Commission, 'Responsibility of states for internationally wrongful acts', Draft articles, Text adopted by the Commission at its fifty-third session, 23 Apr.–1 June and 2 July–10 Aug. 2001, subsequently adopted by the United Nations General Assembly through Resolution A/RES/56/83 of 12 Dec. 2001.

[43] ICRC, Rule 149 (note 41). Also expressed in Convention (IV) Respecting the Laws and Customs of War on Land and its Annex: Regulations Concerning the Laws and Customs of War on Land, The Hague, 18 Oct. 1907, Article 3; and International Law Commission (note 42), Chapter II, articles 4–11.

[44] Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 12 Aug. 1949, Article 1.

[45] International Law Commission (note 42).

[46] International Law Commission (note 42), p. 85.

[47] Amoroso, D. and Benedetta, G., 'Who is to blame for autonomous weapons systems' misdoings?', eds E. Carpanelli and N. Lazzerini, *Use and Misuse of New Technologies* (Springer Press, 2019), p. 224.

[48] International Law Commission (note 42), articles 30 and 31.

[49] Grave breaches are defined, among other places, in Geneva Convention (IV) (note 39); and Additional Protocol I (note 11), articles 85, 86, 147. War crimes are defined in the Rome Statute of the International Criminal Court, opened

circumstances give rise to a war crime may be found in interactions between IHL and international criminal law (ICL).[50] Although this report is centred around IHL, the applicability of the ICL framework, as set out in the Rome Statute, is also taken into consideration. The relevance of this framework, in relation to AWS and in general, has been established and confirmed by multiple actors, and it would be incomplete to consider the rules on individual responsibility without considering at least the Rome Statute.[51]

Based on both frameworks, four elements form the basis of individual criminal responsibility, and all four need to be established for a war crime to arise. First, the conduct must be attributable to one or more agents engaging in conduct that relates to an armed conflict. Artificial agents, such as machines, cannot bear individual criminal responsibility under the Rome Statute of the International Criminal Court.[52]

Second, the conduct, circumstances or consequences must constitute a serious violation of IHL amounting to a grave breach or war crime, such as the wilful killing of a protected person. This requirement relates to the material (or objective) element of the crime. The scope and content of the material elements of war crimes (including grave breaches) originate in IHL.

Third, the requisite mental element (also known as *mens rea*) must be established. Depending on the applicable law, this aspect, which relates to the subjective element of the war crime, may concern whether the alleged perpetrator acted 'wilfully' (under IHL instruments) or whether the alleged perpetrator had or lacked sufficient 'knowledge and intent' concerning the conduct, circumstances or consequences (under the Rome Statute).[53]

Fourth and finally, criminal responsibility and liability for punishment arise only if the proscribed conduct was carried out through one of the recognized modes of individual criminal responsibility. Under the Rome Statute, for example, modes of responsibility include committing a war crime; ordering, soliciting or inducing the commission of a war crime; and aiding, abetting or otherwise assisting in the commission of a war crime.[54] Further, pursuant to the Rome Statute, a commander or other superior is responsible under certain circumstances for war crimes committed by forces or subordinates under their effective command or authority and control, as a result of their failure to properly exercise control over such forces or subordinates.[55]

IV. Conclusion: The development and use of AWS is not unlimited

The development and use of technologies in the area of AWS do not take place in a legal vacuum. States are obliged to respect and ensure respect for IHL rules applicable to AWS. Those rules relate to specific and general prohibitions and restrictions on the employment of AWS as weapons, means and methods of warfare. The rules also include the obligations to conduct legal reviews of AWS and to provide legal advice,

for signature at Rome on 17 July 1998 and at New York on 18 Oct. 1998, entered into force 1 July 2002; see International Criminal Court, *Rome Statute of the International Criminal Court* (last amended 2010), 2011, Article 8. See also ICRC, Rule 151 (note 39).

[50] International criminal law is guided by the framework set out in the Rome Statute (note 49).

[51] See e.g. CCW Convention, GGE LAWS, 'Switzerland's commentary on operationalizing the guiding principles at a national level, as requested by the chair of the 2020 Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW)', Aug. 2020; CCW Convention, Commentaries by Portugal (note 9); and CCW Convention, GGE LAWS, 'Chairperson's summary', Working paper, CCW/GGE.1/2020/WP.7, 19. Apr. 2021.

[52] Rome Statute (note 49), Article 25(1).

[53] Geneva Convention (IV) (note 39); Additional Protocol I (note 11), Article 85(3)(4); and Rome Statute (note 49), Article 30.

[54] Rome Statute (note 49), Article 25(3)(a)–(c); Geneva Convention (I) (note 44), Article 49; Additional Protocol I (note 11), Article 86; and ICRC, Rule 151 (note 39).

[55] Rome Statute (note 49), Article 28.

including instructions to armed forces before and during armed conflicts, relating to AWS. Finally, states are required to take measures to suppress violations of IHL involving AWS.[56]

Arguably, for the use of an AWS to reflect respect for IHL, at least three conditions must be met. First, it must be possible to reliably *foresee* whether the operation, performance or consequences of the AWS, in its anticipated circumstances and context of use: (*a*) are specifically prohibited in an IHL treaty or in customary IHL; (*b*) would cause superfluous injury or unnecessary suffering; (*c*) would have indiscriminate effects; or (*d*) would be intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment. Second, it must be possible to *administer* the operation, performance and effects of the AWS so as to ensure respect for the rules governing the conduct of hostilities, notably the principles of distinction, proportionality and precautions. Third, it must be possible to *trace* the operation, performance and effects of the AWS back to the specific individuals involved in the system's employment.[57]

Questions remain about how these conditions can be secured and, in particular, what type and degree of human–machine interaction are required. When, where, how and to what extent may humans rely on the technologies underlying AWS to foresee, administer and trace the system as required by IHL? Specifically, what are the implications for compliance with IHL rules on the means and methods of warfare, conducting legal reviews and providing legal advice? What about the ability to hold individual and states responsible for IHL violations? These questions are addressed in the subsequent chapters.

[56] Malik, S., 'Autonomous weapon systems: The possibility and probability of accountability', *Wisconsin International Law Journal*, vol. 35, no. 3 (2018), p. 63.

[57] This is also reflected in Lewis (note 6), p. 7; and Verdiesen, I., Santoni de Sio, F. and Dignum, V., 'Accountability and control over autonomous weapon systems: A framework for comprehensive human oversight', *Minds and Machines*, vol. 31 (2021), p. 11.

3. Key issues concerning the rules on weapons, means and methods of warfare

How the rules on weapons, means and methods of warfare should be interpreted and applied has always been a matter of debate among states and experts.[58] AWS require a revisitation of old and unresolved issues, such as what constitutes a 'specific' military objective or what constitutes 'feasible precautions'.[59] AWS also bring to the surface new and unique questions of interpretation and application. These flow from the impact autonomy could have on the way humans perform obligations under IHL, across the four dimensions outlined in the introduction: *Who* may or must be responsible for respecting IHL provisions governing an AWS, and what may the persons protected by those provisions demand of those people (personal dimension, section I); what is the *content and nature* of IHL provisions governing an AWS, and what are the circumstances in which those rules apply (material dimension, section II); at what *points in time* may or must IHL provisions governing an AWS be respected (temporal dimension, section III); and in relation to what *locations* may or must IHL provisions governing an AWS be respected (geographical dimension, section IV)?

I. The personal dimension: Who is responsible for respecting IHL provisions and in relation to whom?

Advances in autonomy in weapon systems are bound to transform the way humans interact with the battlefield and make decisions about the use of force. The first dimension in which autonomy introduces transformation and poses novel questions about the interpretation and application of IHL relates to the people involved in developing, administering and assessing AWS and the people affected by the use of AWS. The use of AWS invites a revisitation of the fundamental assumptions about who is responsible for respecting IHL provisions and in relation to whom.

Who may or must respect IHL provisions? Natural versus artificial agents

The first and most fundamental interpretative question that AWS raise is whether humans are the only valid agents for the exercise and implementation of IHL rules on the conduct of hostilities, or whether artificial agents, including AWS, may act as agents for the partial or full exercise and implementation of those rules.[60] This question, which in many ways ignited the debate on autonomous weapons, is critical as it frames the entire legal debate on AWS.[61] It conditions the type and degree of human–machine interaction that is needed to ensure lawful use of AWS, it has implications for how AWS need to be reviewed prior to employment and, finally, it affects how responsibility may be incurred for an IHL violation involving AWS.

[58] See e.g. Saul, B. and Akande, D. (eds), *The Oxford Guide to International Humanitarian Law* (Oxford University Press: Oxford, 2020); Practice relating to ICRC, Customary IHL Database, 'Rule 17. Choice of means and methods of warfare'; Schmitt, M. and Schauss, M., 'Uncertainty in the law of targeting: Towards a cognitive framework', *Harvard National Security Journal*, vol. 10, no. 1 (2019); and Boothby, W. H., *Weapons and the Law of Armed Conflict*, 2nd edn (Oxford University Press: Oxford, 2016).

[59] The notion of military objects is expressed, among others, in Additional Protocol I (note 11), Article 52(2); and ICRC, Customary IHL Database, 'Rule 8. Definition of military objectives'. The obligation to take feasible precautions is mentioned, among others, in Additional Protocol I (note 11), Article 57(1).

[60] Lewis (note 6), p. 11.

[61] Human Rights Watch and the International Human Rights Clinic, the Human Rights Program at Harvard Law School, *Losing Humanity: The Case Against Killer Robots* (Human Rights Watch: 2012); and ICRC, Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects', Geneva, 26–28 Mar. 2014.

Regarding responsibility for war crimes, the Rome Statute provides a clear answer, addressing only 'natural persons'. However, in terms of respecting IHL provisions more broadly, IHL instruments are not as clear on the matter.[62] Three interpretations have been proposed within the context of the CCW discussion, which themselves reflect different presumptions about the nature and character of IHL and what IHL rules demand. At one end of the spectrum, some states and experts believe that IHL is a human-centric framework, which was 'undoubtedly conceived with States and individual humans as agents for the exercise and implementation of the resulting rights and obligations in mind'.[63] According to this approach, the responsibility to comply with IHL rules, notably the rules on distinction, proportionality and precautions, resides solely with natural (human) agents. Humans, therefore, need to be the ones acting as agents for the exercise and implementation of IHL rules, including by making the decisions and judgements demanded by these rules. This understanding may be seen as reflecting a deontological approach to the nature and character of IHL, by placing particular importance on ensuring the non-delegable role of humans in the deliberative processes through which IHL-regulated effects are produced.

At the other end of the spectrum, some states and experts believe that IHL's first and foremost concern is to avoid or at least minimize unlawful effects. Within this approach, the nature of the agents exercising and implementing IHL rules is less important than the avoidance of unlawful effects and the maximization of relatively humanitarian effects. According to this reasoning, the question of whether responsibility may reside with an 'artificial' agent, namely an AWS, is primarily a technical issue that concerns what effects can or cannot be achieved with technology. In theory, at least some of the agency necessary to exercise and implement IHL obligations may reside in machine processes. This interpretation reflects a more utilitarian approach to the nature and character of IHL rules governing AWS, emphasizing the net humanitarian effect with less focus on whether that effect was produced by reliance on human agency.[64]

The third approach, which falls somewhere between these two poles, presumes that humans alone are responsible for exercising and implementing legal agency, but that humans may—and perhaps ought to—rely on technical systems when that reliance is more likely to result in avoiding or at least minimizing unlawful effects and maximizing humanitarian effects.[65]

Nevertheless, in the list of 11 guiding principles that the GGE adopted in 2019, Guiding Principle (d) states that: 'Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system.'[66] This guiding principle makes clear that it is the user of the weapon that is responsible and accountable for complying with IHL, not the weapon itself. It is to be noted, however, that the principle is concerned with human *responsibility* rather

[62] Rome Statute (note 49), Article 25.

[63] CCW Convention (note 27); and CCW Convention, Working paper by Venezuela (note 3).

[64] See e.g. Lewis, L. 'Killer robots reconsidered: Could AI weapons actually cut collateral damage?', *Bulletin of Atomic Scientists*, 10 Jan. 2020; CCW Convention, 'Israel considerations on the operationalization of the eleven guiding principles adopted by the Group of Governmental Experts', Aug. 2020; CCW Convention, GGE LAWS, 'Working paper of the Russian Federation: National implementation of the guiding principles on emerging technologies in the area of lethal autonomous weapons systems', Unofficial translation, Sep. 2020; and CCW Convention, GGE LAWS, 'US commentaries on the guiding principles', 1 Sep. 2020.

[65] See e.g. CCW Convention, GGE LAWS, 'German commentary on operationalizing all eleven guiding principles at a national level as requested by the chair of the 2020 Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW)', 24 June 2020; CCW Convention, Commentary by Australia (note 3); CCW Convention, GGE LAWS, 'United Kingdom expert paper: The human role in autonomous warfare', CCW/GGE.1/2020/WP.6, 18. Nov. 2020; and Work, R., 'Principles for the combat employment of weapon systems with autonomous functionalities', Center for a New American Security, 28. Apr. 2021.

[66] Guiding Principle (d) quoted from CCW Convention (note 2), Annex IV.

than human *agency*. It does not solve the question of whether, how and to what extent humans may rely on autonomous systems when complying with IHL provisions—this issue remains debated (see below).

Who bears responsibility? Individual versus systemic responsibility

A common feature of modern warfare is that the decision-making process leading to the use of force may be distributed across a large number of actors at the strategic, operational and tactical levels, before and during an attack.[67] The case of AWS pushes this trend to an extreme, as the preprogrammed nature of an AWS supposes that its effects will not only be determined by decisions made by multiple people along the military command-and-control chain (users at different levels and weapon operators) but also by engineers and technicians during the development phase. In this context, the second interpretative question that AWS raise is whether IHL demands that a single person be responsible for the decision to employ an AWS and the resulting effects or whether that responsibility may reside with multiple people. If the latter, how then do individual contributions to the ultimate decision to employ an AWS and to administer its operation interact in a systemic multi-agent model (see below)?

As it currently exists, IHL does not necessarily provide a clear answer. Yet this question is critical, not only for the framing of the human–machine interaction debate (especially who needs to exercise and implement legal agency), but also for the debate on state and individual responsibility (who can be held accountable for IHL violations; see the detailed discussion in chapter 5). Broadly speaking, two interpretations have been put forward in the GGE debate.

One interpretation is that the responsibility for deciding to employ and for administering an AWS needs to reside with a single person—typically framed as a commander. From this perspective, the person who authorizes the activation and launch of an AWS and who administers its operation is responsible for exercising and implementing legal agency relating to the AWS, including by making the value judgements demanded by the IHL rules governing the conduct of hostilities. The commander is the ultimate decision maker, even if the judgements may be partly implemented by other people, such as weapon operators, and built on instructions provided by higher levels of command or even the automated systems. Unlike the preceding agents, the commander is in a better position to make the context-dependent, legal assessments required to exercise IHL obligations.[68]

Another interpretation is that the responsibility for exercising and implementing legal agency, including by making the evaluations demanded by the principles of distinction, proportionality and precautions, may reside with multiple people—and possibly systems of people—in the command-and-control chain.[69] From this perspective, all members of the command-and-control chain who contribute to the targeting process are seen as exercising and implementing legal agency. Thus, the responsibility for complying with the principles of distinction, proportionality and

[67] Ekelhof and Persi Paoli (note 7); Bo, M., 'The human–weapon relationship in the age of autonomous weapons and the attribution of criminal responsibility for war crimes', Conference paper presented at We Robot 2019, University of Miami Law School, Apr. 2020, p. 10; and Amoroso and Benedetta (note 47).

[68] CCW Convention (note 27); CCW Convention, Commentary by Germany (note 65); CCW Convention, GGE LAWS, 'Commentaries on national implementation of the guiding principles on LAWS', Commentary by Spain, Aug. 2020; and Henderson, I., Keane, P. and Liddy, J., 'Remote and autonomous warfare systems: Precautions in attack and individual accountability', ed. J. D. Ohlin, *Research Handbook on Remote Warfare* (Edward Elgar Press: Cheltenham, UK, and Northampton, MA, USA, 2016), p. 23.

[69] Ekelhof and Persi Paoli (note 7); Schulzke, M., 'Autonomous weapons and distributed responsibility', *Philosophy and Technology*, vol. 26 (June 2013); CCW Convention, Considerations by Israel (note 64); CCW Convention, GGE LAWS, 'Reflections by the Bolivarian Republic of Venezuela on emerging technologies in the area of lethal autonomous weapons systems (LAWS) and the mandate of the group of governmental experts (GGE)', Sep. 2020; and CCW Convention, Commentary by Germany (note 65).

precautions may be shared across multiple human agents. The question of whether this systemic multi-agent model complicates the task of ascribing individual responsibility for alleged criminal violations remains debated (see chapter 5). Critics argue that it implies a diffusion of responsibility that could open the door for blame avoidance or lessening: if everyone is partly responsible, no one is fully responsible. On the other hand, proponents of this model argue that multi-layered decision-making structures are already a reality of modern military decision making and operations, and that they are not problematic as long as the command-and-control chain is clear about how individual decisions interact (who decides over who, what and when) and contains the possibility to foresee the effects of an AWS, to continuously administer an AWS and to trace who made what decisions regarding the use of an AWS.[70]

What can other subjects of IHL demand of the users of AWS?

The third interpretative question focuses on the perspective of those subject to an attack, as well as bystanders who might be affected by an attack involving an AWS. What does IHL say about what claims such people can make about the type and degree of human–machine interaction that a belligerent should use involving an AWS? This evaluation also needs to consider what and who can be targeted, as different potential targets may demand different things from IHL.

The question of what people protected by IHL can demand of the users of AWS has been raised by some states and civil society organizations, including with reference to the content and legal status of the Martens Clause.[71] Here, the claim is made that any parties to Additional Protocol I can demand that decisions to use force remain under direct human control, even if the person making the demand is not a party to the conflict or even an object of the attack. Notwithstanding any other rules of IHL, it is argued that the principles of humanity and the dictates of public conscience provide a legal basis to demand that the users of AWS exercise human judgement and agency in decisions and operations involving the use of force. It is further argued that the principles set out in the Martens Clause should be regarded as peremptory norms of international law (also called *jus cogens*), and that they thereby impose obligations on states that prevail over any other rules and cannot be modified merely by the will of the parties.

Human Rights Watch (HRW) has played a substantial role in the elaboration of the claim that AWS, if not under sufficient human control, would violate the principles of humanity and the dictates of the public conscience.[72] HRW argues that compliance with the principle of humanity requires the ability to feel compassion and emotions and to be guided by ethical standards, which, they assert, AWS would lack.[73] Further, it claims that AWS contradict the dictates of public conscience, as a number of governments and experts and a sample of the general public have expressed moral discomfort vis-à-vis the possibility that the decision to use force, particularly against people, could be delegated to machine processes, regardless of how sophisticated the technologies might be.[74]

[70] See e.g. Australia's position paper on 'systems of control': CCW Convention, GGE LAWS, 'Australia's systems of control and application for autonomous weapon systems', 20 Mar. 2019, CCW/GGE1./2019/WP2.

[71] CCW Convention, Joint commentary (note 25); ICRC (note 26); CCW Convention, UK expert paper (note 65); and ICRC, 'ICRC position on autonomous weapon systems', 12 May 2021.

[72] Human Rights Watch, 'Heed the call: A moral and legal imperative to ban killer robots', 2018.

[73] Human Rights Watch (note 72), pp. 2, 19.

[74] Human Rights Watch (note 72), p. 3; CCW Convention, GGE LAWS, Statement by Greece, 9. Apr. 2018, p. 2; CCW Convention, GGE LAWS, 'Statement by Brazil', 26. Mar. 2019, p. 2; Open Roboethics initiative, 'The ethics and governance of lethal autonomous weapons systems: An international public opinion poll', 9 Nov. 2019; and Millar, J. and Moon, A., 'How to engage the public on the ethics and governance of autonomous weapon systems', Conference paper presented at We Robot 2016.

However, that view has been debated on several accounts. For example, some states and experts question the ethical basis on which the argument is made. They challenge HRW's assumption that the lack of emotions of AWS threatens the principle of humanity. Instead, they argue from a utilitarian standpoint that an AWS without feeling may result in more precise and controlled targeting, which consequently leads to greater respect for the principles of humanity.[75] Russia, for instance, has expressed the view (although not expressly in relation to the Martens Clause) that 'in addition to their technological advantages (accuracy, speed, effectiveness), such weapons neutralize human-caused risks (operator's mistakes due to his or her mental or physiological state, ethical, religious or moral attitudes), and thus reduce the probability of unintentional attacks against civilians and non-military targets'.[76]

Regarding the dictates of public conscience, it is argued that states, experts and opinion surveys are not necessarily unanimous on the question and that some people may find it ethically defendable to use AWS.[77] What constitutes 'public conscience' could also be questioned. What public? Whose conscience? How to guard against bias? Establishing what type and degree of human involvement could be demanded from users of AWS in light of the principle outlined in the Martens Clause would arguably be difficult.[78]

II. The material dimension: What is the nature and content of key IHL provisions?

While the GGE has agreed that human responsibility for decisions on the use of weapon systems must be retained, the question of what is required of states and other parties to armed conflicts to comply with IHL provisions in carrying out attacks with AWS remains debated. As it is beyond the scope of this report to discuss the content of all IHL provisions, the focus is on those that are currently most relevant in relation to the AWS debate, considered as two main sets of issues. First, there is the question of whether and to what extent the judgements and evaluations demanded by IHL can be automated. Second, there is the question of the informational basis on which targets can be identified and attacked.

To what extent can IHL-mandated evaluations be automated?

As with many other IHL rules, the cardinal rules governing the conduct of hostilities, especially concerning distinction, proportionality and precautions, presume the application of evaluative decisions and value judgements. With the introduction of AWS, this has raised the questions as to whether these judgements can be carried out only by humans and to what extent humans may rely on technical indicators.[79] Although it is generally agreed in the GGE that the use of force must reflect at least some manifestation of human agency and human intent, the extent to which it is permissible under IHL to allow evaluative judgements partially or fully implemented by AWS remains open to interpretation. There are currently two different standpoints.[80]

On the one hand, there are states and experts that believe this is ultimately a technical issue: the question is not if, but how, and to what extent, the IHL evaluation and judgement demanded by the rules on distinction, proportionality and precautions

[75] Evans (note 26), p. 730.
[76] CCW Convention, Working paper by Russia (note 64), pp. 5–6.
[77] Horowitz, M. C., 'Public opinion and the politics of the killer robots debate', *Research and Politics*, Jan.–Mar. 2016.
[78] CCW Convention, Chairperson's summary (note 51).
[79] Henderson, Keane and Liddy (note 68), pp. 12–13.
[80] CCW Convention, Chairperson's summary (note 51).

can be formalized in technical terms. From their standpoint, it is possible to envisage the formulation and configuration of information proxies or technical indicators that AWS could use to exercise IHL-demanded evaluations in a way that remains consistent with the intention of the person(s) responsible for deciding to employ and administer an AWS. With regard to the principles of distinction and proportionality, that would involve programming in advance target profiles and a decision method that would allow the systems to: (*a*) distinguish what is and what is not a 'military objective' (e.g. through indicators of the operational context and features of targetable and non-targetable people/objects); and (*b*) assess under what conditions force may or may not be applied (e.g. using parameters for threat level or context-based requirements for precautions).[81] With regard to the principle of precautions, the systems could, by design, be equipped with fail-safe mechanisms, the ability to give warnings, and self-destruct, self-deactivation or self-neutralization mechanisms.[82]

From this standpoint, the extent to which it is feasible to turn the assessments demanded by the principles of distinction, proportionality and precautions into functions reliant on data, sensors and algorithms is context-dependent.[83] It depends on the characteristics of the military objectives, the characteristics of the environment of use (e.g. how complex and dynamic it is) and what is technically feasible in light of their respective level of complexity. The more complex the target type and circumstance of use, the more technically challenging it will be to model the IHL-demanded assessment in technical terms. Consequently, for the proponents of this interpretation, the limits on autonomy and the question of what type and degree of human–machine interaction is needed for IHL compliance are ultimately issues of technological affordances in specific circumstances and limitations.[84]

If it can be predicted that an AWS can implement an evaluation—such as whether a person is a militarily active combatant who has not been rendered *hors de combat*—accurately and consistently with the intention of the human user in the operational circumstances, then the system can operate in full autonomy and no direct human supervision or intervention is needed after activation.[85] Alternatively, if the context of use is too complex to be modelled or if it creates a risk that an AWS might mischaracterize civilian objects as military objectives, then human supervision and intervention might be needed to compensate for the system's technical limitations.

On the other hand, some states and experts consider that the rules on distinction, proportionality and precautions demand value- and context-based judgements which do not lend themselves to machine automation.[86] While it is acknowledged that technology may help humans distinguish between targetable and non-targetable people and objects in some circumstances and to avoid or minimize adverse effects, they consider that the criteria underlying the principles of distinction, proportionality and precautions necessarily demand evaluative decisions and value judgements by humans. They question or reject the possibility that such human judgement can be 'baked' into AWS.

With regard to the principle of distinction, this standpoint notes that it would be challenging to identify metrics which would allow the system to reliably distinguish

[81] CCW Convention, Commentaries by the USA (note 64).

[82] As suggested e.g. by the USA and France. See CCW Convention, Commentaries by the USA (note 64); and CCW Convention, GGE LAWS, 'Operationalization of the 11 guiding principles at the national level', Comments by France, Aug. 2020, (5)(iii).

[83] See e.g. CCW Convention, UK expert paper (note 65); and CCW Convention, Commentary by Japan (note 3).

[84] CCW Convention, UK expert paper (note 65); CCW Convention, Working paper by Russia (note 64); CCW Convention, Commentary by Australia (note 3); and CCW Convention, Commentaries by the USA (note 64).

[85] Work (note 65).

[86] CCW Convention, Commentary by Switzerland (note 51); CCW Convention, Working paper by Venezuela (note 3); and CCW Convention, Joint commentary (note 25).

between people who are civilians, civilians directly participating in hostilities, militarily active combatants, and fighters *hors de combat*.[87] It also considers that the presumption of civilian status 'in case of doubt' would demand contextual assessments which would be difficult to preprogramme.

Regarding the principle of proportionality, it points out that what constitutes a military advantage is not based on quantifiable hard metrics, but is relative to the circumstances prevailing at the time of the attack.[88] This interpretation presumes the application of evaluative decisions and value judgements that cannot be turned into algorithmic formulas. These include the necessarily contextual judgements underlying the assessment of the 'excessiveness' of expected incidental harm in relation to the anticipated 'military advantage', and the prohibition of the destruction of civilian property except where 'imperatively' demanded by the necessities of war.[89] On this basis, the proponents consider that an AWS cannot be permitted to make proportionality assessments once deployed and, for that reason, it is the user's responsibility to make the proportionality assessment ahead of the attack.[90]

Regarding the principle of precautions, they note that it places a number of obligations on 'those who plan or decide upon an attack', therefore the evaluation it demands cannot be delegated to an AWS or substituted with preprogrammed precautionary (e.g. fail-safe and self-destruct) mechanisms. Some states have put forward the argument that while increased autonomy may improve the ability to comply with the principle of precautions, the same technology simultaneously increases the demands on human operators with respect to taking feasible precautions during the use of force.[91]

These two standpoints offer different perspectives on what the rules governing the conduct of hostilities demand from humans and permit from technology. Nevertheless, they are not necessarily incompatible in the sense that—at the core—they both expect the user of an AWS to foresee whether its use could result in unlawful effects. Where they diverge is on the role that humans (and concomitantly technology) might or must play in the foreseeability and administration of the likely effects. For the former, the limiting factor is technology, and human involvement in the performance of IHL-mandated evaluations is a practical response to technological limitations that may be overcome in the future. For the latter, humans will have to continue to make judgements demanded by IHL, regardless of how sophisticated the technology becomes.[92]

What kind of indicators can be relied on?

Whether evaluative tasks can be delegated to AWS and whether they may be used to apply force against human targets also depends on what type of information is

[87] Henderson, Keane and Liddy (note 68), pp. 5, 8–9; International Humanitarian Law Research Initiative, *Commentary to the HPCR Manual on International Law Applicable to Air and Missile Warfare* (Harvard University: Cambridge, MA, 2010), pp. 267–68; ICRC, 'ICRC position on autonomous weapon systems' (note 71); and Bo, M., 'Autonomy in weapons and/or targeting: The responsibility gap in the ICC Statute in light of the *mens rea* of the war crime of attacking civilians', *Journal of International Criminal Justice* (Mar. 2021), p. 5.

[88] The assessment of whether incidental civilian casualties and damages to civilian objects are likely to be excessive in view of the anticipated military advantage must be undertaken before deciding to launch an attack, on the basis of information available *ex ante*, i.e. at the time of planning and/or executing the attack.

[89] Hague Convention (IV) (note 43), Regulations, Article 23(g); Geneva Convention (IV) (note 39), Article 53; and Additional Protocol I (note 11), Articles 50(3), 52(3).

[90] Human Rights Watch and the International Human Rights Clinic (note 61); Sharkey, N., 'Saying "No!" to lethal autonomous targeting', *Journal of Military Ethics*, vol. 9, no. 4 (Dec. 2010); and Sharkey, N., 'Towards a principle for the human supervisory control of robot weapons', *Politica & Società*, vol. 36, no. 2 (2014).

[91] See CCW Convention, Commentary by Switzerland (note 51), p. 1.

[92] On the practical implication of these views for the exercise of human control, see Boulanin, V. et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*, SIPRI and ICRC Report (SIPRI: Stockholm, June 2020), pp. 8–10.

considered sufficient when exercising obligations under IHL, notably concerning distinction, proportionality and precautions. While some states argue that automated information may help improve the user's ability to strike a specific target, others point to risks associated with the use of automated information.[93] Thus, it remains to be clarified on what informational basis a target may be identified as a lawful target or a protected object or person under IHL. A key question is to what extent IHL permits the use of information proxies and technical indicators.[94]

The use of AWS requires the formulation in advance of technical indicators that will allow the systems to detect, identify, select and prioritize targets once the system is activated.[95] The informational basis on which these indicators may be programmed remains a debated legal question, not just in the context of the CCW process on AWS.[96] The issue of whether combatants and military objects may be targeted based on information proxies, such as location and biometric or behavioural features, has also been discussed in the relation to the debate on targeted strikes, for example.[97]

Under IHL, central concepts such as combatants, non-combatants, civilians and military objectives are not necessarily precisely defined.[98] Such people have been conceptualized in relation to their (non-)membership in the armed forces and their (non-)participation or (non-)contribution in the conduct of hostilities. Therefore, the basis on which a person—or object—may be identified, and consequently encoded, as a legitimate target remains subject to different interpretations and practice.[99]

For some states, a military objective can be framed in geographical terms: any individual or object located in defined areas of heavy fighting may be presumptively considered a legitimate target. Such types of military objectives are sometimes referred to as 'kill boxes'.[100] Some states also consider that biometrical and behavioural markers (e.g. assigned gender, height, age and gait) may be legitimately used as proxies to (help) determine whether an individual is a combatant or civilian for purposes of targeting.[101]

This view has been challenged and criticized by other states and a number of civil society organizations, which question the feasibility and the legality of distinguishing combatants and non-combatants based on geographical location or 'physiological' or 'social' indicators.[102] IHL does not explicitly prohibit the use of such indicators to identify targets, but reliance on them alone is arguably not sufficient to meet the legal obligations under the principles of distinction, proportionality and precautions.[103]

[93] This is the view of the USA, for example, on the one hand, and Austria, Belgium, Brazil, Chile, Ireland, Germany, Luxembourg, Mexico and New Zealand, on the other hand. See CCW Convention, Commentaries by the USA (note 64), p. 1; and CCW Convention, Joint commentary (note 25), p. 4.

[94] Lewis (note 6), p. 13.

[95] What is understood here is that AWS do not have free will. They can only act within the limits of their programming.

[96] Moyes, R., 'Target profiles', Article 36 Discussion Paper, Aug. 2019.

[97] Ackerman, S., 'US to continue "signature strikes" on people suspected of terrorist links', *The Guardian*, 1 July 2016; and Davis, L. E., McNerney, M. and Greenberg, M. D., 'Clarifying the rule of targeted killing: An analytical framework for policies involving long-range armed drones', Rand Corporation Research Report, 2016.

[98] ICRC, Customary IHL Database, 'Rule 3. Definition of combatants'.

[99] Melzer, N., 'Interpretive guidance on the notion of direct participation in hostilities under international humanitarian law', *International Review of the Red Cross*, vol. 90, no. 872 (Dec. 2008).

[100] Beauchamp, S., 'The moral cost of the kill box', *The Atlantic*, 28 Feb. 2016; Air Land Sea Application (ASLA) Center, 'Kill box: Multiservice, tactics, techniques, procedure for kill box planning and employment', June 2018; and Work (note 65).

[101] Chamayou, G., *A Theory of the Drone* (New Press: New York, 2011); and Heller, K. J., '"One hell of a killing machine": Signature strikes and international law', *Journal of International Criminal Justice*, vol. 11, no. 1 (Mar. 2013).

[102] Brehm, M., *Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law* (Geneva Academy of International Humanitarian Law and Human Rights: Geneva, 2017), pp. 61–62; and Pejic, J., 'Extraterritorial targeting by means of armed drones: Some legal implications', *International Review of the Red Cross*, vol. 96, no. 893 (2015), pp. 22–28.

[103] Brehm (note 102), pp. 61–62; and Moyes (note 96).

A related issue is how precise the target profile needs to be for the user to be satisfied that the AWS will be used lawfully. Of particular relevance here is the prohibition against indiscriminate attacks, which demands that the effects should be limited and directed at a specific military objective. This calls for a clarification of what defines a 'military objective', and notably how specific it needs to be. For this purpose, IHL already distinguishes between three types of military objects, namely those by nature, location, and purpose or use.[104] These widely accepted categories serve as useful frameworks when discussing—and agreeing on—the circumstances in which technical indicators may be relied on, what types of restrictions may be placed on target profiles, and what the temporal and geographical scope of an operation may be.

It is widely accepted that distinct military objectives cannot be amalgamated into one single objective as reflected by Article 51(5) (a) of Additional Protocol I to the 1949 Geneva Conventions, as well as the 1996 CCW Amended Protocol II on mines, booby-traps and other devices, which states that: 'Several separated and distinct military objectives located in a city, town, village or other area containing a similar concentration of civilians or civilian objects are not to be treated as a single military objective.'[105]

However, the extent to which this interpretation restricts the way targets can be encoded into an AWS remains a question that deserves further discussion. Does it mean that an AWS cannot be programmed to attack multiple and distinct military targets or different locations in the same mission, without authorization or guidance from the user? Or does it mean an AWS can be aimed at different targets and locations as long as they are part of a larger, identifiable and coherent military objective? The question of how broadly or narrowly states define a 'single military objective' is likely to be critical.

States, as well as experts consulted as part of this project, stressed that, as far as objects are concerned, the type of military objective was key when considering these questions.[106] In their view, military objectives by nature, such as a military base, would probably allow for more reliance on technical indicators, in contrast to military objectives by location or by purpose or use, such as a border area, bridge or school transformed temporarily to military headquarters.[107] Thus, the more complex the battle environment and nature of targets, the less reliance on technical indicators would be lawful. However, this area remains relatively unexplored in relation to AWS and states are encouraged to elaborate their views on under what circumstances different types of automated information are permitted.

III. The temporal dimension: At what points in time should IHL provisions be respected?

AWS are categorized as time-delay weapons by definition, as there can be a substantial time lag between the decision to use the force and the effects. As with every time-delay weapon, they raise a number of issues related to the temporality of IHL-mandated provisions. When do the various obligations underlying respect for the principles of distinction, proportionality and precautions begin and end? Can they be fulfilled in advance and, if yes, how far in advance does the law permit these decisions

[104] ICRC, Customary IHL Database, 'Rule 8. Definition of military objectives'.

[105] Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices (Protocol II) to the CCW Convention, opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983, as amended 3 May 1996; and Additional Protocol I (note 11), Article 51(5)(a).

[106] CCW Convention, Joint commentary (note 25), p. 4.

[107] This view is supported by the ICRC, which believes that the use of AWS should be limited to targeting military objectives 'by nature' only; see ICRC, 'ICRC position on autonomous weapon systems' (note 71).

to be made? And under what circumstances would the evaluations demanded by IHL provisions need to be performed after activation?

When do restrictions on attacks begin and end?

The principles of distinction, proportionality and precautions are, like many of the IHL rules regulating the conduct of hostilities, formulated as restrictions on attacks. What constitutes an 'attack' is therefore defining for when legal obligations under IHL apply and need to be performed.[108] However, the definition of attack in Article 49 of Additional Protocol I does not specify when an attack begins and ends.[109] This is not a new issue but the use of AWS has stressed the need to clarify that aspect.[110]

AWS, like time-delay weapons before them, raise interpretative questions about the temporal scope of an attack.[111] Does an attack using AWS start at the programming phase when targeting parameters are defined? Does it start when the weapon is activated? Or does it only start when the weapon has reached the target area and is searching for targets? At the other end of the process, does an attack only end once the weapon is deactivated or has released its payload? How is it possible to discern the end of an attack and/or the beginning of another attack, and to what extent do lethal effects define the temporal delineations?[112] These questions have practical implications for the type and degree of human–machine interaction that needs to be exercised to ensure compliance with IHL, and the topic is yet to be discussed in detail in the GGE and in the relevant literature. Two approaches could be explored moving forward.

One approach to discerning the beginning of an attack would be to consider at what point in the operation a person or object becomes directly endangered. Adopting this approach would require an assessment of what direct endangerment means.[113] For instance, in the case of a mine, would the point of direct endangerment be when a mine is laid or at some later point in time when a person or vehicle approaches the mine? Similarly, regarding AWS, direct endangerment could be interpreted as occurring at various points in time, including: (*a*) at the point when the person or object is selected for force application; (*b*) earlier, at the point when the person or object is identified as a potential target; or (*c*) even earlier, at the point when the AWS is activated and begins attempting to match input data to the target profile library, because then all persons or objects within the operational area of the AWS are potentially directly endangered.

Another approach to discerning the beginning of an attack is to identify the specific act that triggers the attack. CCW Amended Protocol II appears to adopt this approach: Article 3(8) replicates in part Additional Protocol I, Article 51(4) and (5), but instead of using the term 'indiscriminate attacks' it refers to 'indiscriminate use'.[114] This change in terminology could indicate that the use of a weapon was considered equivalent to

[108] The use of the term 'armed attack' under international law on the use of force is not discussed here. Military instruction manuals in many countries define an attack differently; see Schmitt, M., '"Attack" as a term of art in international law: The cyber operations context', eds C. Czosseck, R. Ottis and K. Ziolkowski, *Proceedings of the 4th International Conference on Cyber Conflict* (NATO CCD COE Publications: Tallinn, 2012).

[109] Commentary of 1987 to Additional Protocol I, Commentary on Article 49, para. 1879.

[110] Elaborating the scope of the notion of an 'attack' does not imply that IHL provisions are limited to those temporal and spatial limitations. IHL provisions extend to acts prior to, during and after an attack, including planning, deciding on, carrying out, preventing, repressing and reporting alleged violations of the rules on attacks.

[111] The use of certain technologies of warfare has, in the past, generated an examination of which acts qualify as an attack under IHL. This includes the use of mines, drones and cyber operations in armed conflict.

[112] The temporal lack of clarity around 'attack' has been raised, among others, by the USA in 2019: 'These requirements (IHL requirements, *ed.*) address "attacks", rather than the firing or activation of weapon systems as such. For example, 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.' See CCW Convention, GGE LAWS, 'Implementing international humanitarian law in the use of autonomy in weapon systems', Working paper by the USA, Mar. 2019.

[113] Commentary of 1987 to Additional Protocol I, 'Definition of attacks and scope of application', para. 1881.

[114] Protocol II to the CCW Convention (note 105), Article 3(8).

the concept of attack. Article 3(8) goes further, stating that 'indiscriminate use is any *placement* of [mines, booby traps and other weapons within the scope of Amended Protocol II]' (emphasis added). Applying this approach to AWS, it is necessary to consider what action is the equivalent of 'placement'. Of the available options, the 'activation' of an AWS is probably most similar to the placement of a mine. One possible exception to this approach would be the case of AWS that are activated outside their operational area, but only begin to search for targets once they reach their operational area.

Adopting the first approach above, an attack would end whenever a person or object ceases to be directly endangered. In attacks involving an AWS, this could be when the AWS: (*a*) is deactivated or (self-)destroyed; (*b*) exhausts its payload; or (*c*) ceases searching for targets (if it is programmed to do so). In some cases, it may also be difficult to discern whether an attack has ended and a new attack commenced, or whether the initial attack was simply suspended—in the language of Additional Protocol I, Article 57(2)(b)—and resumed. This would also be the case when a user provides new instructions to the AWS, for example, by altering the target profile. While the distinction may be difficult in practice, it has significant legal implications. A new attack presumably demands new IHL assessments, whereas a suspended attack may not.

Existing definitions arguably allow for a broad interpretation of the notion of attack in relation to AWS. However, while a broad definition may be adopted, outer limits will still need to be clarified with respect to the unique characteristics of AWS.

When should the mandated evaluations and processes be implemented?

Compliance with the rules of distinction, proportionality and precautions requires consideration of the circumstances prevailing at the time of an attack. The use of an AWS presupposes that the user assesses *in advance* whether the system can and will be used in compliance with these rules. This prospect raises three practical questions: (*a*) how far in advance an assessment may be made; (*b*) under what circumstances the user may assume that an *ex ante* assessment will continue to be valid after an AWS is activated; and (*c*) whether an ongoing or new legal assessment needs to be made after an AWS has been activated.

States are still to articulate detailed positions on this set of questions. The dominant framing view seems to be that these questions cannot be answered in the abstract, as they depend on the characteristics of the environment of use: whether it is predictable, dynamic or populated. The assumption is that the more dynamic and unpredictable the environment is, and the more likely the presence of civilians and civilian objects is, the closer in time to the use of force that the legal assessment needs to be.[115] Alternatively, if the environment is known to be relatively predictable—because it is static or controlled—and does not include the presence of military personnel or civilians, then it can be assumed that the IHL-mandated evaluations could be made more easily in advance. The risk that new events would undermine their validity during the weapon system's mission would be lower. From this standpoint, the crux of the problem relates to whether and how the user can reasonably foresee how variation in the intended circumstances of use might result in prohibited decisions or effects.

An alternative interpretation is that compliance with the principle of precautions, including the requirement for 'constant care', supposes that the IHL-mandated evaluation should be exercised *throughout* the operation of an AWS.[116] This has been inter-

[115] During GGE meetings in 2020, the chair noted that, among states, 'human control/involvement/judgement needed to be reasonably temporally proximate to an attack, to remain valid'; see CCW Convention, Chairperson's summary (note 51), p. 9.

[116] Additional Protocol I (note 11), Article 57(1).

preted by some states and experts as an indirect requirement for maintaining the AWS under some form of direct supervision and maintaining the possibility to intervene to cancel or suspend the system's operation.[117] This conclusion has been challenged by other states and experts who consider that precautionary duties are conditioned on what is feasible at the time of the attack. For them, the principle concerns precautions that 'are practicable or practically possible taking into account all circumstances ruling at the time, including humanitarian and military considerations'.[118] In their view, this requires a context-based assessment of the foreseeable effects of the available weapon systems based on their technical features and the expected circumstances of their use, and consideration of alternative weapons and tactics, if these would avoid or minimize the likelihood or extent of incidental civilian harm. This view does not elaborate on the critical issue of what 'at the time' means. Is it at the point of activation of the weapon? Or is it when the weapon has reached the target area? This circles back to the aforementioned discussion on the temporal scope of an AWS attack.

IV. The geographical dimension: In relation to what locations must users respect IHL provisions?

The geographical relationship between the subject and the object of an attack has been a recurring theme in military ethics since antiquity.[119] The case of AWS fits in that regard into a larger debate on remote warfare and the increasing distancing between decisions to use force and their effects on the battlefield. From the perspective of respect for IHL, with AWS the question is less about where the decisions are made and more about what location or locations are implicated by the decision to use or limit the use of AWS and what must be known about the location(s) where force might be applied. This raises two sets of issues for consideration: first, regarding where an attack starts, extends and ends; and second, regarding how the user of an AWS should interact with the environment of use.

Where does an attack start, extend and end?

The first set of issues relates to the spatiality of an attack (and, perhaps, military operations more broadly). What is so unique with AWS—contrary to remotely controlled armed drones and missile technologies—is that the user does not necessarily know in advance what the system will specifically attack, or when and where. That uncertainty is a challenge for IHL compliance, since the principles of distinction, proportionality and precautions are premised on sufficient awareness and assessment of the circumstances prevailing at the time of the attack.

In general, states and experts seem to agree that spatial limitations on the operation of an AWS may help the user to reduce unpredictability and have more certainty that the system will not result in prohibited decisions or effects. The major question, however, is to determine what IHL says about where an attack begins, how far it extends and where it ends. Mobile AWS could potentially have the capability to travel and operate over large distances. How then to account for the distance that an AWS could travel to reach the target area? How wide, broad or well defined can the target area be? States have yet to speak with great clarity regarding what limits IHL imposes with respect to the distance that an AWS may travel as part of an attack.

[117] CCW Convention, Working paper by Venezuela (note 3); and CCW Convention, Joint commentary (note 25), p. 2.

[118] Protocol II to the CCW Convention (note 105); Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons (Protocol III) to the CCW Convention, opened for signature 10 Apr. 1981, entered into force 2 Dec. 1983, Annex I, Article 1(5); and ICRC, Customary IHL Database, 'Rule 15. Principle of precautions in attack'.

[119] Gros, F., *State of Violence: An Essay on the End of War* (Chicago University Press: Chicago, 2006).

It seems broadly accepted by states that case-by-case evaluations are required, as a range of factors need to be considered. First, the nature of the environment needs to be considered.[120] Second, whether the intended environment might include elements that are protected under IHL, such as civilians, civilian objects and the natural environment. Third and finally, how the characteristics of the AWS itself interact with the environment and variations within it, for example, what the system can perceive, the extent to which it can cope with changes in weather conditions, and what the environmental effect of an attack will be. [121]

How should the user of an AWS interact with the environment of use?

The second set of issues relates to the type and degree of interaction that the user of the weapon should have with the environment of use. This issue can be broken down into two related questions. The first is what is the user expected to know about the environment in order to respect IHL provisions before and during an attack? Arguably, the highest standard of knowledge that is demanded by IHL is captured by the principle of precautions, which requires those who plan or decide on an attack to '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'.[122] What 'everything feasible' concretely means is subject to debate. It would be helpful in that regard if states could further articulate their views on the types of measures that they would deem appropriate and what knowledge of the environment of use the user is expected to possess.

The second question is how and to what extent is the user expected to interact with the environment once the weapon has been activated? The principle of precautions demands that a party has the possibility 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 may be expected to be disproportionate'.[123] Some states and experts have interpreted that obligation as an indirect requirement for the user of an AWS to keep it under some form of direct supervision and control, enabling the user to maintain the ability to intervene and veto an attack.[124] It has even been argued that of the three cardinal principles, the principle of precautions might be the one that places the most restrictive limits on the development and use of AWS with respect to exercising human control.[125]

However, that view has been challenged in the context of the GGE debate. Some states consider that it would be impractical and inconsistent with the way weapons are currently used to infer a demand for active human supervision and control from the principle of precautions. Such an interpretation would make the use of 'fire and forget weapons'—missiles and other types of smart munitions that cannot be recalled after launch—potentially unlawful. They also question the relative value of having a human operator in a supervisory role, especially for cases of use that may involve fighting at

[120] There are legal provisions and customs that pertain to the environment of use, some related to naval warfare, air warfare and warfare on land. See e.g. San Remo Manual on International Law applicable to Armed Conflicts at Sea, 12 June 1994; and Program on Humanitarian Policy and Conflict Research at Harvard University (HPCR), *Manual on International Law Applicable to Air and Missile Warfare*, Bern, 15 May 2009 (Harvard University: Cambridge, MA, 2009).

[121] Boulanin et al. (note 92).

[122] Additional Protocol I (note 11), Article 57(2)(a)(i).

[123] Additional Protocol I (note 11), Article 57(2)(a)–(b); and ICRC, Customary IHL Database (note 8), rules 15–19.

[124] Boulanin et al. (note 92).

[125] Thurnher, J., 'Means and methods of the future: Autonomous systems', eds P. A. L. Ducheine, M. N. Schmitt and F. P. B. Osinga, *Targeting: The Challenges of Modern Warfare* (Asser Press: The Hague, 2016).

machine speed.[126] Taking the example of existing automated air defence systems, they point out that in some operational conditions artificial agents may be better placed than humans at undertaking precautions.[127] From that standpoint, the question of whether human supervision and intervention is needed is ultimately one of what can be expected from the technology and humans in a specific context of use.

V. Conclusion: Human–machine interaction in the conduct of hostilities

In 2020, the report 'Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control', by SIPRI and the International Committee of the Red Cross (ICRC), found that human control measures could take multiple forms.[128] These could be targeted at the design of the weapon, the environment it is used in and the way that the user interacts with it. The study also found that the necessary combination of measures is likely to be context-dependent and determined by the characteristics of the systems and of the environment. These findings have been positively received by states, as reflected in the 2020 chair's summary of the GGE discussions, which includes some of the elements in its recommendation for the future elaboration of the guiding principles.[129]

The issue of how states can determine what combination of measures is required to respect and ensure respect for IHL remains a vexed issue. This chapter has shown that it raises fundamental questions about interpretation and application across multiple dimensions: personal, material, temporal and geographical. Each dimension merits attention on its own, but it may be in the combination of two or more of these dimensions that the most important questions about the type and degree of human–machine interaction demanded for IHL compliance can be found. What combination would be intolerable under IHL? At what point does the combination of distribution of responsibility between multiple humans, reliance on automated processes and distancing in time and space become such that it is no longer feasible to reliably foresee, effectively administer and sufficiently trace the operation, performance and effects of an AWS? Some states have already elaborated on why and how the use of certain types of autonomous weapons, such as close-in weapon systems, can be used in compliance with IHL across these four dimensions. It would be helpful if states could further articulate what combinations would reflect respect or lack of respect for IHL in the conduct of hostilities.

The use of concrete scenarios could also be helpful for states to explore the outer limits of what is permitted under IHL. There are a number of variables that states could combine in different scenarios to explore what would be permitted, required or prohibited. For example, what combination of personal, material, temporal and geographical dimensions would make the use of a swarm of anti-personnel AWS in a populated area lawful or unlawful? Could the nature of the target demand human supervision regardless of technological capabilities? What would the user need to know about the system and the environment of use in order to authorize a launch and decide on the necessary level of human–machine interaction after it? And what role does the timing of operation play? These are just some of the questions that states needs to consider in their ongoing discussions on the conduct of hostilities and human–machine interaction.

[126] Scharre, P. and Horowitz, M. C., 'Meaningful human control in weapon systems: A primer', Center for a New American Security (CNAS) Working Paper, Mar. 2015.
[127] CCW Convention, Working paper by the USA (note 112); and CCW Convention, UK expert paper (note 65).
[128] Boulanin et al. (note 92).
[129] CCW Convention, Chairperson's summary (note 51).

4. Key issues concerning legal reviews and legal advice

It is widely recognized, including by the GGE, that the rules on legal reviews and legal advice in articles 36 and 82 of Additional Protocol I to the 1949 Geneva Conventions are of critical importance to respect and ensure respect for IHL in the development and use of AWS.[130] Legal reviews are traditionally focused on the inherent characteristics of a weapon, while legal advice focuses on how employable weapons, means and methods of warfare may or may not be used in compliance with IHL in particular military operations. However, as a number of states and experts have already noted, the presence of autonomy may pose novel questions concerning their implementation.[131] The fact that AWS are preprogrammed to execute targeting tasks traditionally undertaken by humans during a specific attack necessarily influences how IHL provisions related to prospective assessments, ongoing administration and post-operation assessment should be interpreted and applied, as well as how they might interact with each other. This chapter explores the key issues that AWS pose for compliance with the conduct of legal reviews (section I) and the provision of legal advice (section II). Generic issues of interpretation and application have been extensively discussed elsewhere and are not covered in detail here, rather the focus in on AWS-specific issues.[132]

I. Legal reviews of new weapons, means and methods of warfare

Although the GGE has established that the conduct of legal reviews is an important mechanism to ensure the lawful development and use of AWS, the group has also recognized that AWS present novel challenges to this process. As noted by the chair of the GGE in 2020, this is particularly due to the 'possible unpredictability and self-learning capabilities' of AWS.[133] Further, as a number of states and experts have already pointed out, the process is challenged by the fact that legal reviews are national procedures which are not subject to international oversight and for which there are no agreed standards regarding how they should or shall be conducted.[134] In short, the question of how and on what basis an AWS will be reviewed is left largely to the discretion of states, resulting in the concern that they might apply vastly different methods and standards in their reviews. Intending to foster a greater common understanding of what IHL demands from the legal review of an AWS, this section discusses some of the complexities that states may encounter when seeking to perform the obligation in general, and in relation to AWS specifically. These are discussed

[130] This is particularly true for the conduct of legal reviews, which has been extensively discussed in the GGE process. Comparatively, the conduct of operational legal advice has received much less attention. The extensive focus on legal reviews is reflected in Guiding Principle (e), adopted by the GGE: 'in accordance with States' obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law'; see CCW Convention (note 2), Annex IV.

[131] Farrant, J. and Ford, C. M, 'Autonomous weapons and weapon reviews: The UK Second International Weapon Review Forum', *International Law Studies*, vol. 93, no. 389 (2017); Boulanin, V., 'Implementing Article 36 weapons reviews in the light of increasing autonomy in weapon systems', SIPRI Insights on Peace and Security no. 2015/1, Nov. 2015; and Chengeta, T., 'Are autonomous weapon systems the subject of Article 36 of Additional Protocol I to the Geneva Conventions?', *International Law & Policy*, vol. 23, no. 1 (2016), pp 66–99.

[132] ICRC (note 33); McClelland (note 36); Brown, G. and Metcalf, A., 'Easier said than done: Legal reviews of cyber weapons', *Journal of National Security Law & Policy*, vol 7, no. 115 (2014); and Copeland, D. P., 'Legal review of new technology weapons', eds H. Nasu and R. McLaughlin, *New Technologies and the Law of Armed Conflict* (Asser Press: The Hague, 2014).

[133] CCW Convention, Chairperson's summary (note 51), p. 5.

[134] CCW Convention, GGE LAWS, 'Strengthening of the review mechanism of a new weapon, means and method of warfare', Working paper by Argentina, CCW/GGE.1/2018/WP.2, 4 Apr. 2018; CCW Convention, GGE LAWS, 'Weapons review mechanisms', Working paper by the Netherlands and Switzerland, CCW/GGE.1/2017/WP.5, Nov. 2017; and ICRC (note 61), p. 8.

across the four dimensions outlined in the introduction: personal, material, temporal and geographical.

The personal dimension: Who should conduct the review

A first, practical complexity relates to who should conduct the legal review. Article 36 does not prescribe whether the review process should be conducted by a single (legal) specialist or a committee. States that do conduct legal reviews use different models of compliance.[135] The case of AWS supposes that whatever format is adopted, technological expertise will need to be mobilized alongside legal expertise. This is because the (il)legality of a use of force by an AWS is intrinsically linked to the technical characteristics of its autonomy.[136]

As reflected in the ICRC's guide to legal reviews, it is already considered best practice to use a multidisciplinary approach that involves different areas of expertise in reviews, notably technical and medical expertise.[137] A number of states already do this as part of their review process.[138] Some states directly invite technical experts to join a review committee, while others rely on technical consultation with either specific experts involved in the design of the weapon or independent experts.

However, mobilizing adequate technical expertise may be a challenge for some states, for example, due to the lack of an established military industrial base or limited technical expertise on autonomous systems and artificial intelligence (AI) within governmental agencies. Notably, states that develop and produce weapons have a comparative advantage here over those that just import them. Therefore, as part of the obligation to review the legality of weapons, means and methods of warfare involving emerging military technologies, states are encouraged to explore the possibilities for greater cooperation and information sharing on technical aspects associated with AWS. Cooperation could take the form of technical training for legal reviewers, or the exchange of information about trends in technological development and methodologies for testing and evaluation.

The material dimension: What needs to be reviewed and on what legal basis

The second and perhaps most significant set of complexities relates to the question of how the presence of autonomy may impact the material scope of the review: what needs to be reviewed and on what legal basis.

The technological scope: Autonomy and technological interdependencies

The first issue is of a practical nature and relates to the extent to which the review should consider the technological and environmental interdependencies that underpin autonomy, and ultimately the performance, behaviour and effects of an AWS. At the most basic level, autonomy is enabled by a suite of sensors through which the system perceives the world (including targets), computer hardware and software through which the system transforms data perceived from the environment into purposeful plans of action (including what to target and under what conditions), and actuators or effectors through which the system acts in its environment of use (including

[135] Park, W. H., 'Conventional weapons and weapons reviews', eds T. Gill et al., *Yearbook of International Humanitarian Law*, vol. 8 (Asser Press: The Hague, 2005), p. 107. For a comparison of state practice see Boulanin and Verbruggen (note 32), p. 9.

[136] For a comparison of state practice see Boulanin, V. and Verbruggen, M., *Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies* (SIPRI: Stockholm, Nov. 2017).

[137] ICRC (note 33).

[138] For a comparison of state practice see Boulanin and Verbruggen (note 32), p. 9.

engines and weapon payload).[139] The characteristics of all these components, and their interactions with each other and the anticipated environment of use, need to be considered in a legal review to generate a picture of the foreseeable performance, behaviour and effects of a weapon.

A practical issue, however, is that these components may not necessarily be instantiated by a single physical object or be self-contained on board the weapon platform. AWS components can also be physically distributed and interconnected, across a system of systems.[140] Further, the performance and reliability of the system may be affected by design decisions that can be difficult to trace or comprehend without adequate expertise or documentation. Notably, that includes the type and volume of data, and the type of computational or AI methods that have been used to develop the target recognition software. These elements are essential to foresee what the systems can or cannot perceive from the environment of use.

Therefore, it would be constructive for states to express their views on what a review process would need to entail in order to reliably discern whether or not the technological and environmental interdependencies underlying anticipated employments of an AWS would reflect compliance with IHL.

The legal basis: Weapons, means and methods of warfare

The second issue relates to how AWS should be categorized for the purpose of legal review. Under IHL, legal reviews are to be conducted for new weapons, means and methods of warfare. A pertinent question for all states, therefore, is whether AWS are to be categorized merely as weapons and/or means of warfare, or also as methods of warfare.[141] This is relevant because states may view the distinction as relevant to the scope of their obligation to conduct legal reviews or may apply different legal review procedures to weapons and means of warfare on the one hand, and methods of warfare on the other. More importantly, it may help determine whether IHL-required human–machine interactions should be considered in the review process.

AWS most likely need to be reviewed both as weapons and as methods of warfare, given that the presence of autonomy necessarily impacts the way the weapon can be used in compliance with IHL. The ICRC views it as best practice to consider the normal or expected use of a weapon in the legal review, since a weapon or means of warfare cannot be assessed in isolation from how it is used.[142] In concrete terms, this means that the review would have to consider the strategies and 'tactics, techniques or procedures' involved in using AWS to conduct hostilities against an adversary.[143]

The novel, and difficult, question is how the reviewer should consider the interdependent elements underlying the technology, the anticipated environment of use and the various humans involved in the determination of whether the employment of an AWS may be unlawful in some or all circumstances. Ultimately, the reviewer will be confronted with the set of conduct-of-hostilities issues that were raised above concerning the rules on weapons, means and methods of warfare (see chapter 3). A key question is whether the review should presume that the AWS itself may exercise and implement IHL provisions, or whether the review should presume that those

[139] Boulanin, V. and Verbruggen, M., *Mapping the Development of Autonomy in Weapon Systems* (SIPRI: Stockholm, Nov. 2017), pp. 11–12; and UNIDIR, 'Framing discussions on the weaponization of increasingly autonomous technologies', UNIDIR Resources, 2014.

[140] Boulanin and Verbruggen (note 139), p. 12.

[141] 'Weapons', 'means' and 'methods' of warfare are not defined terms in IHL. For discussion about definitions in the context of legal reviews, see ICRC (note 33).

[142] ICRC (note 33), p. 10; and Commentary of 1987 to Additional Protocol I (note 30), Commentary on Article 36, para. 1466.

[143] Sassoli, M., Bouvier, A. and Quintin, A., *How Does Law Protect in War? Cases, Documents and Teaching Materials on Contemporary Practice in International Humanitarian Law*, vol. III, 3rd edn (ICRC: Geneva, 2011), p. 280; and McClelland (note 36), p. 405.

may be exercised and implemented only by natural persons. This question is critical because it has methodological implications, and it may determine the outcome of the review. In the former case, the review would consider the system's technical ability to comply with the principle of distinction, proportionality or precautions. In the latter case, the review would instead focus on how the system's technical characteristics would facilitate or impede the user's ability to employ it in compliance with IHL in a particular environment of use. This is a notable difference of approach.

The legal scope: Applicable fields of international law

A third, fundamental issue is what rules of international law should form part of legal reviews. Should it only be the rules of IHL? If so, which? If not, which other fields of international law should be considered? These questions predate the debate on AWS, but have been brought to the surface by the question of autonomy.

With regard to the applicable fields of international law, Article 36 is framed broadly: reviewers are to have regard not only for the rules in Additional Protocol I, but for all the international law rules applicable to the High Contracting Parties, whether treaty-based or customary. The discussions in the GGE about the conduct of legal reviews of AWS tend to focus on IHL because the CCW Convention is an IHL-rooted treaty, but concerns over AWS extend beyond the IHL framework. International human rights law (IHRL) and ICL are among other fields of international law that arguably should be considered in legal reviews.[144] Some states consider IHRL as providing particular protection against the use of AWS. They invoke the target's right to life and right to dignity as a possible legal basis for demanding that the decision to use force is not delegated to AWS or autonomous targeting processes. The question of whether, under what circumstances and to what extent IHRL places limits on the legality of AWS would arguably be important in the context of a legal review process. However, states currently do not systematically take into consideration IHRL in the legal review process.[145] These diverging practices reflect enduring differences of interpretation of the applicability of IHRL during armed conflict. The same differences have led some states to reject the aforementioned argument that the right to life and right to dignity could be invoked to demand human agency in the use of force. It would be useful for states to articulate their views on the relationship between IHL and IHRL and discuss what uses, cases or situations would legitimately trigger consideration of protection under IHRL in legal reviews.

Regarding IHL specifically, while its centrality for the review process is uncontested, states and experts have expressed different views over the years as to what rules are relevant and how those rules need to be interpreted and applied. In the case of AWS, the debate has focused on whether—and if so, how—the review process should consider rules governing the conduct of hostilities, as well as the Martens Clause. States and experts have, at the outset, different understandings of the pertinence of the rules governing the conduct of hostilities for the review process. Some find the rules relevant and consider that if the reviewer is not satisfied that the state can employ the weapon, means or method of warfare in compliance with these rules, the review should not allow the weapon to be used, or at least limits should be placed on its use.[146] Others consider that these rules are not necessarily determinative of the lawfulness of a weapon, means or method of warfare.[147] They argue that most weapons can be used

[144] ICRC (note 61), p. 21; CCW Convention, Commentaries by Portugal (note 9); CCW Convention, Commentary by the Netherlands (note 3); CCW Convention, Joint commentary (note 25), p. 5; and Boulanin and Verbruggen (note 32).

[145] For a comparison of state practice see Boulanin and Verbruggen (note 32), p. 9.

[146] See the Netherlands in Boulanin and Verbruggen (note 32), p. 9.

[147] See the USA in Boulanin and Verbruggen (note 32), p. 21; Schmitt, M. and Thurnher, J., '"Out of the loop": Autonomous weapons systems and the law of armed conflict', *Harvard National Security Journal*, vol. 4 (2013), p. 243;

lawfully in some circumstances and misused in others, and that the responsibility for respecting and ensuring respect for IHL rules on the conduct of hostilities rests with the user, not the legal reviewer. In their view, legal reviews are conducted on the basis that, if approved for use, a weapon, means or method of warfare will be employed in hostilities in accordance with the relevant IHL rules.

The strict separation made in the latter approach is questionable in the case of AWS. The use of autonomy for the execution of targeting tasks implies that some aspects of the assessments required for IHL compliance will be baked into the AWS and, therefore, determined much further in advance of an application of force than when using other types of weapons. The ability of the user of an AWS to employ it in accordance with the principles of distinction, proportionality and precautions will be greatly conditioned on design choices. For that reason, it is recommended that the reviewer include consideration of the rules governing the conduct of hostilities. Such an inclusion would most likely be beneficial in terms of IHL compliance, since it could generate insight and recommendations that would make the user better equipped to determine how, where and when it may or may not use such a weapon.

Whether a legal review of AWS should take into consideration the Martens Clause has also been debated. As flagged above in the overview of the limits on AWS under IHL, states continue to have a different understanding of the legal status of this clause (see chapter 2). According to the ICRC's guide on legal reviews, the Martens Clause needs to be considered when there are no other relevant treaty or customary rules. On that basis, some states argue that the clause would not need to be considered since the development and use of AWS is already regulated by IHL. Others consider that the Martens Clause provides guidelines for the evolution of customary or treaty law and, for that reason, it invites the review to weigh in on discussions such as the one in the CCW on AWS, which could lead to future developments in international law. The question of how a legal review of an AWS may interpret the principle set out in the Martens Clause remains vexed, however, as the normative debate on AWS remains unsettled.

The temporal dimension: When and how often the review should be conducted

The third set of complexities relates to the timing of the legal review and the duration of its findings. According to the wording of Article 36, legal reviews are to be conducted in 'the study, development, acquisition or adoption of a new weapon, means or method of warfare'. It is generally agreed that reviews should occur as early as possible and be repeated as necessary.[148] In practice, the following triggers for a legal review are acknowledged: (*a*) when a state is acquiring a weapon, means or method of warfare for the first time, even if it has been employed by others; (*b*) when a state adheres to a new international law obligation relevant to the use of that weapon, means or method of warfare; (*c*) when a state sufficiently modifies an existing weapon, means or method of warfare; or (*d*) when new information comes to light about the performance or effects of a weapon, even after the weapon has been employed by a state.

The case of AWS poses two practical questions. The first is how should the review process account for the fact that AWS, as a software-based technology, will frequently undergo software updates? Reviews will probably need to be repeated multiple times and at shorter intervals than for technologies not reliant on complex software systems. The crux, however, is to determine what type or degree of software modification

Human Rights Watch and the International Human Rights Clinic (note 61); and CCW Convention, Commentaries by the USA (note 64).

[148] ICRC (note 33).

would trigger a new review.[149] Some states and experts have suggested that software modifications which alter the functioning of the weapon or its behaviour in a way that affects the application of IHL would most likely require a new review.[150] What that means in practice, however, remains debatable.[151] It would, therefore, be useful if states could further articulate their views on the parameters that might be deemed critical for that assessment.

The second practical question relates to how the review process should deal with the prospect of AWS becoming more adaptive with the progressive integration of machine-learning capabilities. The use of machine learning could open up the possibility for AWS to 'learn' (in the sense of being trained) from experience, that is, to partially parameterize themselves during or after each mission, based on data they collect during the military operation.[152] This prospect begs practical questions. How should legal reviews address such a prospect? Under what conditions may legal reviews consider such a capability lawful or unlawful? Would it require the introduction of continual re-review procedures that allow the state to conduct legal reviews of weapons in the field?[153] In the GGE discussion, a number of states have already expressed views on what they would consider to be the outer limits: self-learning capabilities that allow AWS, during use, to override mission parameters set by human commanders would be unlawful (and make no operational sense), and therefore should not be deployed.[154] The permissibility of learning capabilities below that threshold has not yet received the attention it deserves. It would be helpful if states and experts could form and express views on the foreseeable application of machine learning, notably the target-recognition and evaluation systems of AWS, which are retrained with new operational data between each military operation.

The geographical dimension: What needs to be known about the environment of use

The fourth set of complexities relates to the environment of use. AWS apply force based partly on triggers in their environment. Much of the determination of the system's anticipated performance, operation and effects depends on the foreseeable interactions between its technical dependencies and variables in the circumstance of use. This raises two practical methodological issues: (*a*) what the reviewer should know and foresee about the environment; and (*b*) how empirical evidence about the environment should be gathered from it.

[149] ICRC, *International Humanitarian Law and the Challenges of Contemporary Armed Conflicts: Recommitting to Protection in Armed Conflict on the 70th Anniversary of the Geneva Conventions* (ICRC: Geneva, 2019), p. 24; and Lewis (note 6).

[150] Brown and Metcalf (note 132), p. 133; Vazquez, A., 'LAWS and lawyers: Lethal autonomous weapons bring LOAC issues to the design table, and judge advocates need to be there', *Military Law Review*, vol. 228, no. 89 (Mar. 2020), p. 17; and CCW Convention, Commentary by Germany (note 65).

[151] E.g. Farrant and Ford raise the question as to whether a 'missile's software was upgraded to a neural network that would allow the missile to identify targets more accurately? Is this a new weapon that would trigger a weapon review, or is this simply a software upgrade but the same weapon system? What if the neural network is continuously learning?'; see Farrant and Ford (note 131), p. 404.

[152] Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edn (Prentice Hall: Englewood Cliffs, NJ, 2014), p. 5; Hagström, M., 'Military applications of machine learning and autonomous systems', ed. V. Boulanin, *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, vol. I, Euro-Atlantic Perspectives (SIPRI: Stockholm, May 2019); and ICRC, 'Autonomy, artificial intelligence and robotics: Technical aspects of human control', 20. Aug. 2019, p. 14.

[153] Hahn, E., 'Legal challenges with the review of weapons with autonomous capabilities', Presentation held at the SIPRI International Conference on Article 36 Review and Emerging Technologies', Stockholm, 21 Sep. 2018 (unpublished).

[154] CCW Convention, GGE LAWS, 'Contribution of Austria to the chair's request on the guiding principles on emerging technologies in the area of LAWS', Sep. 2020, p. 3; CCW Convention, Commentary by Germany (note 65), p. 2; and CCW Convention, Commentary by Switzerland (note 51), p. 4.

What the reviewer needs to know and foresee

Information about the environment of use is generally made available to legal reviewers as part of the description of the intended use of systems. In the case of AWS, information about the intended and expected environment of use is of paramount importance for the evaluation of the foreseeable effects of a weapon. The question then is how much and what type of information the reviewer should have in order to make the assessment.

Some practitioners consulted by the authors argued that it ultimately depends on the individual case of foreseeable use. The reviewer should demand more information about the operational parameters when it is deemed that environmental factors will play a significant role in the legal evaluation. Some experts have also pointed out that there is always a risk that information relevant to the question of legality may not be available to the reviewer, because either it is not able to be anticipated (e.g. the behaviour of an adversary) or its influence on the effects of the weapon is not known to be a relevant factor.[155]

However, to support the development of best practice, it may be useful if states could articulate their views on which environmental variables or conditions may demand particular attention from reviewers. These include weather conditions, clutter, civilians and civilian objects, and enemy countermeasures, as well as how far in terms of distance the person(s) responsible for administering an AWS may be from its operation and how dependent the person(s) might be on the communication link with the system or other systems to maintain situational awareness.

How to gather empirical evidence from the environment

Another methodological issue concerns how the review process needs to gather empirical evidence about the performance and foreseeable effects of a weapon in its intended circumstance of use. Two questions need to be distinguished in this regard. The first relates to where and from whom information should come. The extent to which the review process may solely rely on the weapon manufacturer or the country where the system was developed and acquired is a recurring question in the implementation of a legal review. Some states and experts believe that it may be sufficient to rely on documentation provided by the manufacturer or exporting countries. Others, including the ICRC, recommend that the review process rely on independent information, by gathering testing and evaluation conducted by the state or a trusted third party, for example. Although the conduct of independent testing and evaluation is widely considered as best practice, some states and experts do raise specific concerns in the case of AWS. Notably, the costs entailed by testing may be an inhibiting factor for states that have limited financial resources.[156] The conduct of computer simulation can help to reduce the costs of testing and evaluation, but it also raises questions about the basis on which empirical evidence is to be generated. Some states argue that computer simulations can evaluate the foreseeable performance, operation and effects of AWS in a larger number of situations than operational physical testing.

The second question relates to what empirical evidence should be considered regarding the performance and foreseeable effects of the weapon. The effectiveness of testing and evaluation, whether physical or computer simulated, is indeed closely

[155] McFarland refers to these as 'unknown unknowns'; see McFarland, T., *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law* (Cambridge University Press: Cambridge, 2020), p. 173.

[156] McClelland (note 36), p. 142; and Backstrom, A. and Henderson, I., 'New capabilities in warfare: An overview of contemporary technological developments and the associated legal and engineering issues in Article 36 weapon reviews', *International Review of the Red Cross*, vol. 94, no. 886 (June 2012).

connected to how realistically it represents the intended conditions of use. In that regard, technical experts acknowledge that it is impossible to include or model all the possible environmental variables—there will always remain some 'unknown unknowns'.[157] It would therefore be helpful if states could articulate their views on what environmental variables, as well as human factors, should be considered in testing and evaluation. Over time, this could contribute to the development of best practice for the conduct of testing and evaluation of AWS.[158]

Cross-cutting challenges

A challenge that cuts across the dimensions discussed above relates to the standard which the reviewer should use as a baseline for the legal review. Three issues can be distinguished here. First, many of the legal elements that the review needs to consider are not expressed in notions that are open to objective metrics (e.g. superfluous injury). In practice, this means that reviewers might exercise different interpretations of the legal outer limits.

Second, from a technical standpoint, the assessment of whether the foreseeable effects of AWS would be prohibited under IHL might not be answered by a simple yes or no proposition. For example, the question of whether AWS would be indiscriminate by nature 'cannot be achieved by a single pass/fail test, but rather it is subject to statistical confidence bounds'.[159] The question then is what statistical score is acceptable? Can an AWS that fails to identify the target 5 per cent of the time pass a legal review? According to some legal review practitioners there is no simple answer, as the failure rate needs to be judged on a case-by-case basis.[160] In the CCW debate, the question has been raised as to whether human performance should or should not be considered as a benchmark for assessing the foreseeable performance, operation and effects of an AWS. Some argue that a comparison with human performance might work in favour of a greater reliance on automated targeting technology, whereas others claim that any tolerance to failure should be lower for autonomous systems than humans.

Third, uncertainty, which is a design feature of AWS, is hard to quantify.[161] What level of certainty should the review—and ultimately the user—have about the foreseeable effect of the weapon in the circumstance of use? The question of whether information proxies can be designed for such an evaluation may deserve more attention.[162]

II. Legal advice

No matter the level of quality of a legal review, respecting and ensuring respect for IHL cannot be guaranteed through legal reviews alone. This is true for most weapons, means and methods of warfare, but for AWS in particular. AWS raise a series of legal complexities and their lawful use will often be difficult to determine on a general basis, as they are highly situation-dependent. In that context, Article 82, which requires states to make legal advice available 'when necessary', is of critical importance.[163]

[157] McFarland (note 155), p. 174.

[158] Boulanin (note 131), pp. 15–16; Boulanin and Verbruggen (note 136); and Goussac, N., 'Safety net or tangled web: Legal reviews of AI in weapons and war-fighting', Humanitarian Law and Policy blog post, 18 Apr. 2019.

[159] Backstrom and Henderson (note 156).

[160] Several practitioners, including those consulted as part of this project, have pointed the increased practice of producing 'yellow-lighted' reviews, that is, where systems are approved but only under certain conditions, notably AWS.

[161] Geiss, R., 'State control over the use of autonomous weapon systems: Risk management and state responsibility', eds R. Bartels et al., *Military Operations and the Notion of Control Under International Law* (Asser Press: The Hague, 2020), p. 442.

[162] McFarland (note 155), p. 174.

[163] See e.g. Vazquez (note 150); and Lewis (note 6).

Indeed, the provision of legal advice is likely to be an essential condition for the lawful use of AWS. However, compared to Article 36, Article 82 has been the subject of limited discussion in the GGE.[164] What it takes to perform the obligation of legal advice in relation to AWS remains unclear in certain respects. This section, therefore, discusses the interpretative questions that can be raised about its implementation across the four identified dimensions: personal, material, temporal and geographical.

The personal dimension: Who should receive legal advice and from whom

Who should receive legal advice

The primary recipients of the legal advice required under Article 82 are 'military commanders at the appropriate level'. They need to receive legal advice for two reasons: (*a*) to help them comply with IHL in specific operations; and (*b*) to develop and implement sound instructions to the armed forces on the application of IHL.

A practical question, however, relates to what constitutes 'the appropriate level' of command for the provision of legal advice in the case of AWS. This question is pertinent from the perspective of the discussion on whether the responsibility to make an IHL-mandated evaluation of the use of AWS resides with one or multiple commanders (see chapter 3). The ICRC's 1987 commentary related to Additional Protocol I notes that the legal consultation system should be implemented both close to the troops, for 'the essential purpose of participating in their instruction', and close to the headquarters, 'for consultation in the true sense of the word'.[165]

The current practice of states varies. Some endeavour to deploy legal advisers at almost all levels of command, some at a rather high level of command (at headquarters or military academies), while others prefer to keep them at the lowest level (close to the operations). For example, according to the United States Department of Defense, 'legal advisors actively participate in the entire planning process from joint intelligence preparation of the operational environment development to mission analysis to course of action, development and recommendation, through execution'.[166]

Who should provide legal advice

Article 82 does not specify the qualifications or background of the legal advisers, and the word 'qualified' was even removed during the drafting process of Additional Protocol I. As a result, legal advisers in Germany have a law background, for example, in Cameroon and France they hold a military rank, and in the UK they are a mixture of civilian and military experts.[167] This means legal advisers may possess varying knowledge of applicable law, and such lack of standardized practice could be a concern in the case of AWS.

To be effective, a legal adviser needs to understand the law and how the law applies in relation to the specific AWS being contemplated or used. As previously discussed, this is a complex exercise that requires both legal knowledge and technical understanding of a system's functioning and dependencies, in order to understand what might affect its performance and effect in the circumstance of use and, thereby, its legality. Unlike the individual(s) in charge of a legal review, the person dispensing legal advice may not have the time or physical ability to turn to a technical expert to gather additional

[164] Bolt, A., 'The use of autonomous weapons and the role of the legal advisor', ed. D. Saxon, *International Humanitarian Law and the Changing Technology of War* (Martinus Nijhoff: Leiden, 2013), p. 139.

[165] Commentary of 1987 to Additional Protocol I, 'Legal advisors in armed forces', para. 3351.

[166] Legal advisors should participate 'at all levels of command'; see US Armed Forces, 'Joint Operations', Joint Publication 3-0, 11 Aug. 2011, p. viii.

[167] Zidar, A. and Gauci, J. P. (eds), *The Role of Legal Advisers in International Law* (Brill-Nijhoff: Leiden, 2017), pp. 339–41.

knowledge to make their judgement. Therefore, it is critical that legal advisers receive proper technical training in addition to their legal education.

The question of whether technological progress would allow the substitution of legal advisers with artificial decision support systems has been raised in the relevant literature.[168] Although states have not addressed this specifically, it circles back to the contentious discussion of whether IHL-mandated evaluations necessarily demand human agency (see chapter 3).

The material, temporal and geographical dimensions: The content, time and location of legal advice

Article 82 states that legal advice may be made available to advise military commanders on: (*a*) the application of IHL in the conduct of military operations; and (*b*) providing adequate instructions to subordinates on the application of IHL. These two components pose different sets of issues with regard to the material, temporal and geographical dimensions of the obligation.

Providing legal advice to military commanders

There is little doubt that the provision of legal advice is necessary to help military commanders navigate the legal complexity that the use of AWS may involve. Nevertheless, there are at least two challenging issues related to how 'when necessary' may be interpreted. First, there is the issue of the temporal scope. Is the provision of legal advice restricted to the context of a military operation or does it extend further up in the design and development phase, such as the programming of an AWS?[169] This question is important because it determines the actual content of the legal advice. It also raises a practical question about the relationship between the legal review process and the provision of legal advice: where does the responsibility of the legal reviewer cease and that of the 'operational' legal adviser start?

The second issue is how 'when necessary' concretely applies in the operational use of AWS. At what juncture in the mission planning and execution should legal advisers be available? Is it sufficient to have legal advisers present before the activation phase, or is it necessary to continue to have them available during the operation of the weapon? Should they be available at all times? These questions circle back to the fundamental ones about when the evaluative judgements demanded by IHL need to be implemented (see chapter 3).

There seems little doubt that legal advisers need to be available in the mission programming phase.[170] The ICRC's 1987 commentary notes that legal advisers should assist in the 'preparation and development of plans, the choice of means, the determination of objectives, and the measures taken to achieve them'.[171] Whether a commander should be able to seek legal advice once an AWS has been activated and launched remains a divisive question, as states have expressed different views on what type of human agency must be exercised over the post-launch administration of an AWS. For some states that would require a case-by-case evaluation, while for others it would be best practice to have a legal adviser available throughout the mission.

On top of these issues of temporality, there are practical issues related to the complexity of the advice that a legal adviser needs to provide. For example, ensuring

[168] Bolt (note 164).

[169] Vazquez (note 150), p. 119.

[170] This argument is put forward by Lieutenant-Colonel Alexander Bolt, legal officer in the Office of the Judge Advocate General for the Department of National Defence and the Canadian Forces; see Bolt (note 164).

[171] Commentary of 1987 to Additional Protocol I, 'Legal advisers in armed forces' (note 165), Commentary on Article 82.

that advisers are properly trained to understand how AWS might perform in specific circumstances of use. What type and level of information does a legal adviser need to possess about the systems and environment of use to make a legal assessment? How is it possible to achieve reasonable certainty over the foreseeable effects of a weapon? In that regard, it would be helpful if states could elaborate their views on the variables that legal advisers would most likely need to consider and the competencies that they would need to possess in order to advise the commander on the legality of an attack involving AWS.

These could include variables such as: (*a*) how many civilians are expected to be in the vicinity of the target subject to attack or the environment in which the attack is taking place; (*b*) the range of the weapon, how its route will unfold and its navigation technology; (*c*) the length (in time) of the mission; (*d*) who or what the autonomous weapon can attack and with what means (in relation to the given rules of engagement and general targeting law); (*e*) whether the target could cease to be a military objective during the weapon's operation; and (*f*) whether changes in the environment would make it necessary to revisit the assessment of proportionality.

These variables may be useful to facilitate the role of legal advisers and help limit the risk of 'overburdening' them.[172] The risk of overburdening legal advisers when using complex and highly context-dependent, lawful AWS was also emphasized by some experts consulted in this project. Therefore, it is recommended that states consider and address this concern when elaborating their views on the role of legal advisers when using AWS.

Providing instructions to armed forces

The provision of legal advice may also help military commanders provide important instruction and guidance to the armed forces. With respect to AWS, the practical implications of this component ought to be further explored by states. Considering the fact that legal advisers could add value to the design of AWS, as well as to the training of operators, this arguably remains 'an unanswered question'.[173]

With regard to the design of AWS, it has been proposed that legal advisers could be mobilized to help the armed forces articulate technical and operational requirements to the manufacturers of weapons and components.[174] This could range from providing advice on the formulation of contracts and design specifications, to providing advice to programmers on the formulation of technical indicators that may or may not be used in the design of target recognition systems (e.g. understanding the nuances of the principle of distinction). In that regard, legal advisers could complement the role of legal reviewers, playing a proactive role in ensuring that the design requirements and choices fall within the boundaries of what is permitted by IHL and thereby helping foster IHL compliance.

With regard to the training of AWS users, legal advisers could play a role in the design of training manuals and courses, ensuring that aspects of IHL compliance are properly taken into consideration already at the training stage. More importantly, this could help improve users' awareness of what is demanded from them in the exercise of IHL provisions, and under what conditions the use of such systems may result in IHL violations. As a result, the provision of legal advice at that juncture could also facilitate the work of the legal advisers who are tasked with providing advice to military commanders in specific AWS operations.

[172] Boothby, B., 'Lethal autonomous weapons: What are the challenges posed to Article 36 weapons reviews?', International Institute of Humanitarian Law, 39th round table on current issues of international humanitarian law, Sanremo, 8–10 Sep. 2016, p. 2; and Henderson and Backstrom (note 156), p. 497.

[173] Vazquez (note 150), p. 119.

[174] Vazquez (note 150), p. 119.

III. Conclusion: Legal assessments of AWS

Legal reviews and legal advice are important mechanisms to ensure that AWS are capable of being used in compliance with IHL, and the performance of both obligations deserves further attention in the GGE. This is for several reasons, but notably because the distinction between the two mechanisms may become blurred when using AWS. Legal reviews of AWS may demand greater insight into the operational environment, while operational legal advice may require a greater understanding of the nature of the weapon system.[175] Recognising this overlap could lead to a better respect for IHL: the streamlining of legal advice throughout the development, acquisition, adoption and use of AWS, thus reducing the risk that issues fall through the cracks. However, this blurring of roles also risks creating an opportunity for responsibility shifting, where neither legal reviewer nor legal adviser considers themselves responsible for identifying and addressing violation risks. To avoid this outcome, states may wish to develop guidelines that address the respective responsibilities of legal reviewers and legal advisers, and their interaction with one another. This chapter has shown that there remain several legal and practical questions on which states could seek greater clarity and common understanding in order to ensure adequate legal assessments before, during and after the employment of an AWS, and thereby promote compliance with IHL.

[175] Vazquez (note 150).

5. Key issues concerning frameworks for state responsibility and individual criminal responsibility

An assumption underlying existing international responsibility frameworks is that, in an armed conflict, a legally wrongful act or omission can be traced back to at least one person, whether that person is an agent of a party to the conflict or is unaffiliated (see chapter 2). If unlawful conduct cannot be traced back, responsibility cannot be imposed. However, the way in which AWS are designed and intended to be operated may pose technical and legal challenges to that task.[176] As discussed in previous chapters, autonomy opens up the possibility for IHL provisions to be exercised by a complex web of human and artificial agents, based on automated processes and in expanded and more complex geographical and temporal circumstances. These factors, among others, have led some states and NGOs to raise concerns that the use of AWS could lead to an 'accountability gap': the risk that conduct potentially amounting to an IHL violation cannot be satisfactorily attributed, discerned or scrutinized and, as a result, an individual or state responsible for an IHL violation is not held to account or punished for it.[177]

As reflected in GGE Guiding Principle (b), it is human agents who are responsible for decisions on the use of AWS, whether in their capacity as individuals or state agents. This chapter, therefore, focuses on the question of how responsibility for IHL violations involving AWS may be traced back to individuals and states.[178] It considers how the rules on state responsibility and individual criminal responsibility (sections I and II, respectively) apply to AWS and to what extent the existing frameworks address the issues that AWS may raise (section III). Each section investigates the applicability of these frameworks across the four dimensions outlined in the report's introduction—personal, material, temporal and geographical—and identifies specific questions that merit further attention and elaboration in the GGE's discussions.

It should be noted that this chapter focuses on the responsibility frameworks applicable to states and individuals.[179] Where state conduct is involved, these frameworks are meant to be complementary and operate in parallel. However, the precise legal parameters pertaining to states under IHL are not fully synonymous with those pertaining to individuals, as individuals are also subject to the rules under individual criminal law (ICL). Further, this chapter does not examine the institutions and mechanisms available for holding an individual or state accountable for an IHL violation (including courts and tribunals, whether national, regional, international or hybrid)—these questions are examined elsewhere in the expert literature.[180]

[176] See e.g. Verdiesen, Santoni de Sio and Dignum (note 57), p. 13.

[177] Human Rights Watch and the International Human Rights Clinic, the Human Rights Program at Harvard Law School, *Mind the Gap: The Lack of Accountability for Killer Robots* (Human Rights Watch: 2015); Chengeta, T. 'Accountability gap: Autonomous weapon systems and modes of responsibility in international law', *Denver Journal of International Law and Policy*, vol. 45, no. 1 (Apr. 2020), pp. 4–11; Government of Pakistan, Statement to the UN Human Rights Council on behalf of the Organization of the Islamic Conference, 30 May 2013; Government of the Republic of Korea, Statement to the CCW Convention Meeting of Experts on LAWS, 13 Apr. 2015; and Verdiesen, Santoni de Sio and Dignum (note 57).

[178] CCW Convention (note 2), Annex IV, principles (b) and (d).

[179] It thereby leaves out other relevant discussions on e.g. corporate responsibility or responsibility of international organizations.

[180] During discussions at the GGE, the terms 'responsibility' and 'accountability' have, at times, been used interchangeably. Nevertheless, it is important to recall that they refer to related but distinct concepts. 'Human responsibility' can be considered as encompassing moral and ethical considerations, as well as legal obligations and expected conduct. 'Accountability' can be considered to relate to legal liability and legal consequences. The concepts are mutually reinforcing with clear and distinct lines of responsibility improving the accountability and attribution process. See CCW Convention, Joint commentary (note 25). For literature addressing accountability, see

I. State responsibility

The rules on state responsibility for internationally wrongful acts have not featured strongly in the GGE discussions, yet they are worthy of further discussion for many reasons.[181] First and foremost, they provide a framework for identifying the elements necessary to help prevent violations of IHL. For example, if states perform their general obligation to respect and ensure respect for IHL, there will arguably be fewer breaches of IHL, either carried out by state agents or unaffiliated individuals. Also, state responsibility may be more relevant than the rules on individual criminal responsibility for two reasons: (*a*) state responsibility applies whenever a state breaches IHL, not just in cases of grave breaches; and (*b*) states are responsible for omission, that is for merely failing to perform their international obligations, whereas higher, or at least different, standards of intent and knowledge must be met to incur individual criminal responsibility.

However, the precise contours of the state responsibility framework are not easy to detect and the introduction of AWS has not made it any easier. Two questions demand particular attention. On the one hand, there is the question of what the rules on state responsibility demand of a state (acting through all of its agents) for the lawful development and use of AWS: *who* needs to take *what* measures to respect and ensure respect for IHL, *when* and *where*? On the other hand, there is the question of *what* constitutes acts or omissions that are attributable to a state in the context of AWS, and *how* conduct (either act or omission) may be effectively traced back to a state.

Further elaboration of views by states seems warranted. This section helps identify the key questions going forward, considering how the four dimensions apply and interact with respect to AWS, and whether the existing framework is up to the task of attributing, discerning and scrutinizing state conduct involving AWS.

The personal dimension: Identifying who can exercise and implement a state's obligations under IHL

States are abstract entities that act through human agents, typically the members of its armed forces and people empowered with governmental authorities. As discussed above, AWS challenge the presumption that natural persons—that is, human agents—are the only valid agents in terms of exercising and implementing a state's IHL-related rights and obligations (see chapter 3). While the GGE has acknowledged the need to 'retain human responsibility', it is not necessarily clear whether an AWS may be conceptualized as exercising and implementing at least part of a state's IHL-related legal agency, or whether all of a state's legal agency must be placed in its human agents. States should elaborate their views on this aspect to clarify how state responsibility applies vis-à-vis the activities of an AWS.

Another issue on which clarification would be warranted is the question of what—and how many—human agents are responsible for respecting a state's IHL obligations in the use of AWS. Guiding Principle (d) states that 'Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW' requires a 'responsible chain of human command and control', but it leaves what

e.g. Sassoli, M., 'State responsibility for violations of international humanitarian law', *International Review of the Red Cross*, vol. 84, no. 846 (2002); Crawford, J. and Olleson, S., 'The character and forms of international responsibility', ed. M. Evans, *International Law* (Oxford University Press: Oxford, 2018); and Hammond, D. N., 'Autonomous weapons and the problem of state accountability', *Chicago Journal of International Law*, vol. 15, no. 2 (2015).

[181] With some exceptions, see e.g. CCW Convention (note 27); CCW Convention, GGE LAWS, 'Possible outcome of 2019 Group of Governmental Experts and future actions of international community on Lethal Autonomous Weapons Systems', Working paper submitted by Japan, CCW/GGE.1/2019/WP.3, 22 Mar. 2019; CCW Convention, Commentaries by the USA (note 64); and CCW Convention, Joint commentary (note 25).

constitutes such a chain unaddressed.[182] Some states, such as Australia, have explained their approaches to devising and administering a system to comply with applicable IHL across all the relevant stages concerning an AWS, from design to employment. These systems typically involve numerous individual human agents (including engineers, weaponeers, legal advisers and a commander) acting in combination as agents of the state.[183] It would be beneficial if states could elaborate their views on who the responsible agents are, and perhaps more importantly: (*a*) what should be demanded of agents (e.g. knowledge, training and facilities) to satisfactorily secure respect for a state's obligations under IHL; and (*b*) what standard of behaviour by agents would make state responsibility apply.

The material dimension: The scope, nature and content of a state's obligations under IHL

As discussed above, states are obliged to respect and ensure respect for IHL (see chapter 2). The first, and self-evident, aspect of this obligation is that state agents are required to respect the primary rules of IHL, including the principles of distinction, proportionality and precautions.[184] The second aspect pertains to the requirement that states take appropriate measures to ensure respect for IHL.[185] These include disseminating information about IHL in military instructions, making legal advisers available when necessary and supervising the implementation of IHL obligations by subordinates.[186] In the GGE, it may be useful for states to elaborate on what particular measures are required to 'respect and ensure respect for IHL' in the development and use of AWS. That may include addressing what is required to exercise due diligence: to prevent, suppress and punish violations by instituting sufficient measures to reliably foresee, administer and trace the operation, performance and effects of AWS.

To trigger state responsibility, there needs to be a breach of one of the state's international obligations, either through commission or omission. What constitutes a breach of a state's IHL obligations in the development and use of AWS still needs to be clarified in certain key respects. A critical question is whether a state could be held responsible for not (or poorly) implementing the measures listed above, to the extent that it could have prevented conduct using an AWS that resulted in a prima facie violation of IHL (e.g. launching an attack anticipated to result in disproportionate harm to the civilian population). The USA, for example, has expressed the view that state responsibility would apply for violations committed by individuals if it can be shown to be the result of inadequate supervision or training.[187] However, establishing that causal link maybe difficult in the case of AWS, since the presence of autonomy is likely to blur the line connecting users to the effects of force. It may be generally difficult to discern and trace back whether the cause of a harmful incident lies in the behaviour of the user or of the technology.

States and experts have expressed concern that the technical complexity, opacity and uncertainties surrounding how an AWS might interact with the environment of use (so-called edge cases or corners cases) might make it easier for states to evade

[182] CCW Convention (note 2), Annex IV.

[183] CCW Convention (note 70).

[184] 1949 Geneva Conventions common Article 1: 'The High Contracting Parties undertake to respect and to ensure respect for the present Convention in all circumstances'.

[185] See 1949 Geneva Conventions common Article 2(1): 'In addition to the provisions which shall be implemented in peacetime'; Geneva Convention (I) (note 44), Article 45; and Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, opened for signature 12 Aug. 1949, entered into force 21 Oct. 1951, Article 46; and Additional Protocol I (note 11), Article 80.

[186] ICRC, Rule 141 (note 38); and Hathaway, O. A. et al., 'Ensuring responsibility: Common Article 1 and state responsibility for non-state actors', *Texas Law Review*, vol. 95, no. 539 (2017), p. 576.

[187] See US Air Force Pamphlet, §21, referenced in ICRC, Rule 149 (note 41), footnote 22.

responsibility in two possible ways.[188] The first is for the state to claim, on the basis that most accidents do not entail responsibility, that no breach occurred because an unforeseen event caused an accident with the AWS. The second is for the state to admit that a violation occurred with an AWS, but to invoke *force majeure* as grounds to preclude the wrongfulness of a breach caused by 'an unforeseen event, beyond the control of the State'.[189] The question of what states should be expected to foresee in the use of AWS is critical in that context.

From a legal standpoint, consequences or effects that were reasonably foreseeable by one or more state agents, but were either (*a*) not foreseen or (*b*) foreseen but ignored, would presumptively engage state responsibility. According to some states and legal specialists, including those consulted for this report, consequences or effects that are not reasonably foreseeable (i.e. so-called true accidents) do not necessarily engage the responsibility of the state, except perhaps concerning a relatively small subset of state conduct.[190]

It would therefore be beneficial if the GGE could clarify what types of 'failures' in AWS would qualify as unintended or foreseeable accidents, respectively. Arguably, this could simplify the task of distinguishing between a 'normal' accident and non-fulfilment of duty by agents of the state. However, some experts have claimed that such a measure might not be sufficient. From their standpoint, the inherently unpredictable nature of AWS—and the corresponding risks associated with their use—demand that the framework of state responsibility extends to a strict liability regime.[191] This would arguably incentivize states to limit the use of AWS to restricted and controlled environments of use. The prospect of adding another liability regime is outside the scope of this project, but is nonetheless worth exploring further elsewhere.

The temporal dimension: At what points in time do a state's obligation to respect and ensure respect for IHL begin and end

The obligation of states to respect and ensure respect for IHL extends beyond solely the time frame of a military operation. Article 36 of Additional Protocol I, for example, applies as early as in the study and development of a new weapon system (see chapter 4). A critical question is whether state responsibility can be engaged for decisions made as early as in the design stage, and this has not been sufficiently addressed by states in the GGE. A key legal issue in that context is how far back in time states are bound by their the obligation to take feasible precautions. Some experts consulted for this project argued that it extends as far back as to the procurement and design stage, while others claimed that the obligation only applies at the operational stage. These conflicting interpretations may stem from the reference in Article 57(1)(a) to 'those who plan and decide upon an attack', which may be interpreted as only applying to commanders' duties during an attack; whereas the general provision under Article 57

[188] Holland Michel, A., 'Unknown unknowns: Data issues and military autonomous systems', UNIDIR, 2021; and Geiss (note 161).

[189] International Law Commission (note 42), Article 23; *Force majeure*: '1. The wrongfulness of an act of a State not in conformity with an international obligation of that State is precluded if the act is due to force majeure, that is the occurrence of an irresistible force or of an unforeseen event, beyond the control of the State, making it materially impossible in the circumstances to perform the obligation. 2. Paragraph 1 does not apply if: (a) the situation of force majeure is due, either alone or in combination with other factors, to the conduct of the State invoking it; or (b) the State has assumed the risk of that situation occurring.'

[190] Dörmann, K., and Serralvo, J., 'Common Article 1 to the Geneva Conventions and the obligation to prevent international humanitarian law violations', *International Review of the Red Cross*, vol. 96, no. 895/896 (2014), p. 730; and CCW Convention, Commentaries by the USA (note 64), p. 5, (7).

[191] Crootof, R., 'War torts: Accountability for autonomous weapons', *University of Pennsylvania Law Review*, vol. 164, no. 6 (May 2016); Fuzaylova, E., 'War torts, autonomous weapon systems, and liability: Why a limited strict liability tort regime should be implemented', *Cardozo Law Review*, vol. 40, no. 3 (2019); and experts consulted as part of this project.

refers to the obligation to take 'constant care' in 'military operations', which arguably applies more broadly and over a longer period.

Further discussions on the temporality of the obligation to take all feasible precautions could clarify how far back in time state responsibility may be engaged for conduct related to the development and use of AWS.[192] From a more practical standpoint, it is likely that tracing a violation of IHL back to inadequate supervision, training or design decisions will present some challenges in the absence of specific AWS standards. It would, therefore, be beneficial if states could further articulate what should be deemed norms and best practice for the responsible development and use of technologies in the area of AWS.

The geographical dimension: In relation to what locations must a state perform its IHL obligations?

The characteristics of environments of use are important, because they may trigger different obligations for state agents involved in the development and use of AWS. The rules on land warfare and those on naval warfare, for example, may place different limitations on the ways in which AWS can be designed and used. Consequently, they provide different bases on which state responsibility can be engaged. Therefore, it could be beneficial if the GGE debate paid greater attention to the legal obligations that apply to different environments. Such clarification matters because the use of AWS will likely take place across different domains and with greater geographical distance between the state agent and the effects of the use of force.

Looking ahead

Overall, it could be argued that the framework of state responsibility contains unexplored potential when identifying and establishing limits on the development and use of AWS. Indeed, the rules on state responsibility allow for an evaluation of whether existing primary IHL rules and accountability institutions provide a sufficient basis to govern the development and use of AWS. Therefore, GGE discussions ought to focus partly on defining with greater specificity what is required to respect and ensure respect for a state's obligations under IHL in relation to conduct involving AWS. Specifically, states should form and express views on: (*a*) which agent or agents are responsible for respecting and ensuring respect for a state's IHL obligations in the development and use of AWS; and (*b*) at what points in time, in relation to what locations and regarding what specific sets of activities and decisions involving an AWS state responsibility may arise. Clarifying uncertainties surrounding the already abstract provisions of exercising due diligence, taking constant care and taking all feasible precautions will be critical in that context.

States should also express views on the nature of IHL violations, including how the concept of 'true accidents' relates to the use of AWS. Moreover, states should address how the ability to foresee, administer and trace conduct involving AWS can be guaranteed through their state agents, in order to ensure compliance with IHL. Answering these questions can help states detect existing limits with greater specificity and identify potential areas for further regulatory development.

[192] Amoroso and Benedetta (note 47), p. 225.

II. Individual criminal responsibility

Although the applicability of individual criminal responsibility has been subject to greater attention in the GGE than state responsibility, a range of questions remains to be answered. As outlined above, individual criminal responsibility is based on four elements: (*a*) a serious violation of IHL; (*b*) a material element; (*c*) a mental element; and (*d*) that the conduct was carried out through one of the established modes of responsibility (see chapter 2). Without the establishment of all four elements, individual criminal responsibility cannot be imposed. Moreover, these four elements need to be attributable to an individual. If the conduct cannot be traced back to an individual, it is not possible to impose criminal responsibility.

A central issue for the discussion on AWS and individual criminal responsibility is whether, and to what extent, the presence of autonomy undermines the possibility to establish these four elements and connect them to one or more specific individuals. As previously discussed, the presence of autonomy impacts the way in which individuals may engage in unlawful conduct across multiple dimensions: personal, material, temporal and geographical. This section explores how AWS affect the ability to: (*a*) attribute conduct to specific individuals; (*b*) discern what conduct was engaged; and (*c*) trace and scrutinize the temporal and geographical circumstances in which the conduct took place.

Personal dimension: Who can be held responsible?

The question of whether AWS can be held responsible for IHL violations has already been answered by states. As reflected in both the GGE's 2019 guiding principles and the Rome Statute, it is well established that individual criminal responsibility applies only to humans, not artificial agents.[193] The debate has consequently moved on, to the question of how the responsible individual(s) may be identified—with two dimensions to that question.

First, as discussed previously, there is the interpretative issue of use of force decisions. The question is whether, from a legal perspective, a use of force decision ultimately resides with one individual—the commander or another responsible person—or can be distributed across multiple individuals, including as part of a command-and-control chain. Neither IHL nor the Rome Statute provides a clear answer, yet this question is important because it determines whether states should focus their efforts on identifying unlawful conduct on the part of the commander or also of the multiple people necessarily involved in the use of AWS—spanning from planners to weapon programmers and operators.[194]

Second, there is the practical difficulty of connecting individual conduct to the four elements on which individual criminal responsibility is based. This is particularly challenging in events where multiple individuals may contribute to the employment of an AWS. At a practical level, there is the difficulty of tracing decisions made by different individuals at different points in time, as well as how these individual decisions may have interacted. In addition, there is the difficulty of discerning how individuals' decisions and their consequences correspond to the provisions as set out in the Rome Statute. How can they be materially connected to unlawful conduct, can

[193] Guiding Principle (b): 'Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system', CCW Convention (note 2), Annex IV. The Rome Statute (note 49), Article 25(1) addresses 'natural persons'.

[194] See e.g. Chengeta (note 177); CCW Convention, Commentary by Switzerland (note 51); CCW Convention, Commentaries by Portugal (note 9); CCW Convention, Commentary by Germany (note 65); Schulzke, M., 'Autonomous weapons and distributed responsibility', *Philosophy and Technology*, vol. 26, no. 2 (2013), p. 213; Henderson, Keane and Liddy (note 68); and Amoroso and Benedetta (note 47), pp. 215, 219.

the mental element be established, and does the conduct satisfy one of the modes of responsibility? As the following subsections show, it may be difficult to prove that the decisions and actions of the designer and programmer—as well as the commander and their subordinate—satisfy the different elements required to establish individual criminal responsibility.

Material dimension: Establishing the elements of a war crime

The presence of autonomy in weapon systems raises both old and new issues with regard to how the elements of a war crime can be established and linked together. Recalling the elements of a war crime mentioned above, these include a material element, a mental element and a mode of responsibility.

A novel feature of autonomy is that it limits the number of situations in which both the material element and the mental element can be connected, which in turn reduces the legal basis on which individual criminal responsibility can be established and imposed. AWS are, by definition, preprogrammed weapons: a user determines in advance the parameters of a mission and the type of targets. While it cannot be excluded, it is unlikely that a user will intentionally programme an AWS to target civilians or produce indiscriminate and disproportionate effects. Therefore, AWS attacks that result in civilian harm or disproportionate effects are more likely to be associated with a failure by the user to take necessary precautions (e.g. failing to seek information on the objects or persons subject to attack or failing to take into account the specific environment of use) than with an intentional and knowing violation of the principle distinction or proportionality.[195]

If the legal basis for individual criminal responsibility is limited to failure to take necessary precautions, the number and range of war crimes that may arise in relation to the employment of AWS will be significantly reduced. That reduction in turn would impact what states can, or must do, to suppress violations of IHL.[196] Indeed, IHL and ICL instruments treat the violation of the principle of precautions somewhat differently. IHL instruments do not provide a clear basis on which to characterize a violation of the principle of precautions as a war crime, whereas under the Rome Statute such a violation can be part of the war crime of 'intentionally targeting civilians'.[197] This discrepancy may have practical consequences with respect to national investigations and prosecutions of war crimes involving the use of AWS. States that have not incorporated the war crime listed in the Rome Statute into their respective domestic systems may not have a legal basis to prosecute a failure to take necessary precautions. This may significantly limit the ability of such states to impose individual criminal responsibility with respect to attacks involving the use of AWS.

Furthermore, the presence of autonomy reopens an unresolved legal debate about what standard of behaviour satisfies the mental element, or *mens rea*. For the user of an AWS to be held responsible for a war crime, it is not sufficient that the conduct results in civilian harm. Under Additional Protocol I, the proscribed conduct amounts to a war crime if it can be established that the user (e.g. the commander, designer or developer) acted wilfully, in violation of a relevant provision of that instrument, and caused death or serious injury to body or health, including by: (*a*) making the civilian population or individual civilians the object of attack; (*b*) launching an indiscriminate

[195] Bo (note 67), p. 18.

[196] It is argued that the individual responsibility gap, among others, lies in the difficulties of ascribing responsibility for omissions. E.g. Bo points to the fact that 'Within the Rome Statute there is no general provision on commission by omission and the status of criminal responsibility by omission is more uncertain'; See Bo, M., 'Meaningful human control over autonomous weapon systems: An (international) criminal law account', Opinio Juris, 18 Dec. 2020.

[197] Rome Statute (note 49), Article 8(2)(b)(i); and Doermann, K., *Elements of War Crimes Under the Rome Statute of the International Criminal Court* (Cambridge University Press: Cambridge, 2003), pp. 131–32.

attack affecting the civilian population or civilian objects in the knowledge that such attack will cause excessive loss of life, injury to civilians or damage to civilian objects; or (*c*) making a person the object of attack in the knowledge that the person is *hors de combat*.[198] Tracing back and assessing whether the user acted intentionally or wilfully depends partly on what the user knew—or should have known—in the circumstances at the time. Due to the lack of foreseeability associated with the use of AWS, this task may become increasingly challenging.[199]

An old but relevant IHL question in this context is whether negligent or reckless behaviour may qualify as 'wilful' behaviour.[200] States have expressed different views on that issue in debates that precede the CCE process on AWS.[201] It would be beneficial if states could return to this question in the specific context of AWS for two reasons. First, this question has direct implications for the level of unpredictability that may be deemed tolerable in the use of AWS.[202] In the expert literature, it is argued that if states were to interpret the concept of wilfulness too narrowly (i.e. equate it to direct intent), that would create more room for AWS users to engage in risk-taking behaviour. Alternatively, if they were to interpret it more broadly to include recklessness and negligence, risk-taking behaviour would be disincentivized, as it 'could have important effects in terms of increasing standards of precautions and deterrence' with respect to the use of AWS.[203] Second, this question of interpretation has concrete implications for the possibility to hold individuals involved in the programming of AWS criminally responsible. While it may be unrealistic (and practically impossible) to establish that a designer intentionally designed an AWS to target civilians indiscriminately long before the launch of an attack, the designer may be held responsible for negligent or reckless behaviour.[204] From that standpoint, addressing and clarifying the law in relation to risk-taking behaviour could be critical to ensure individual criminal responsibility in the use of AWS.[205]

Moreover, the diffused process associated with the use of AWS has led a number of states and experts to call for an examination of the adequacy of the existing modes of responsibility as set out in the Rome Statute.[206] One issue that states should seek to clarify is how the mode of command responsibility applies in cases where the actions of the (human) subordinate deploying an AWS result in violations of IHL. On what basis can it be established that the commander (*a*) failed to exercise effective control over the subordinate, and (*b*) knew—or should have known—that the subordinate would use the AWS in violation of IHL and failed to prevent or stop their actions?[207]

[198] Additional Protocol I (note 11), articles 11, 85.

[199] According to Bo, 'Unintentional attacks against civilians stemming from the unforeseen decisions of an autonomous process would escape responsibility under the current legal framework'; see Bo (note 87), p. 25. See also Amoroso and Benedetta (note 47), p. 220.

[200] See McDougall, C., 'Autonomous weapon systems and accountability: Putting the cart before the horse', *Melbourne Journal Of International Law*, vol. 20, no. 1 (2019), p. 10; Ohlin, J. D., 'Targeting and the concept of intent', *Michigan Journal of International Law*, vol. 35, no. 1 (2013), pp. 79, 86; Eser, A., 'Mental elements—Mistake of fact and mistake of law', eds A. Cassese, P. Gaeta and J. R. W. D. Jones, *The Rome Statute of the International Criminal Court: A Commentary* (Oxford University Press: New York, 2002), p. 899; and Henderson, Keane and Liddy (note 68), p. 17.

[201] Some states have already identified *mens rea*, negligence and recklessness as possible challenges with respect to AWS; see CCW Convention, Joint commentary (note 25), p. 5.

[202] Bo (note 67), p. 25.

[203] Bo (note 67), p. 25.

[204] CCW Convention, Joint commentary (note 25).

[205] Bo (note 87).

[206] Rome Statute (note 49), Article 25(3)(a)–(c).

[207] Chengeta (note 177), p. 50.

Temporal dimension: When did the violation take place?

The case of AWS supposes that decisions about attacks and the use of force may be made a long time in advance, as early as in the development phase of the weapon, and this temporal aspect raises several issues.

One issue it raises is the question of how far back in time individual criminal responsibility may be traced. In the GGE, this has been discussed in relation to the action of the individual engaged in the development, programming and design of an AWS.[208] Some states and experts have argued that designers and programmers should also bear responsibility for unlawful harm resulting from the use of AWS in armed conflict, as they are—through their design choices—in a position to significantly influence an operator's ability to use an AWS in compliance with IHL (e.g. by deciding the kinds of actions the system can carry out). From a legal standpoint, it has been argued that their involvement could qualify under the modes of responsibility listed in Article 25(3)(c) of the Rome Statute: 'For the purpose of facilitating the commission of such a crime, [the person] *aids, abets or otherwise assists* in its commission or its attempted commission, including *providing the means for its commission*.'[209]

However, that view has been challenged by a number of states and experts, who point out that while war crimes may, in theory, be committed by any person, not only by military actors, there are several barriers to 'developer responsibility'.[210] One barrier is that their activities may take place outside the temporal context of an armed conflict.[211] Another is the challenge of their conduct or the consequences of their actions satisfying the material and mental elements of a war crime, including acting with intent and knowledge.[212]

Nevertheless, the question of developer responsibility may gain greater prominence as autonomous technologies develop. Due to the exponentially increasing complexity of AWS, states may find it necessary to reconsider where the balance of responsibility lies along the chain that connects those who create a weapon with those who use it. States will need to consider whether they wish to ensure the possibility of holding those who design or develop an AWS responsible for possible war crimes. Given the barriers that exist to holding designers or developers responsible, doing so may require the development of a specific normative framework suited to the particular characteristics of designer/developer behaviour, namely that their conduct occurs before—sometimes long before—the employment of a weapon system (i.e. it is preparatory or facilitative).

Another issue is more practical and relates to how conduct that engages individual criminal responsibility may be traced back in time. The performance of an AWS may result from multiple decisions at multiple points in time. Therefore, discerning at what

[208] CCW Convention, GGE LAWS, 'Autonomy in weapon systems', Working paper by the USA, CCW/GGE.1/2017/WP.6, 10 Nov. 2017; ICRC (note 149); Thurnher, J., 'Examining autonomous weapon systems form a law of armed conflict perspective', eds H. Nasu and R. McLaughlin, R., *New Technologies and the Law of Armed Conflict* (Asser Press: The Hague, 2014), p. 225; Corn, G. S., 'Autonomous weapon systems: Managing the inevitability of "taking the man out of the loop"', Social Science Research Network, 14 June 2014; and McFarland, T. and McCormack, T., 'Mind the gap: Can developers of autonomous weapons systems be liable for war crimes?', *International Law Studies*, vol. 90, no. 361 (2014), p. 375.

[209] Rome Statute (note 49), Article 25(3)(c).

[210] See e.g. the discussion of developer responsibility: McDougall, C., 'Autonomous weapon systems and accountability', *Melbourne Journal of International Law*, vol. 20 (2019), p. 13; McFarland and McCormack (note 208), p. 375; and CCW Convention, Statement by the USA (note 208). See also ICRC (note 149); Sassoli, M., 'Autonomous weapons and international humanitarian law: Advantages, open technical questions and legal issues to be clarified', *International Law Studies*, vol. 90, no. 308 (2014), p. 325; Thurnher (note 208), p. 225; and Corn (note 208). In contrast, Sparrow has argued that to 'hold the programmers responsible for the actions of their creation, once it is autonomous, would be analogous to holding parents responsible for the actions of their children once they have left their care'; see Sparrow, R., 'Killer robots', *Journal of Applied Philosophy*, vol. 24, no. 1 (2007), p. 70.

[211] McFarland and McCormack (note 208), pp. 372–74; and Amoroso and Benedetta (note 47), p. 219.

[212] McFarland (note 155), pp. 153–161; and Amoroso and Benedetta (note 47), p. 219.

point in time individuals formed and acted with intent and knowledge to commit a war crime may be difficult. Consequently, it would be useful for states to discuss practical measures that could allow them to trace and scrutinize the decisions and actions taken over time to foresee and administer the operation of a weapon, for example, in the form of an (electronic) paper trail for the decision chain and on-board black-box systems that record sensor input, communications and decisions.[213]

Geographical dimension: Where did the violation take place?

The final dimension relates to how the nature of the environment and the type and degree of interaction between the user and the environment of use affect the possibility to establish both the material and the mental element of individual criminal responsibility related to AWS. As previously discussed, the characteristics of the environment of use are relevant for the evaluation of humanitarian risks associated with the deployment of an AWS. Changes in environmental conditions and enemy countermeasures could affect the performance of the system and lead to fatal consequences. Two types of risks may emerge in this connection. One is that military commanders might be able to mask their misconduct more easily by pointing to purported technical failures. Another is that an enemy could induce technical failures resulting in potentially unlawful harm (e.g. through cyber, electronic or spoofing attacks, or data poisoning).[214]

In that context, it would be helpful if states could identify what standards of knowledge and behaviour are required to foresee and administer the operation, performance and effects of an AWS in its intended environment of use. These standards would make it easier to discern, after the fact, whether a harmful incident resulted from an accident or whether it was an intentional act by the AWS user.[215] To that end, states may want to consider the following questions: what level of knowledge and understanding of the environment is expected from the user? How should knowledge and understanding be applied to define the spatial limits for an AWS mission? At what point can it be considered that the spatial envelope was too broadly defined and might amount to unlawful conduct in terms of the material element (e.g. a violation of the principle of distinction or proportionality) and the mental element (e.g. being aware that a prohibited consequence will occur in the ordinary course of events)?

Looking ahead

In summary, the presence of autonomy in weapons has the potential to transform core assumptions underlying how humans respect or violate IHL, including as relates to modes of responsibility or mental elements of war crimes. It could, in turn, also affect some of the conditions necessary to assess and impose individual responsibility for IHL violations. As this section shows, there are several difficult (and important) questions that need to be considered in order to attribute, discern or scrutinize conduct involving an AWS that may amount to a war crime. In that regard, the debate on human–machine interaction provides an opportunity for states to identify concrete

[213] See e.g. Tubella, A. et al., 'Governance by glass-box: Implementing transparent moral bounds for AI behaviour', International Joint Conference on Artificial Intelligence, Ume University, Aug. 2019; and Gubrud, M. and Altman, J., 'Compliance measures for an autonomous weapons convention', International Committee for Robot Arms Control (ICRAC) Working Paper, May 2013.

[214] It remains an open question as to whether users can be held responsible for harm resulting from an adversary's attempt to trick the systems into failing. Scharre, P., 'Autonomous Weapons and Operational Risk', Center for a New American Security (CNAS) Working Paper, Feb. 2016.

[215] CCW Convention, Commentary by the Netherlands (note 2), p. 2; CCW Convention, Joint commentary (note 25); CCW Convention, Commentaries by Portugal (note 9); and CCW Convention, GGE LAWS, 'Commonalities in national commentaries on guiding principles', 2020, §9.

measures for tracing conduct associated with AWS. Such measures could help to address the possible ways in which people might evade individual responsibility for a harmful incident resulting from AWS use.

III. Conclusion: Retaining human responsibility

While the primary rules of IHL guide the lawful development and use of AWS, the secondary frameworks of responsibility help to ensure that states and individuals respect those rules and that legal institutions exist to hold violators to account. Although this report addresses both frameworks in the same chapter—state responsibility and individual criminal responsibility—it is important to distinguish between the two. Some incidents can constitute war crimes but do not engage state responsibility, some can engage state responsibility but not attract war crime liability, and some can trigger both or neither.

The central issue here is whether the existing state responsibility and individual criminal responsibility frameworks provide sufficiently clear and applicable institutions and norms in order to attribute, discern and scrutinize conduct involving an AWS, including the imposition of individual and/or state responsibility when an IHL violation has occurred. As discussed in this chapter, these issues relate to core assumptions concerning notions of the legal agency of states and individuals in armed conflicts. The extent to which existing responsibility frameworks are adequate with respect to an AWS depends partly on how the frameworks are interpreted and applied. Substantial discussions on these issues have, in comparison with discussions on the primary rules of IHL, been relatively sparse in the GGE. The questions posed in this chapter are therefore aimed at guiding future discussions by identifying key challenges and questions to be answered.

A cross-cutting challenge that states need to address in greater detail is what combinations of the four dimensions underlying conduct involving AWS—personal, material, temporal and geographical—would be prohibited, either because they reflect a lack of respect for IHL or because they preclude imposition of state or individual criminal responsibility when an IHL violation occurs. Building on the 11 guiding principles already agreed in the GGE, states could provide more clarity on whether, for military operations involving AWS, the responsibility of the state can and should be placed on a single human agent; and, if so, what should be demanded of that person (e.g. in terms of their legal, technical and situational knowledge, training or facilities) to satisfactorily secure respect for IHL provisions applying to states.

As a touchstone for those discussions, it may be useful to elaborate on the preconditions necessary, across various potential combinations of the four dimensions, to ensure that: (*a*) one or more human agents can and will reliably foresee the effects and performance of an AWS in the anticipated circumstances of use and throughout the anticipated timeline of operation; (*b*) the AWS can and will be satisfactorily administered during its operation; and (*c*) the performance, effects and operation of the AWS can and will be traced to one or more responsible human agents after the fact.

Sharing views on these aspects may help strengthen debates in the GGE about existing legal principles, rules and standards concerning foreseeability, causal control and responsibility related to military operations involving AWS. In that context, the ability to trace conduct will require particular attention. While it has not featured prominently in the GGE debate, it is arguably an essential condition for securing respect for IHL. Both an inability to trace and an absence of tracing preclude an assessment of IHL compliance and thereby impair the potential to impose responsibility and hold people and states accountable for violations.

6. Key findings and recommendations

This report has presented the findings of SIPRI's one-year research project on 'Autonomous Weapon Systems and International Humanitarian Law'. It has been designed to support both national and international discussions on AWS, including at the GGE. It has mapped relevant rules of IHL and identified key questions of IHL interpretation and application with regard to the development and use of AWS. The report has been informed by desk research conducted by the authors, as well as a series of virtual expert discussions that SIPRI held in June 2020 under the Chatham House Rule. This chapter summarizes the key findings (section I) and recommendations (section II) from the report.

I. Key findings

1. The development and use of AWS is not unlimited

IHL already places limits on the development and use of AWS. It prohibits any weapon—including AWS—that: (*a*) has characteristics prohibited by a weapon treaty or customary law; (*b*) is of a nature to cause superfluous injury or unnecessary suffering; (*c*) is indiscriminate by nature; or (*d*) is intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment.

The employment of an AWS that has not been subject to the aforementioned prohibition is nevertheless regulated by the general prohibitions and rules on weapons, means and methods of warfare. Of particular relevance for AWS are the rules governing the conduct of hostilities, notably the principles of distinction, proportionality and precautions. These rules prohibit AWS attacks that are not specifically directed at military objectives or that may be expected to cause civilian harm and damage which would be excessive in relation to the concrete and direct military advantage anticipated. The principle of precautions demands that those who plan or decide on an attack using AWS should: (*a*) take all feasible precautions to avoid, and in any event minimize, incidental harm to civilians and damage to civilian objects; and (*b*) cancel or suspend the attack if it becomes apparent that the target is not a military objective, or that the attack may result in excessive incidental harm.

IHL requires states to respect and ensure respect for IHL, including by preventing, suppressing and punishing IHL violations, regardless of what weapons, means or methods are used. Under the framework of state responsibility for internationally wrongful acts, states bear responsibility for violations of IHL arising from the employment of an AWS as long as the conduct is attributable to the state through its agents. Under the framework of individual criminal responsibility, including both IHL and ICL, individuals bear responsibility for grave violations amounting to war crimes that they commit, contribute to, order or fail to prevent. To hold states or individuals responsible, it must be technically possible to trace whether a violation of IHL is the result of conduct and/or decisions made by specific individuals (including state agents), and to what extent unlawful conduct occurred or unlawful consequences arose due to the culpable fault of one or more natural persons.

2. Respect for IHL presupposes the ability to foresee, administer and trace the operation, performance and effects of AWS

In order to respect IHL, the development and use of AWS arguably must satisfy three conditions. First, it must be possible to reliably *foresee* whether the effects of an AWS

would in some or all circumstances contravene specific and/or general prohibitions and restrictions on weapons, means and methods of warfare. The employment of an AWS whose operation, behaviour and effects cannot be sufficiently predicted is likely to be unlawful. A legally problematic lack of foreseeability could be caused by a design feature, which makes the system's behaviour inherently unpredictable (e.g. online learning capabilities), or if due diligence was not followed in the development and acquisition process, meaning the system's performance and operation were not assessed through adequate testing and evaluation. A legally problematic lack of foreseeability may also result from a user's decision to deploy an AWS in an environment of use that is itself not sufficiently predictable. From that standpoint, the conduct of legal reviews and the provision of legal advice in the development and use stages of AWS are critical (yet insufficient on their own) measures for IHL compliance.

Second, it must be possible to satisfactorily *administer* an AWS during its operation, in a manner that is consistent with the rules governing the conduct of hostilities. The use of an AWS whose operation, behaviour and effects cannot be limited according to IHL, notably according to the principles of distinction, proportionality and precautions, would be unlawful. Limits can be placed on an AWS through controls on the system's parameter of use, structured forms of human–machine interaction during operational use and some environmental controls.[216]

Third, it must be possible to *trace* the operation, behaviour and effects of an AWS back to the relevant human agent(s). The employment of an AWS that cannot be satisfactorily attributed, discerned or scrutinized would preclude assessing and imposing state responsibility and/or individual criminal responsibility for violations. While the GGE has established that responsibility to comply with IHL remains with humans and cannot be delegated to machines, it remains to be clarified how human responsibility ought to be secured through human–machine interaction. This is a critical interpretative question that states need to further articulate their views on.

3. IHL compliance depends partly on threshold questions concerning the required type and degree of human–machine interaction

The three conditions underlying respect for IHL—the ability to foresee, administer and trace the operation, behaviour and effects of AWS—require a holistic approach to human–machine interaction, one that is characterized by the application of practical measures at multiple junctures in the design and use of AWS. Following the adoption of Guiding Principle (c) in 2019, a number of states called for a greater investigation into what type and degree of human–machine interaction is demanded for IHL compliance, specifically because these demands can also stem from ethical or operational considerations.

This report finds that IHL does not provide a general answer to that query. In fact, it calls on states to elaborate their presumptions about the very nature of IHL, and consider the question of whether IHL is: (*a*) solely an effects-based regime, permitting militaries to use any combination of humans and machines to undertake military action as long as the anticipated or actual effects are not unlawful; or (*b*) a regime that also mandates evaluations and judgements by human beings in the conduct of military operations. In that regard, the report has identified a number of threshold questions that can be used to guide states when elaborating on their interpretation of what IHL provisions require, permit or prohibit from humans and technology, respectively (see appendix A). These pertain to respect for IHL provisions across four dimensions: personal, material, temporal and geographical.

[216] Boulanin et al. (note 92).

The first dimension—personal—relates to *who* may or must respect IHL provisions. Are humans the only valid agents who may do so, or can artificial agents such as AWS be permitted to exercise some or all IHL obligations? Does responsibility for the decision to employ an AWS, and the resulting effects, reside with one single person or multiple people? If the latter, how do individual contributions to the ultimate decision to employ an AWS, and to administer its operation, interact in the systemic multi-agent model? These questions are critical for the framing of the debate on human–machine interaction (who needs to exercise what form of control and judgement?), as well as the debate on state and individual responsibility (who can be held responsible for an IHL violation?).

The second dimension—material—relates to *what* type and degree of human–machine interaction the substantive and procedural rules of IHL require, permit or prohibit. Does IHL mandate, allow or bar legally required value judgements from being entrusted partly or fully to machine processes? This interpretative question has practical implications for legal reviews (e.g. what should be reviewed, on what basis and at what standard?) and the provision of legal advice (e.g. can legal advice be automated, who should get it, when and where?). It is also highly relevant for the debate on state responsibility and individual criminal responsibility (e.g. when can a state and its human agents be held responsible for overly relying on automated input?).

The third and fourth dimensions—temporal and geographical—relate to *when* and *where* (in relation to what locations) IHL provisions need to be respected. How far in advance may IHL-mandated evaluations be made? What limits does IHL impose on how much time may lapse between the activation of an AWS and when its operation must be suspended or terminated? How do the characteristics of the environment of use impact the legal parameters of AWS use? These questions are critical for the debate on the extent to which humans need to maintain situational awareness and to act as legal agents in exercising and implementing judgement once a system has been activated and launched.

A final and cross-cutting question is *how* the lack of foreseeability introduced by the use of autonomy should be addressed and managed. What level of predictability is demanded by IHL? How should unpredictability be evaluated in advance (as part of the legal review and legal advice), controlled during use, and assessed afterwards (to trace individual and state responsibility for IHL violations)? The combination of these different sets of questions is essential to the determination of what would make an AWS unlawful per se, but also to the evaluation of how IHL provisions should be respected in the development and use of AWS: who must do what, when, where and how.

II. Recommendations

Following the key findings of this report, there are three recommendations addressed to states and non-governmental experts that could contribute to the intergovernmental debate on AWS at the GGE and in other relevant forums. These recommendations aim to support more focused and constructive discussions on the legal challenges posed by AWS, including deliberations on the development of aspects of the normative and operational framework applicable to AWS.

1. Deepen and sharpen discussions on what respecting and ensuring respect for IHL means for AWS

With the adoption of the GGE's 11 guiding principles, states have provided answers to some of the legal questions posed by the development and use of AWS. However, the findings outlined above show that these principles remain insufficient. Many questions remain about what IHL requires, permits or prohibits with regard to the development and use of AWS. Some of them pertain to old legal debates, others are new and specific to AWS. The list of questions that structure this report provides a framework for states to elaborate their views on a range of central matters, particularly the issue of what is and should be expected from humans, AWS and their interactions in order to secure respect for IHL (see appendix A). To deepen the discussions, states should consider how to best utilize the agenda of the GGE, as well as other venues, formats and modalities. In that respect, scenario exercises could help states explore what combinations of human–machine interaction are off limits. These could cover the four dimensions discussed in the report—personal (who), material (what), temporal (when) and geographical (where).

2. Share views and experiences about practical measures that could enhance respect for IHL in the development and use of AWS

A practical focus of future discussions on the normative and operational framework applicable to AWS should be on identifying measures and best practice to help ensure that AWS are used in compliance with IHL. In this regard, states should share their views and experiences of what standards of knowledge and behaviour are expected in the development and use of an AWS to: (*a*) allow the user to foresee whether the operation, performance and effects of the system would be lawful; (*b*) ensure that the user can administer the weapon's operation and limit its effects as required by IHL; and (*c*) ensure that the consequences of employing an AWS can be satisfactorily traced back to an individual and/or a state (e.g. what documentation or information records would be necessary to attribute, discern or scrutinize unlawful conduct).

3. Further elaborate the legal and ethical bases for human–machine interaction in relation to IHL compliance

Compliance with IHL, along with ethical and security considerations, is a critical benchmark for assessing the acceptability of AWS and the need for human–machine interaction. However, with regard to certain key aspects, what respect for IHL entails can be subject to different interpretations. Notably, states' understanding of what type and degree of human–machine interaction is demanded for IHL compliance could vary as a result of different ethical inclinations. In that context, states need to clarify their ethical presumptions about why particular forms of human–machine interaction may be warranted in relation to IHL compliance: whether the choice of human–machine interaction only needs to be guided by the need to limit the risk of producing unlawful effects (from a utilitarian/effects-based perspective), or whether it also needs to ensure human agency and responsibility in the exercise of IHL obligations (from a deontological/process-based perspective). Answering that question is essential for clarifying whether some type and degree of human–machine interaction would always be needed regardless of the characteristics of the weapon system and environment of use. The question is also relevant for discussion of whether the CCW process should look beyond the sole case of AWS and aim more broadly to develop norms to preserve human agency in exercising IHL obligations.

Appendix A. List of key legal questions

This appendix is intended as a practical tool for states and legal experts to elaborate their views on the type and degree of human–machine interaction that is demanded for compliance with international humanitarian law (IHL) in the use of autonomous weapon systems (AWS). It contains the key questions that the report identified with regard to the interpretation and application of: (*a*) the rules on weapons, means and methods of warfare; (*b*) the obligation to conduct legal reviews and provide legal advice; and (*c*) the frameworks of state responsibility and individual criminal responsibility. The questions are sorted on the basis of the four dimensions considered in the report: personal, material, temporal and geographical. Each individual dimension merits attention, but it may be in their combination that the most central questions can actually be found. The combination of these different dimensions is essential to the determination of what would make an AWS unlawful per se, but also to the evaluation of how humans should be exercising their legal obligations in the development and use of an AWS: who must do what, when, where and how.

Questions concerning the rules on weapons, means and methods of warfare

Personal:

- Who may or must be responsible for respecting IHL provisions governing AWS? Are human agents the only valid agents, or can artificial agents also be permitted or required to perform IHL obligations? Must the responsibility for the decision to employ an AWS reside with one single person or can it reside with multiple people?
- What can be demanded from IHL by the people who are the potential *objects* of military action, in terms of the type and degree of human–machine interaction?
- What is the relevance of the Martens Clause when discussing human agency in the use of force, and what limits (if any) does it place on the use of AWS in the conduct of hostilities?

Material:

- What IHL-mandated evaluations (if any) can be entrusted partly or fully to machine processes?
- What kind of socio-technical indicators can be relied on to make IHL-mandated evaluations?

Temporal:

- When do the various obligations underlying respect for the principles of distinction, proportionality and precautions begin and end?
- How far in advance of an AWS activation does the law permit IHL-mandated evaluations to be implemented? Under what circumstances would the evaluations demanded by IHL provisions need to be performed after activation?

Geographical:

- In relation to what locations must IHL provisions be respected?

- What spatial limits (if any) does IHL impose on where AWS may travel and be used?
- How should the user interact with the environment of use to respect IHL provisions before and during an attack?

Questions concerning legal reviews and legal advice

Personal:

Legal review
- Who should be involved in the legal review of AWS? What sets of expertise are required?

Legal advice
- Who should receive legal advice? What constitutes the 'appropriate' level of command?
- Who should provide the legal advice?

Material:

Legal review
- What should be reviewed, considering the techno-environmental interdependencies that underpin autonomy?
- On what legal basis should AWS be reviewed? What are the applicable fields of international law to consider in legal reviews in general, and for AWS in particular? Should the rules on the conduct of hostilities be considered in legal reviews of AWS, and to what extent?

Legal advice
- What makes the provision of legal advice necessary in relation to AWS?

Temporal:

Legal review
- When and how often should a review be conducted? What type or degree of software modification would trigger a new review?

Legal advice
- At what juncture in the development, acquisition, and deployment and use of AWS would the provision of legal advice be necessary?
- How should the legal review process and the provision of legal advice complement one another in relation to AWS?

Geographical:

Legal review
- What is the legal reviewer expected to know and foresee about the intended environment of use?
- What empirical evidence should be gathered, from where, by whom and how?

Legal advice
- What type and level of information does a legal adviser need to possess about the systems and the environment of use to make a legal assessment?

Questions concerning the frameworks for state responsibility and individual criminal responsibility

Personal:

State responsibility

- Which (and how many) state agents are expected to exercise and implement a state's obligations under IHL?

Individual criminal responsibility

- Does a use of force decision ultimately reside with one single person, such as a commander, or can it be distributed across multiple individuals in a command-and-control chain and what does it entail in relation to the imposition of individual criminal responsibility?

Material:

State responsibility

- What particular measures are required to 'respect and to ensure respect for IHL' in the development and use of AWS?
- What types of acts or omissions would engage state responsibility in the context of AWS?
- What types of 'failures' in AWS would qualify as unintended or foreseeable accidents, respectively?

Individual criminal responsibility

- What conduct amounts to a war crime when using AWS?
- How should *mens rea* be interpreted in the context of AWS? Notably, to what extent should risk-taking, negligence and recklessness be considered in the interpretation?
- To what extent do the modes of responsibility apply to the range of agents involved in the development and use of AWS?

Temporal:

State responsibility

- At what points in time is state responsibility engaged? For example, when does the obligation to take 'constant care' and 'feasible precautions' begin?

Individual criminal responsibility

- How far back in time may acts or omissions related to AWS give rise to individual criminal responsibility?

Geographical:

State responsibility

- As AWS are likely to take place across different domains, in relation to what types of environments (e.g. land, naval or air) must a state perform its IHL obligations?

Individual criminal responsibility

- What are the geographical limits (if any) to imposing individual criminal responsibility related to AWS?

About the authors

Dr Vincent Boulanin (France/Sweden) is a Senior Researcher leading SIPRI's research on emerging military and security technologies. His focus is on issues related to the development, use and control of autonomy in weapon systems and the military applications of artificial intelligence (AI). He regularly presents his work to and engages with governments, United Nations bodies, international organizations, research institutes and the media. Before joining SIPRI in 2014, Boulanin completed a doctorate in political science at the École des Hautes Études en Sciences Sociales [School of Advanced Studies in the Social Sciences], Paris. His recent publications include *Responsible Artificial Intelligence Research and Innovation for International Peace and Security*, SIPRI Report (2020, co-author); *Artificial Intelligence, Strategic Stability and Nuclear Risk*, SIPRI Report (2020, co-author); and *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*, SIPRI/ICRC Report (2020, co-author).

Laura Brunn (Denmark) is a Research Assistant working on emerging military and security technologies. Her focus is on how emerging military technologies, notably autonomous weapon systems, affect compliance with—and interpretation of—international humanitarian law. She has a background in Middle Eastern Studies and International Security and Law, and wrote her master's thesis on remote warfare's implications for the protection of civilians, by analysing the United States targeting cycle in its aerial campaign against the Islamic State. Before joining SIPRI, Bruun worked at Airwars in London, where she monitored and assessed civilian casualty reports from US and Russian airstrikes in Syria and Iraq. Her recent publications include *Responsible Military Use of Artificial Intelligence: Can the European Union Lead the Way in Developing Best Practice?*, SIPRI Report (Nov. 2020, co-author).

Netta Goussac (Australia) is an Associate Senior Researcher within Armament and Disarmament at SIPRI, with particular expertise in legal frameworks related to the development, acquisition and transfer of weapons. Before joining SIPRI in 2020, she 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. She has provided legal and policy advice related to new technologies of warfare, including autonomous weapons, military applications of artificial intelligence (AI), and cyber and space security. Since 2017, Goussac has participated in the United Nations Group of Governmental Experts on lethal autonomous weapon systems. Her recent publications include *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*, SIPRI/ICRC Report (2020, co-author); and 'Safety net and tangle web: Legal reviews of AI in weapons and warfighting', Humanitarian Law and Policy Blog (ICRC, 18 Apr. 2019).

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