**2020 12th International Conference on Cyber Conflict**
20/20 Vision: The Next Decade
T. Jančárková, L. Lindström,
M. Signoretti, I. Tolga, G. Visky (Eds.)
2020 © NATO CCDCOE Publications, Tallinn



# Making the Cyber Mercenary – Autonomous Weapons Systems and Common Article 1 of the Geneva Conventions

**Aleksi Kajander**
MA Candidate
Tallinn Law School
Tallinn University of Technology
Tallinn, Estonia
aleksi.kajander@gmail.com

**Agnes Kasper**
PhD, Senior Lecturer
Tallinn Law School
Tallinn University of Technology
Tallinn, Estonia
agnes.kasper@taltech.ee

**Evhen Tsybulenko**
PhD, Senior Lecturer
Tallinn Law School
Tallinn University of Technology
Tallinn, Estonia
evhen.tsybulenko@taltech.ee

**Abstract:** Common Article 1 of the Geneva Conventions requires that states 'respect and ensure respect for' the Geneva Conventions 'in all circumstances'. In the new 2016 Commentary to the Convention, the existence of not only a negative obligation, but also a positive obligation of third countries to a conflict to prevent violations was confirmed. Hence, third countries must do everything 'reasonably in their power to prevent and bring such violations to an end'.

The use of autonomous weapons systems (AWS) is imminent in the future, as demonstrated by the Pentagon committing to spend $2 billion on research, with similar research programmes taking place in other countries. The buying and selling of these AWS is an equally impending part of the future. Consequently, inevitably a state that

is buying or being supplied with AWS will use them in a conflict. Therefore, suppliers of such systems will have to comply with the aforementioned positive obligation.

This paper will examine the positive obligation's impact on the state supplying AWS to a conflict. This includes the question of whether it will be their responsibility at the manufacturing stage to ensure that the system cannot violate the Geneva Conventions and – because autonomous systems are somewhat uncontrollable and unpredictable as they will also learn rather than only carrying out pre-programmed commands – whether the supplying state will be obligated to maintain a permanent tether to the supplied AWS to monitor them. The implications of tethering the supplied AWS may go well beyond ensuring compliance with international humanitarian law (IHL), and may include multiplying the leverage of the supplying state by turning the systems into 'cyber mercenaries'.

**Keywords:** *autonomous weapons, Geneva Convention, international humanitarian law, IHL*

## 1. INTRODUCTION

The development of autonomous technology is raising questions and shifting paradigms in a variety of fields such as transport, business and even governance. The military is no exception to this trend, as the possibilities for the military uses of autonomous technology are becoming increasingly apparent. However, as in other fields, the existing framework of laws was not created with autonomous systems in mind, and therefore its application to such systems is unclear. In the case of the military application of autonomous weapons systems (AWS), the application of the existing rules is literally a matter of life and death.

The Geneva Conventions have long been a cornerstone of international humanitarian law (IHL), and their application and interpretation have had fundamental effects on conflicts since their introduction.[1] They are now having to be examined in a new light, which creates new legal questions about their application.

An updated Commentary was released on the First Geneva Convention in 2016, which confirmed the existence of a positive external obligation under Common Article 1,

1 Lindsey Cameron, Bruno Demeyere, Jean-Marie Henckaerts, Eve La Haye, Heike Niebergall-Lackner, 'The updated Commentary on the First Geneva Convention – a new tool for generating respect for international humanitarian law' (2015), ICRC 97,1210.

whereby the High Contracting Parties 'undertake to respect and ensure respect for' the Convention in 'all circumstances'.[2] This positive obligation requires that the High Contracting Parties do 'everything reasonably in their power to prevent and bring such violations to an end'.[3]

This positive external obligation reaches a whole new dimension with the introduction of AWS, as a contracting party supplying them could potentially have unprecedented control over their supplied systems, whether by their programming or by the presence of a 'backdoor' enabling remote control. Either would significantly improve their ability to prevent IHL violations. However, the latter type of tethering, if required by Common Article 1, could also bring a new dimension to cyber warfare and have unintended military and political effects. Therefore, backdoors are a double-edged sword in the sense that, while they may bring added compliance, they will bring additional risk factors in the form of unintended third parties gaining access to the AWS.

Therefore, defining the parameters of this positive external obligation will be of utmost importance for states supplying such AWS, as it will impact both the design of those systems and the circumstances in which they can be supplied. This paper aims to analyse the relationship and implications of the positive external obligation in Common Article 1 concerning AWS and the states supplying them, particularly whether the supplying state is obliged to maintain a tether to the supplied systems.

## 2. COMMON ARTICLE 1

At its core, Common Article 1 (CA1) has a two-fold structure, the first part of which is to restate the principle of *pacta sunt servanda*: the binding nature of the treaty and the obligation of the parties to perform the treaty obligations in good faith.[4] This first obligation is evidenced by the wording of the Article, under which all High Contracting Parties (HCPs) 'undertake to respect' the convention in all circumstances. The first obligation is therefore relatively straightforward: to ensure that each party performs their obligations in good faith and respects the Conventions and the entire body of international humanitarian law binding upon that state. The reference to 'all circumstances' clarifies that the obligations of CA1 are always applicable both in peace and in more exceptional circumstances, a view confirmed by the 2016 Commentary.[5]

2 Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (First Geneva Convention), 12 August 1949, 75 UNTS 31, Article 1.

3 International Review of the Red Cross, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 2nd edition* [154].

4 Ibid. 143.

5 Ibid. 185.

However, the second obligation is far more ambiguous, as arguably there are many ways of 'ensuring respect', and moreover, the scope of this obligation may include an external dimension regarding the compliance of other states. Hence, the second obligation, to 'ensure respect' for the Convention in all circumstances, would go beyond the ordinary principle of *pacta sunt servanda* in the sense that the parties are not only obliged to perform their obligation in good faith, but also to ensure that others do so as well.[6] This second obligation derives from the addition of the words 'and to ensure respect' for the Convention, which, read in combination with the first obligation, could conceivably be directed outwards.

There is debate regarding the scope of the obligation to 'ensure respect', whether it is narrow and not directed towards other parties or broad and external as the updated 2016 ICRC Commentary states.[7] In essence, at the time of its adoption the obligation to 'ensure respect' was not considered to be external in nature, as evidenced by the *travaux préparatoires*.[8] However, those in favour of a broad scope argue that, since its adoption, the meaning of the provision has evolved through subsequent practice to include an external dimension.[9] The counterarguments point to the existing contrary state practice and the high standard of Article 31(3)(b) of the Vienna Convention on the Law of Treaties, which in their view requires that all parties accept or acquiesce to the subsequent practice for it to be relevant.[10]

Under the narrow view, the obligation of states to ensure respect contained in CA1 pertains only to their organs and those acting under their effective control.[11] This has severe implications regarding AWS, as without the external dimension of the broad scope it would be sufficient for HCPs to ensure that their AWS respect the Convention. This obligation would nevertheless extend to supplied AWS in the sense that they should not of their own accord encourage IHL violations under CA1.[12] However, should their supplied AWS be misused, CA1 would not provide an obligation to ensure compliance by the systems, as the supplying state does not have effective control over them. Consequently, under the narrow scope the supplying states would only have to ensure that their own AWS and any AWS they have effective control over respect the Convention, and that those supplied do not encourage violations.

6 Ibid. 154.
7 Theo Boutruche, Marco Sassoli, 'Expert Opinion on Third States' Obligation vis-à-vis IHL Violations under International Law, with a special focus on Common Article 1 to the 1949 Geneva Conventions' < https://www.nrc.no/resources/legal-opinions/third-states-obligations-vis-a-vis-ihl-violations-under-international-law/> accessed 21 February 2020.
8 Andrea Breslin, 'Reflections on the Legal Obligation to Ensure Respect' (2017), Journal of Conflict and Security law 22(1), 11.
9 Boutruche, Sassoli (nr 7) 7-8.
10 Tomasz Zych, 'The Scope of the Obligation to Respect and to Ensure Respect for International Humanitarian law', (2009) Windsor Yearbook of Access to Justice 27, 256.
11 Ibid. 270.
12 Ibid. 265.

It ought to be highlighted that should a tether enabling effective control of a supplied AWS exist, then arguably it will be within the scope of the obligation to 'ensure respect' for CA1 for the supplying state, even under the narrow view. However, the narrow view cannot require a supplying state to tether supplied AWS in the first place, as there is no obligation towards ensuring respect in regard to other states. Therefore, the design decision of whether supplied AWS are tethered will determine whether the CA1 obligation will apply after they are exported. Thus, regardless of which interpretation prevails, CA1's obligation to ensure respect will conceivably affect the design of AWS, for if a tether is included, then the supplying state must comply with that obligation even after the system has been supplied.

For the purposes of this paper, the obligation of 'ensuring respect' shall be construed to include an external dimension under the 'accepted' contemporary interpretation[13] and the ICRC 2016 Commentary and the Expert Opinion requested in light of it.[14] This is to enable the analysis of the relationship between CA1 and AWS in its potentially most influential form, that is to say, whether it can require a tether to be included by the supplying state in all AWS it supplies.

As pointed out by the 2016 Commentary, the meaning of the term 'ensure' is to make sure something will occur or in this case will not occur, i.e. violations of the Conventions.[15] Logically, this goes beyond a prohibition on encouraging, aiding or assisting violations of the Convention by parties to a conflict. Therefore, ensuring respect within the meaning of CA1 includes a preventive aspect, whereby the HCPs must take steps to prevent foreseeable violations, both during peace and wartime, which, as mentioned above, is also directed towards other parties such as those in a conflict. The positive obligation also requires that the HCP does 'everything reasonably in their power to […] bring such violations to an end'.[16]

In relation to preventing future violations, there must be a foreseeable risk of them being committed.[17] The actual means by which a state is to carry out this obligation is largely at its discretion, provided the principle of due diligence is adhered to.[18] Hence, the positive external duty to ensure respect is an 'obligation of means', whereby an HCP is not held responsible for a failure of its efforts, provided it did everything reasonably in its power.[19] Consequently, the HCP must first correctly identify foreseeable future violations and then take all measures reasonably in its power to prevent them.

13 Breslin (n 8) 37.
14 Boutruche, Sassoli (n 7) 13.
15 International Review of the Red Cross, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 2nd edition* [145].
16 Ibid. 154.
17 Ibid. 164.
18 Ibid. 165.
19 Ibid.

The 2016 Commentary goes on to refer to the 'unique position' of influence where an HCP takes part in the arming, training or otherwise equipping of the armed forces of a party to a conflict.[20] If we consider autonomous weapons systems in this context, it is apparent that if an HCP is providing such weapons, it is arguably in a unique position to prevent or end violations as it could reasonably have taken a multitude of steps to increase its influence beforehand, such as placing remote kill-switches on the supplied systems. Arguably, this is the first time the use of physical weapons systems in the physical possession of a state to which it has been supplied can be made conditional on complying with IHL, even if conceivably similar conditions could already in the present be attached to the use of cyber capabilities supplied by another state. Therefore, whereas in the case of conventional human-operated weapons the most the supplying party could do directly is to stop further supply, under the new paradigm the threat could be to make existing systems useless, thus greatly increasing the leverage. This would effectively prevent future violations, at least by those AWS that can be disabled. Which both introduces the importance and leads us to the main topic of this paper: what are the implications of CA1 in relation to an HCP supplying AWS, and is the supplier required to maintain a tether enabling control of those supplied systems?

## 3. AUTONOMOUS WEAPONS SYSTEMS

Autonomous weapons systems are no longer contained within the realm of science fiction, as already in the present day there are, for example, missile defence systems that can work entirely autonomously. These include the U.S. Aegis command system and the Phalanx Close-in Weapons System (CIWS), which has a mode where it presumes the human operators are incapacitated and it can engage incoming missiles and aircraft on its own.[21] From this example, we may derive the key aspects for defining an autonomous weapons system: a weapons system that is capable of independently identifying and making the decision to engage targets without human intervention, which closely mirrors the U.S. definition of an AWS.[22] There is much discussion regarding the precise definition; however, for the purposes of this paper, we will use the definition whereby a weapons system is autonomous when it can identify, target and engage without human intervention.

The lack of human influence has led to discussions about the 'responsibility gap'[23] regarding AWS, similar to the discussion about liability for self-driving cars and other vehicles. In both cases, the options that are most often discussed are that either the manufacturer or programmers are liable, or the seller, the operator in limited cases,

20 Ibid. 167.
21 Rebecca Crootof, 'Autonomous Weapon Systems and the Limits of Analogy' (2018) HNSJ 9, 59.
22 Ingvild Bode, Hendrik Huess, 'Autonomous Weapons Systems and changing norms in international relations' (2018) Review of International Studies 44, 399.
23 Marcus Schulzke, 'Autonomous Weapons and Distributed Responsibility' (2013) Philosophy & Technology 26, 206.

or the user (such as in the case of neglect that leads to a failure), or even the machine itself.[24] While each has its pros, cons and limitations, the discussion is too complex to attempt to resolve in this paper.

Nevertheless, a few aspects must be discussed in this regard. Firstly, the question of the possibility of human intervention is crucial for the accountability for the actions of the autonomous system. Arguably, if a person has the possibility of influencing the autonomous system, it is not truly autonomous, as that person will be held responsible for failing to prevent the system from malfunctioning. In the case of autonomous vehicles, there are complex legal and ethical questions of whether such a possibility should even be included, as its inclusion would defeat the point of the autonomous vehicle; the human would still have to supervise it, thereby removing the benefit of, for example, sleeping while travelling.[25]

The same will hold true for AWS, but with the added dimension that now the autonomous system can make decisions to specifically end human life. Therefore, in the case of AWS, the pressure to include such safeguards is increased, but this raises further ethical questions; if the AWS is capable of operating unsupervised in a dangerous situation, is it ethical to endanger your own soldiers' lives by placing them inside the system to monitor its operation?

Secondly, it may be an unfortunate reality that not all AWS can be monitored if they are on the offensive, as it may be beneficial from a military point of view that they abstain from unnecessary communications and are as 'radio-silent' as possible, to prevent their location and destruction by the enemy. Hence, it is conceivable that future AWS may not have any human overrides, which would create the 'accountability gap'.[26] This would mean that the supplying state, if it so desires, could distance itself from supplied AWS in a similar way to 'traditional' weapons operated by humans, by stating that the users have the possibility of influencing them.

A further aspect in relation to AWS, which is closely related, is the unprecedented opportunity to include a pre-programmed 'basic moral code', whereby the AWS would simply refuse to comply with certain commands, such as those in clear violation of the Geneva Conventions. This situation is distinct from present reality, where human combatants may harbour hidden 'characteristics' unknown to their commanders, such as hatred of certain ethnicities, a thirst for revenge in the heat of battle, or hidden mental diseases. The possible presence of these hidden characteristics in human combatants is preventable in AWS, where, despite a potential capacity to learn and

24 Alexander Hevelke, Julian Nida-Rumelin, 'Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis' (2015) Science & Engineering Ethics 21, 620-621 & 623-624.

25 Ibid, 619-630.

26 Marcus Schulzke, 'Autonomous Weapons and Distributed Responsibility' (2013) Philosophy & Technology 26, 206.

adapt, the programming of the system could nonetheless include safeguards like Asimov's Laws of Robotics,[27] i.e. absolute prohibitions that underlie all operations.

Due to this possibility, the state supplying and producing AWS has a concrete and unique opportunity to prevent those systems from violating IHL norms, and thus 'ensure respect' for the Geneva Conventions. Potentially, the AWS could even be used as a 'vigilance system', whereby the AWS observing violations of IHL would either store details of those violations in a black box type of storage or send them to the manufacturer or another relevant entity, such as the Protecting Power or even the ICRC. Similarly, the AWS could store all the orders it has received from its human operators in a log, allowing for retroactive tracing of who gave the command and exactly what the command was, thus identifying commands that would have used the AWS to commit violations of IHL. If such features were to be included, non-physical safeguards should be considered, as suggested in the Guiding Principles of a 2019 draft report by the Group of Governmental Experts for the UN Convention on Certain Conventional Weapons (CCW), to prevent, for example, data spoofing that would reduce the utility of such a log and increase uncertainty related to its integrity.[28] All these possibilities hinge on the producer of the AWS including or being required to include such features into their machines, thereby giving further value to defining the obligations of CA1, as these possibilities, if they are technically feasible at the time, could certainly be included in measures reasonably in the power of the HCP supplying the AWS.

Nonetheless, at the time of writing, though many discussions have taken place about legally regulating AWS, especially in the context of the CCW in the form of a pre-emptive ban such as in the case of blinding laser weapons, at present there are no international legally binding instruments on AWS.[29] Therefore, considering that autonomous weapons such as CIWS are already in use, and many research programmes are underway, it is safe to say the legal practice is lagging.[30]

27 Roger Clarke, 'Asimov's Laws of Robotics: implications for information technology' (1993) Computing Milieux, 55.

28 United Nations, 'Draft Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of lethal Autonomous Weapons Systems' <https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf> accessed 17 April 2020.

29 Ingvild Bode, Hendrik Huess, 'Autonomous Weapons Systems and changing norms in international relations' (2018) Review of International Studies 44, 398-400.

30 Ibid. 400.

## 4. INTERACTION OF AWS AND COMMON ARTICLE 1

### *A. Not all AWS are Created Equal*

Autonomous weapons systems are not mentioned in the 2016 Commentary, nor how the obligations of the Article would interact with them. Nonetheless, based on the discussion in the previous section about the nature of AWS, as they can make decisions to engage targets on their own, it is foreseeable that they could do so in violation of IHL norms. Therefore, the positive obligation of preventing violations when there is a foreseeable risk[31] would apply to such AWS systems.

This presumes that the AWS systems in question can cause harm or use lethal force, meaning that a distinction must be made between AWS systems where it is foreseeable that they may cause violations and those that foreseeably could not. It is reasonable to presume that the armed forces will adopt (unarmed) autonomous vehicles such as cars and trucks, but arguably, as these are not designed to have a combat role, they are unlikely to cause violations of IHL in their normal operations. By contrast, the moment an autonomous vehicle is armed, the situation becomes different, as foreseeably the armament could be misused.

The distinction may be even more difficult if we consider the present example of the already autonomous Goalkeeper CIWS system, which can engage missiles and aircraft on its own. First, we must consider that it is a mounted system that is immobile, and so its operation can be closely monitored by humans, even if the people doing the monitoring do not contribute to the decision-making of the system, and the system can be shut down if it malfunctions. Secondly, the system is designed to engage high-speed targets such as missiles and aircraft with the capacity to identify friend or foe (IFF functionality), meaning that it can distinguish between civilian and military aircraft.[32] Thirdly, the system is short-ranged (2000 metres),[33] which in combination with only targeting high-speed objects such as missiles, and its ability to distinguish civilian aircraft, would mean that the foreseeable violations would be limited to engaging a misidentified civilian aircraft that strayed within 2000 metres of the system. Considering the specification of this system, despite it being a lethal AWS as it is capable of destroying aircraft, it is difficult to identify many foreseeable risks in terms of IHL violations, as it is highly unlikely to interact with protected persons under the Geneva Convention and could violate IHL in highly specific scenarios only. By comparison, a mobile airborne autonomous drone engaging in a persistent

31 International Review of the Red Cross, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 2nd edition*, (2016) [164].

32 Seaforces, 'Goalkeeper close-in weapon systems' <http://www.seaforces.org/wpnsys/SURFACE/Goalkeeper-CIWS.htm> accessed 23 December 2019.

33 Ibid.

campaign of targeted killings[34] would be at a higher risk of foreseeably causing IHL violations, as it could target a variety of ground forces, installations and civilian targets. Consequently, the range of foreseeable violations of IHL that the system is capable of causing is far wider than in the case of an autonomous CIWS system.

Both systems in the above examples can be exposed to cyber threats as they rely on and are operated by computer systems. Hence, it is plausible to consider a scenario where a cyberattack causes the AWS to violate IHL.[35] Although there is currently no obligation on states to foresee and analyse possible misuses of weapons,[36] it may be argued that, given the relative but inherent insecurity of computer systems, it can be reasonably expected that tampering by cyber means will sooner or later take place and affect the normal and expected use of an otherwise legal AWS. While no such binding obligation exists, the topic of cyber security in AWS in the context of non-physical safeguards has been mentioned in the Guiding Principles of a 2019 draft report by the GGE for the CCW Convention as an aspect to consider, thereby suggesting at the very least mounting discussions on the topic that could eventually lead to binding obligations in the future.[37] Potential misuses of AWS by adversaries via exploiting unknown vulnerabilities and resulting in the risk of violations of IHL are hardly foreseeable in advance. However, the same cannot be said about already-known vulnerabilities. Therefore, although analysis of misuse may not be required under IHL or other international law obligations, it is questionable whether the existence of a known vulnerability in an AWS that could potentially lead to violation of IHL would render the risk of that violation foreseeable.

Consequently, the foreseeable risk of violations is highly specific to the type of AWS, and as such, AWS cannot be categorised merely based on their autonomous function or potential lethality, but rather a system-by-system overall risk analysis must be performed. For states party to Additional Protocol I, Article 36 does require that reviews are conducted for each new weapon developed or acquired; however, major military powers such as the United States are not bound by AP I, thereby limiting its reach.[38] Moreover, Article 36 weapons reviews that are conducted are not required to be published and therefore can be subject to secrecy, so arguably this lack of transparency could compromise the effectiveness and truthfulness of the reviews that

34 Michael Carl Haas, Sophie-Charlotte Fischer, 'Evolution of targeted killing practices: autonomous weapons, future conflict and international order' (2017) Contemporary Security Policy 38, 283.

35 Michael N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features, 7.

36 ICRC Commentary on the Additional Protocols, paragraph 1469. Also see Michael N. Schmitt (ed) Tallinn Manual 2.0. on the International Law Applicable to Cyber Operations, 466.

37 United Nations, 'Draft Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of lethal Autonomous Weapons Systems' <https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf> accessed 17 April 2020.

38 Natalia Jevglevskaja, 'Weapons Review Obligation under Customary International Law.' (2018) U.S. Naval War College International Law Studies, Vol 94, 209.

are conducted.[39] Nevertheless, a discussion regarding Article 36 of AP I is beyond the scope of this paper and a comprehensive examination thereof would require a separate article to examine.

A state supplying a system like the Goalkeeper CIWS would arguably have to take fewer preventive steps to inhibit the system from causing IHL violations than a state supplying an autonomous 'killer-drone'. The actual content of the obligations under CA1 would be different based on the types of AWS supplied, and could not be mapped precisely in the abstract. However, it is possible to state abstractly that the HCP should take all measures in ensuring that the AWS cannot cause the foreseeable violations of IHL specific to that system. Such measures should include a misuse risk assessment by identifying and appropriately addressing at least the known cyber vulnerabilities that can lead to violations of IHL.

### *B. The External Positive Obligation of Common Article 1*

Under CA1, HCPs have the positive obligations of both preventing future violations and stopping ongoing violations by a party to a conflict. Consequently, AWS provide the unprecedented opportunity to definitively pre-programme a set of rules that the physical weapons system must follow, such as to prevent violations of IHL. Of course, considering the complexity of both practical situations in a conflict and the legal framework, the correct course of action can be difficult to determine and there has been doubt expressed about whether AWS can ever operate within the correct manner from an IHL point of view.[40] However, arguably that is dependent on the type of system, as outlined above in 4.A.

It would be a gross oversimplification to reduce the situation to programming the system with a simple set of rules such as 'never target non-military infrastructure' or 'never cause the death of a civilian' to definitively prevent violations. While both are in theory protected, in practice the situation may be more complicated and would not necessarily involve a violation of IHL, depending on the proportionality and the military advantage gained. For example, a bridge can be entirely a civilian structure, however, the military advantage of destroying that bridge may justify its destruction, thus abstractly transforming it from a civilian structure to a military target.[41] Similarly, in the case of a targeted killing campaign, if a high-ranking enemy is found who is in the presence of a civilian and a decision to engage would end both their lives,

39 International Review of the Red Cross, *Commentary on the Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of International Armed Conflicts (Protocol 1)*, (1987), 1470.

40 Max van Kralingen, 'Use of Weapons: Should We Ban the Development of Autonomous Weapons Systems?' (2016), The International Journal of Intelligence, Security and Public Affairs, 18:2, 137.

41 ICRC, 'Practice Relating to Rule 10. Civilian Objects' Loss of Protection from Attack' <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule10> accessed 23 December 2019.

conceivably considerations of military advantage and proportionality could justify the killing of the civilian alongside the high-ranking commander.[42]

Both cases highlight that commands that appear almost like a tautology such as 'never kill or cause the death of a civilian' are not always realistically possible to include as overruling laws, in a manner similar to Asimov's Laws of Robotics. Consequently, the task of pre-programming an AWS to such an extent that under absolutely no circumstances could it violate IHL is a herculean task. The supplier of an AWS could likely never eliminate the chance of their AWS causing violations purely based on its programming. Of course, if such a technological feat is possible, feasibly CA1 would require that the supplied AWS would be included with such programming as it would be a measure reasonably in the power of the supplying state. However, we must be realistic and assume it is not possible, at least for all systems, for the near future.

Therefore, from the above conclusion we arrive at the second possibility that could potentially be required under CA1, the question of whether or not the supplying HCP has an obligation to retain the possibility of influencing the AWS or monitoring its activity.

### *C. To Tether or Not to Tether?*

The possibility of influencing the actions and behaviour of AWS using remote-control raises the possibility of HCPs meeting the positive obligation of CA1 by taking control of their supplied AWS. This question is similar to that which has been posed regarding encryption: whether backdoors should be provided to give authorities access.[43] In the case of AWS, the discussion will have the added life-and-death dimension whereby if a backdoor is included and the system is hacked, lives could be lost. The presence of a backdoor also increases the number of actors potentially able to commit IHL violations with the AWS, should a third party be able to hijack the system by exploiting the backdoor. To a degree, this risk could be reduced by limiting the backdoors to only disabling the AWS, which if breached would at least not cause violations, but would hamper the functionality of the AWS considerably.

There could be no better or more immediate way of preventing violations by AWS used by a party to a conflict than remotely disabling those systems being misused. Therefore, from a compliance perspective, the ability to remotely monitor and influence would ensure the respect for the Geneva Convention and other applicable IHL, even if it is a double-edged sword due to the risk of unauthorised access. Several other considerations should be considered when determining whether tethering the AWS should be required as a means of fulfilling the obligations under CA1.

42 ICRC, 'Practice Relating to Rule 14. Proportionality in Attack' <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule14> accessed 23 December 2019.

43 Ronald Rivest, 'Case against regulating encryption technology' (1998) Scientific American, 116-117.

First, let us consider the untethered model, whereby, to begin with, the supplying state severs or does not include all possibility of influencing the supplied system once it has been supplied to another state. The supplying state would be entirely unable to monitor or direct its activities in the future. This would render AWS akin to traditional human-operated weapons systems whose country of origin has no control over how they are used after handing them over. Consequently, the supplying state would have to resort to the traditional means of influence such as diplomatic pressure, economic sanctions and refusing to supply the party in the future.[44]

Under this untethered model, the introduction of AWS changes less in how the HCPs comply with the obligation to 'ensure respect' under CA1. The only meaningful improvement would be the programming of the AWS aimed at preventing the misuse of the AWS, which would be included under the measures that HCPs can reasonably take to prevent foreseeable violations. This would likely not cover all possible situations where violations can occur, and hence would likely not be a panacea. Nonetheless, when compared to the present where the compliance or non-compliance of non-autonomous weapon systems is entirely at the mercy of their crews,[45] it would be an improvement.

The second possibility is the 'tethered' model, which could be described by analogy as a 'Swiss mercenary of old' model. If the supplying state maintains some form of connection, be it the capacity to monitor the activity, direct the activity or have a remote 'kill-switch' for the AWS, the AWS is not truly an asset of the state it has been supplied to, but rather something of a cyber mercenary'. In this sense it is similar to the 'Swiss mercenaries of old', whose service came with conditions in regard to their state of origin (Switzerland) such as that they might be recalled if the Swiss confederacy were to come under attack.[46] Consequently, a prudent user of the Swiss mercenaries would have understood that they could not be entirely relied on in all circumstances. Similarly, if the AWS were tethered to its state of origin it could not necessarily be relied on in all circumstances, such as when those AWS were used to cause violations of IHL or conflict with the supplying state. Especially if there were a conflict with the state supplying the AWS, the user might find that those systems had 'turned traitor', adding a whole new level to cyber warfare, and as such putting them at a great military disadvantage.

That is to say, the AWS could never be 'fully trusted' in the same way as the 'Swiss mercenaries of old', who, while entirely under the command of the local armed forces,

44 Knut Dormann, Jose Serralvo, 'Common Article 1 to the Geneva Convention and the obligation to prevent international humanitarian law violations' (2014) ICRC 96, 725-726; International Review of the Red Cross, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 2nd edition*, (2016) 181.

45 Hin-Yan Liu, 'Categorization and legality of autonomous and remote weapons systems' (2012) ICRC 886, 629-630.

46 John McCormack, *One Million Mercenaries: Swiss Soldiers in the Armies of the World* (Pen and Sword 1993), 62.

would nonetheless have a link to their state of origin. Similarly, tethered AWS would have a remote cyber link to the supplying state, which may be activated at any time, thus transforming them into a 'cyber mercenary' from the supplying state. Therefore, the armed forces of a state buying AWS might be compromised by the presence of these 'cyber mercenaries' in their ranks, which would enable the supplier of those AWS to retain both political and military leverage over the state using those systems. Naturally, this analogy is restricted by the use of the term 'mercenary', as there is a risk of confusion with the term's present legal meaning under which a 'mercenary' does not retain any link to their state of origin.[47] Consequently, this 'cyber mercenary' would arguably require a new term without pre-existing definitions or prejudices. In this vein, a portmanteau between 'autonomous' and 'mercenary' could be used, such as 'autocenary', which could be defined as 'an autonomous weapons system that is tethered to its state of origin or production by means that enable monitoring or remote control'. Nevertheless, despite its limitations, the term 'cyber mercenary' will be used for the purposes of this paper.

Nevertheless, the tethering of AWS to the supplying state would solve one of the key questions of supplying weapons: what if they are ever used against the supplier? For on the one hand, the supplier wants to supply inferior systems so that they cannot compete with their own, but at the same time they must be better than the competing systems which would otherwise be chosen. Maintaining control would give the best of both worlds to the supplier: the systems can be as effective as possible, as the supplier knows that if ever it was used against them, they could disable or control it. Merely being able to monitor its use would allow the supplier to spy on the supplied state's armed forces, and as such gain valuable intelligence. If we accept that only major military powers will be able to produce and develop their own AWS, tethering them to the supplier would multiply their leverage over states that are forced to purchase foreign systems and are thus left with an unreliable military full of 'cyber mercenaries'. The leverage gained by such a tether would be both military and political, as not only does the supplying state have a measure of control over the military of the supplied state, but also political capital. This control could be used to ensure favourable relations with the supplying state by exploiting that leverage given by the tethered AWS.

However, it must equally be remembered that if the supplier can remotely access the AWS, conceivably so could a third party; thus the presence of tethering will increase the vulnerability of the systems to cyber-attack by third parties. This threat is especially elevated by the fact that if such a tether is required by law, third party actors will know that it must be present, therefore justifying a significant investment into attempting to exploit such a tether and the leverage over the military of the supplied state brought with it. If no tether is required, third party actors would have to consider

47 International Convention against the Recruitment, Use, Financing and Training of Mercenaries, 4 December 1989, Article 1 (1) (e) and 1 (2) (d).

if such a tether even exists, and thereby the incentive to invest significant resources into exploiting a potential tether would be reduced.

While tethering the systems to the supplying state might appear the most tempting option to fulfil the positive external obligation under CA1, if such a tethering were to be required it would have significant undesirable consequences for any state purchasing such systems. Therefore, it would be prudent not to be naïve when the tethered model is being advocated under the guise of added or assured compliance with the obligations of both IHL, and especially, CA1. Nonetheless, it is equally possible that hidden backdoors and overrides can never be conclusively eliminated, regardless of whether or not this would be required by CA1, as the potential leverage is so tempting.

## 5. CONCLUSION

The relationship of the positive external obligation of CA1 and AWS can take on a variety of directions; however, the key factor of the relationship is the question of tethering the supplied AWS so that the supplying state can 'ensure respect', as required by CA1, in all circumstances. Certainly, from a legal point of view, a compelling case can be made for requiring such tethering based on the need for HCPs to do 'everything reasonably in their power to prevent and bring such [IHL] violations to an end'[48] under the positive obligation of CA1. Consequently, provided that such a tether is technically feasible, it would be within the reasonable power of the supplying state to include such a backdoor for access, and would significantly aid in preventing both future and ongoing violations.

The choice, however, is more difficult and complex, as the trade-off is either potentially sacrificing compliance by not requiring the tethering, or potentially compromising the armed forces of the supplied states with these autonomous 'cyber mercenaries'(autocenaries) in their ranks in exchange for added compliance. The presence of tethering would also likely significantly increase the risk of the AWS being hijacked by a third party, thereby further adding to the cyber security concerns of the systems. Requiring the tethering of the AWS would have significant political and military implications by further increasing the power of the states supplying AWS, and the potential military leverage gained by cyber warfare for third parties seeking to exploit the tether would be increased.

Moreover, it must be kept in mind that not all AWS are the same and involve similar foreseeable risks of committing violations of IHL. Therefore, the question of 'to tether

48 International Review of the Red Cross, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 2nd edition*, 154.

or not to tether' could be broken down to a case-by-case basis, wherein, for example, an AWS that has a relatively low risk of causing violations, such as a stationary missile defence system, would not be under a tethering requirement, but a higher-risk 'killer-drone' would be. Under such a system-by-system model, however, the legitimate concern can be raised that, if one state supplies both tethered and untethered AWS, how is the receiving state ever going to silence the doubt that on the 'untethered' systems, the tethers are merely hidden? Consequently, further discussions and contemplations are required on the matter, for conceivably at present the positive obligation of CA1 could be used to justify such a tethered system, as it would ensure a higher degree of compliance and respect for the Geneva Conventions and other applicable IHL.

The positive external obligation of CA1 has implications for the use and development of AWS and the states supplying them. The identified primary key issue arising from the relationship between CA1 and AWS is the question of the tethering of AWS to the state of origin. However, as AWS can take a variety of forms with different risk profiles, it is difficult to provide an all-encompassing answer to whether tethering would be appropriate in every case. This uncertainty is compounded by the additional political and military ramifications of tethering, as it would likely result in an increased power imbalance between the state using the AWS and the supplying state. Therefore, in conclusion, the positive external obligation of CA1 has serious implications for AWS in potentially requiring tethering to the supplying state, a question which is best approached on a system-by-system basis owing to the diversity of AWS and their differing risk profiles.

## REFERENCES


Primary Sources

1. Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (First Geneva Convention), 12 August 1949, 75 UNTS 31.

Secondary Sources

2. International Review of the Red Cross, *Commentary on the First Geneva Convention: Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, 2nd edition*, (2016).
3. International Review of the Red Cross, *Commentary on the Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of International Armed Conflicts (Protocol 1)*, (1987).
4. International Review of the Red Cross, *Commentary on the Additional Protocols*, (1987).
5. ICRC, 'Practice Relating to Rule 10. Civilian Objects' Loss of Protection from Attack' <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule10> accessed 23 December 2019.
6. ICRC, 'Practice Relating to Rule 14. Proportionality in Attack' <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule14> accessed 23 December 2019.
7. Seaforces, 'Goalkeeper close-in weapon systems' <http://www.seaforces.org/wpnsys/SURFACE/Goalkeeper-CIWS.htm> accessed 23 December 2019.
8. Bode, I., Huess, H. 'Autonomous Weapons Systems and changing norms in international relations' (2018) Review of International Studies 44, 393-413.

9. Boutruche, T., Sassoli M, 'Expert Opinion on Third States' Obligation vis-à-vis IHL Violations under International Law, with a special focus on Common Article 1 to the 1949 Geneva Conventions' < https://www.nrc.no/resources/legal-opinions/third-states-obligations-vis-a-vis-ihl-violations-under-international-law/> accessed 21 February 2020.
10. Breslin, A. 'Reflections on the Legal Obligation to Ensure Respect' (2017), Journal of Conflict and Security Law 22(1), 5-37.
11. Cameron, L., Demeyere, B., Henckaerts, J-M., La Haye, E., Niebergall-Lackner, H.'The updated Commentary on the First Geneva Convention – a new tool for generating respect for international humanitarian law' (2015), ICRC 97,1209-1226.
12. Clarke, R. 'Asimov's Laws of Robotics: implications for information technology' (1993) Computing Mileux, 53-61.
13. Crootof, R.'Autonomous Weapon Systems and the Limits of Analogy' (2018) HNSJ 9, 51-83.
14. Dormann, K., Serralvo, J.'Common Article 1 to the Geneva Convention and the obligation to prevent international humanitarian law violations' (2014) ICRC 96, 707-736.
15. Haas, M., Fischer, S-C. 'Evolution of targeted killing practices: autonomous weapons, future conflict and international order' (2017) Contemporary Security Policy 38, 281-306.
16. Hevelke, A., Nida-Rumelin, J. 'Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis' (2015) Science & Engineering Ethics 21, 619-630.
17. Jevglevskaja, N. 'Weapons Review Obligation under Customary International Law.' (2018) U.S. Naval War College International Law Studies, Vol 94, 186-221.
18. Liu, H. 'Categorization and legality of autonomous and remote weapons systems' (2012) ICRC 94, 627-652.
19. United Nations, 'Draft Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of lethal Autonomous Weapons Systems' <https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf> accessed 17 April 2020.
20. van Kralingen, M. 'Use of Weapons: Should We Ban the Development of Autonomous Weapons Systems?' (2016), The International Journal of Intelligence, Security and Public Affairs, 18:2, 132-156.
21. McCormack, J., *One Million Mercenaries: Swiss Soldiers in the Armies of the World* (Pen and Sword 1993).
22. Rivest, R. 'Case against regulating encryption technology' (1998) Scientific American, 116-117.
23. Schmitt, M. 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) Harvard National Security Journal Features, 1-37.
24. Schulzke, M. 'Autonomous Weapons and Distributed Responsibility' (2013) Philosophy & Technology 26, 203-219.
25. Zych, T. 'The Scope of the Obligation to Respect and to Ensure Respect for International Humanitarian Law', (2009) Windsor Yearbook of Access to Justice 27, 251-270.