CHARLES UNIVERSITY

Faculty of Law

Petra Rešlová

Meaningful Human Control in Autonomous Weapons

Doctoral Thesis

Thesis supervisor: prof. Marco Sassòli
University of Geneva, Faculty of Law, Department of International Law
Date of completion (manuscript closure): 14. 06. 2022

UNIVERZITA KARLOVA

Právnická fakulta

Petra Rešlová

Smysluplná lidská kontrola v kontextu autonomních zbraní

Rigorózní práce

Vedoucí práce: prof. Marco Sassòli

Katedra: Ženevská univerzita, Právnická fakulta, Katedra mezinárodního práva

Datum vypracování práce (uzavření rukopisu): 14. 06. 2022

DECLARATION

Hereby, I declare that this doctoral thesis is my original work and that I have written it independently. All sources and literature that I have used during elaboration of the thesis are fully cited and listed. I further declare that this thesis has not been used to obtain any other or the same degree.

The text of this thesis has 289 469 characters including spaces and footnotes.

doctoral candidate

In Prague on 20. 06. 2023

PROHLÁŠENÍ

Prohlašuji, že jsem předkládanou rigorózní práci vypracoval/a samostatně, že všechny použité zdroje byly řádně uvedeny a že práce nebyla využita k získání jiného nebo stejného titulu.

Dále prohlašuji, že vlastní text této práce včetně poznámek pod čarou má 289 469 znaků včetně mezer.

rigorozantka

V Praze dne 20. 06. 2023

ACKNOWLEDGEMENT

I would like to thank my thesis supervisor prof. Marco Sassòli for his valuable guidance and supportive approach, which have many times greatly influenced my work on this thesis. Furthermore, I would like to thank JUDr. Milan Lipovský, PhD not only for his consultations on this thesis but especially for introducing me to Philip C. Jessup Moot and the world of public international law.

PODĚKOVÁNÍ

Na tomto místě bych ráda poděkovala prof. Marcu Sassòlimu za jeho vedení a cenné připomínky k této práci, které mě mnohokrát navedly správným směrem. Dále bych ráda poděkovala také JUDr. Milanu Lipovskému, Ph.D. nejen za konzultace k této práci, ale především za to, že jsem si díky němu a Philip C. Jessup Moot Court zamilovala mezinárodní právo veřejné.

Contents

INTRODUCTION

Law is anthropocentric, its aim is not to change the physical laws of nature but to shape human behaviour in a way that is viewed as desirable by society. However, law is neither static nor permanent. It does not exist in a vacuum. Quite the contrary, new circumstances in the world around us constantly challenge the existing legal norms. All different fields of law must react to emerging technologies as they alter the human reality. The more common situation is that new technologies are developed, humans start using them in a certain way, and only then do legal rules either cement or indeed overturn the existing patterns of human behaviour. The less common approach is pre-emptive, where law is a step ahead of the actual development of technology and legal rules are established that prevent the development altogether or regulate its use.[1] With regards to autonomous weapon systems ("AWS") and their lethal subcategory ("LAWS"), many argue that it is the case for a pre-emptive approach, that "fully" autonomous weapon systems have not yet been developed. Thus, the whole debate is future-oriented. Throughout this paper, an argument will be made that autonomy is a spectrum, and some of the currently deployed weapon systems have crossed the imaginary line into autonomy.

However, that does not change the fact that there is a dire need to regulate lethal autonomous weapon systems, irrespective of the definition we may use to describe them. While the research on autonomy in robotic systems is flourishing in many areas, none is deemed as troubling as the development of lethal autonomous weapon systems. It raises various compelling questions, not only legal but also ethical and moral ones. The debate over those questions sparked in 2012 when the US Department of Defence issued an executive order on autonomous weapon systems[2], and the Human Rights Watch published its report on "killer robots."[3] Ten years later, most of those questions are yet to be answered. The most prominent example is the inability to agree on a definition of a "lethal autonomous weapon system". The problem of the definition will be briefly addressed in this paper, as it is inextricably linked to all the questions the development of autonomous weapons poses to the international legal community.

[1] Asaro, P. Jus nascendi, robotic weapons and the Martens Clause. In: *Robot Law*. Cheltenham, UK: Edward Elgar Publishing, 2016, 367–386. p. 368.

[2] US Department of Defense ("US DoD"). Directive 3000.09 on Autonomy in Weapon Systems. 2012. At: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf (last accessed 13 April 2022).

[3] Human Rights Watch. *Losing Humanity: The Case Against Killer Robots*. 2012. At: http://www.hrw.org/reports/2012/11/19/losing-humanity (last accessed 13 April 2022).

However, that is not to say that no progress has been made. The 2019 Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects ("CCW") adopted 11 guiding principles, addressing the "*potential challenges posed by emerging technologies in the area of lethal autonomous weapons systems to IHL.*"[4] The guidance identifies the legal and moral principles that have been agreed upon by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System throughout the years of their debates. Its importance lies in the fact that they move forward the discussion and show agreement on what are so far rather abstract rules and principles, however, which may find their way into hard law rules such as a treaty or a new protocol to the CCW.

One of the rules proposed is the requirement of "meaningful human control", to which the guiding principle (c) is the most relevant, stating that: "[h]*uman-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole*."[5] This principle emerged from discussions over the desirability of a certain level of human control over lethal autonomous weapon systems and the requirements this control should fulfil. Mentioned already in 2014, it has gained widespread support, and the notion of "meaningful human control" has been accepted by most.

The main goal of this paper is to analyse the emerging principle of meaningful human control and explore its elements and requirements. The topic is highly relevant and contemporary. Even though the majority agrees that meaningful human control over LAWS should be required, the principle itself is yet to be defined. This paper builds upon the current understanding of MHC and aims particularly to clarify questions such as where does the principle stem from and how it should be perceived and integrated into State practice.

[4] CCW. Meeting of the High Contracting Parties to the CCW, Final report, Annex III. CCW/MSP/2019/9. 13 December 2019. At: https://undocs.org/Home/Mobile?FinalSymbol=CCW%2FMSP%2F2019%2F9&Language=E&DeviceType=Desktop&LangRequested=False (last accessed 13 April 2022).

[5] Ibid, p. 10.

However, it should be noted that the intention is not to present a clear-cut solution. The debate over LAWS is extremely dynamic and complex and a great number of issues need to be addressed. Even the very particular issue of meaningful human control is unlikely to be solved in a single paper, which also has a limited scope, focusing mostly on IHL rules. The main contribution here is the analysis of elements and factors influencing the quality of human control over LAWS, something that has not been explored in the relevant literature in great detail. This paper thus aims to bring a new perspective to the debate, attempting to define "meaningful human control" by analysing its elements.

Chapter I will provide the necessary background and introduce topics relevant for further discussion on meaningful human control. The definition and categorisation of lethal autonomous weapons will be addressed to provide an introduction to the topic and demonstrate how a definition can broaden or restrict any attempt at legal regulation. Both comparative method and research synthesis will feature in the Chapter, when different approaches to definitions and their common ground will be examined. Various technological characteristics or features will also be addressed, focusing on categories pertinent to determining what makes a system "autonomous". The interplay between various levels of autonomy and levels of human control will be introduced.

Chapter II will argue that it is desirable, if not necessary, to insist on the requirement of meaningful human control. First, the technological limitations of current and future technology will be addressed, particularly object recognition and classification, bias, or unpredictability. Second, the rules of international humanitarian law on the conduct of hostilities presenting challenges to the use of lethal autonomous weapons will be explored, mainly the rules of distinction, proportionality, and several precautionary measures. A descriptive method of research will be used in this Chapter to characterise the relevant rules of IHL and their implications for the deployment of LAWS. Third and fourth, moral and ethical arguments will briefly be touched upon, reacting to concerns about the lack of human involvement in decisions with significant consequences for people's lives and livelihoods. Finally, the issue of individual criminal responsibility for acts carried out by LAWS will be debated.

Chapter III will explore how control is exercised over weapons currently in use. An argument will be introduced that even systems with automated functions may already be setting a precedent for what is considered meaningful in terms of human control. A case study of an air defence system downing a fighter jet will provide the background upon which the current role of a human operator of complex weapon systems will be analysed, particularly the issues that the current practice displays.

Chapter IV will focus on a possible legal basis for the requirement of (meaningful) human control in armed conflicts. Various IHL rules on targeting will be analysed. An argument will be presented that IHL implicitly requires human control to be maintained over LAWS (or indeed any weapon systems using lethal force). It will also be argued that States seem to consider that the requirement of MHC and its characteristics should be agreed upon in the future and do not yet constitute a rule of customary international law.

Chapter V will build upon the analysis provided in the previous parts and elaborate upon the requirement of meaningful human control in detail. First, the diverging views on its definition and criteria will be introduced. Second, the analysis will focus on what should be control exercised over and at which level. Third, the central part of this Chapter will focus on elements which influence how meaningful the control is. Three elements will be elaborated upon: technological, conditional, and decision-making one. It will be argued that a significant number of factors influence the quality and nature of human control, ranging from the predictability of the programme, tasks and targets, and environment to factors influencing human behaviour, such as automation bias, training, situational understanding, or the time to deliberate. In this Chapter, the main method of research is analytical, as it focuses on understanding the cause-effect relationships between factors influencing the quality of human control.

The final Chapter VI will draw conclusions from the examination of all the elements. It will be argued that meaningful human control ought to be exercised over critical functions of lethal autonomous weapon systems. The appropriate level of control should be determined for each particular set of circumstances in a way that ensures compliance of the weapon system with relevant rules of international humanitarian law, as well as the potential responsibility of its operator for all the resulting actions of the weapon system.

I. DEFINITION AND CATEGORISATION OF LAWS

The focus of this paper lies in exploring the requirement of meaningful human control over lethal autonomous weapon systems. However, in every attempt to address LAWS from a legal point of view, a problem crystallises with their definition. Ten years after the emergence of the debate on this topic, a shared international definition has not yet been agreed upon.[6] A recent analysis has identified 12 definitions of AWS proposed by States or key international actors, which focus on different aspects and lead to different legal approaches.[7]

The following chapter will present selected definitions of (L)AWS and introduce the most used categorisations relevant for the purposes of analysing meaningful human control.[8]

1. Defining autonomous weapons

Defining LAWS is a puzzle that has been following any attempt at legal qualification since the academic legal circles have started debating these emerging technologies. The highly technical nature of the issue poses a particular challenge, as does the fact that the whole debate is very future-oriented (on most points at least). The importance of a precise definition of LAWS must not be underestimated, as the definition itself may broaden or reduce the scope of any future regulation. Definitions are way too often drafted too narrowly or focused on capabilities that LAWS do not currently possess and are very unlikely to gain in the future.[9] As Taddeo and Blanchard argue, this approach is detrimental for two reasons. First, future-oriented definitions divert focus from the ethical and legal problems posed by existing or foreseeable AWS. Second, establishing a high threshold will undermine the regulation efforts as it leaves unaddressed other systems currently being developed. It does not enable the correct categorisation of these systems, which are autonomous, but that do not meet the high threshold.[10]

[6] Taddeo, M. and Blanchard, A. *A Comparative Analysis of the Definitions of Autonomous Weapons.* 10 May 2021, p. 4. At: http://dx.doi.org/10.2139/ssrn.3941214 (last accessed 13 April 2022).

[7] Ibid.

[8] LAWS represent a specific subset of AWS with the goal of exerting kinetic force against human beings. Throughout this paper, both terms will be used, as most of the problems related to autonomy apply in general. A distinction will be made when there is a problem specific to LAWS.

[9] Such would be the case of the UK's or France's definition, see section 1.1 below.

[10] Taddeo and Blanchard. *A Comparative Analysis of the Definitions of Autonomous Weapons,* supra note 6, pp 11-12.

This approach contravenes the purpose of the definition of AWS. The definition should serve as a tool to identify autonomous weapon systems and rule out systems that do not fall in this category. The goal of the definition, as the International Committee of the Red Cross ("ICRC") states, is that it "*encompasses some existing weapon systems,* [and so] *enables real-world consideration of weapons technology to assess what may make certain existing weapon systems acceptable - legally and ethically - and which emerging technology developments may raise concerns under international humanitarian law.*"[11]

While the issue of the definition of (L)AWS is not the focus of this paper (and proper analysis is definitely beyond its scope), several examples of definitions are provided to offer a better picture of which weapons are considered the subject of the debate.

1.1 Examples of definitions

The ICRC views AWS as systems that "*select and apply force to targets without human intervention. After initial activation or launch by a person, an autonomous weapon system self-initiates or triggers a strike in response to information from the environment received through sensors and on the basis of a generalized 'target profile'.*"[12]

Some define LAWS as "*systems that, once activated, can track, identify and attack targets with violent force without further human intervention*".[13] Others have argued that autonomy is the "*ability of a machine to perform a task without human input*".[14] Therefore, an "autonomous" system is one that "*once activated, can perform some tasks or functions on its own*".[15]

[11] ICRC. Views of the ICRC on Autonomous Weapon Systems. November 2016, p. 1. At: https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system (last accessed 13 April 2022).

[12] ICRC. Position on Autonomous Weapon Systems. Geneva. 12 May 2021. At: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems (last accessed 13 April 2022).

[13] Bode, I. and Watts, T. *Meaning-less Human Control.* Centre for War Studies, University of Southern Denmark with Drone Wars UK. 2021, p. 12. At: https://dronewars.net/wp-content/uploads/2021/02/DW-Control-WEB.pdf (last accessed 13 April 2022).

[14] Scharre, P. and Horowitz, M.C. *An Introduction to Autonomy in Weapon Systems*. CNAS. February 2015, p. 5. At: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf?mtime=20160906082257&focal=none (last accessed 15 April 2022).

[15] Boulanin, V. and Verbruggen, M. *Mapping the Development of Autonomy in Weapons Systems*. Stockholm: Stockholm International Peace Research Institute. 2017, p. 5. At: https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_ systems_11171.pdf (last accessed 15 April 2022).

The UN Special Rapporteur's report to the Human Rights Council on autonomous weapon systems provides another: "*Lethal Autonomous Robotics (LARs) refers to robotic weapon systems that, once activated, can select and engage targets without further intervention by a human operator. The important element is that the robot has an autonomous 'choice' regarding selection of a target and the use of lethal force*."[16]

The UK understands an autonomous system as "*capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present*."[17] As mentioned above, the high threshold established by the UK to identify AWS would allow an ever-increasing use of AWS insofar as these do not show "understanding higher-level intent and direction".[18]

Germany considers LAWS as "*weapons systems that completely exclude the human factor from decisions about their employment. Emerging technologies in the area of LAWS need to be conceptually distinguished from LAWS. Whereas emerging technologies such as digitalization, artificial intelligence and autonomy are integral elements of LAWS, they can be employed in full compliance with international law*."[19]

From the various examples of definitions of LAWS, several common elements can be extracted: (1) the focus is on weaponised systems that have a direct connection with targeting; (2) the systems are able to sense their environment; (3) based upon the data collected, the system is capable of "making decision" on targeting, based on the rules imbedded in its

[16] Heyns, C. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions.* UN General Assembly, A/HRC/23/47. 9 April 2013, para. 38. At: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (last accessed 13 April 2022).

[17] UK Ministry of Defence. Joint Doctrine Publication 030.2 Unmanned Aircraft Systems. August 2017. At: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf (last accessed 13 April 2022); UK Ministry of Defence. Joint Doctrine Note 2/11, The UK Approach to Unmanned Aircraft Systems. 2011, sec. 205. At: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505JDN_211_UAS_v2U.pdf (last accessed 13 April 2022).

[18] Taddeo and Blanchard. *A Comparative Analysis of the Definitions of Autonomous Weapons,* supra note 6, pp 11-12.

[19] CCW. Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, Annex III, 11 guiding principles. CCW/GGE.1/2020/WP.7. 19 April 2021, para. 1. At: https://documents.unoda.org/wp-content/uploads/2020/07/CCW_GGE1_2020_WP_7-ADVANCE.pdf (last accessed 13 April 2022).

algorithm; (4) human operators are not (directly) involved in the system's final decision; and (5) violent force is used against the target. A comparative analysis of existing definitions of AWS has identified four key aspects as the essential factors to define AWS: autonomy, adapting capabilities of AWS, human control, and purpose of use.[20]

The Group of Governmental Experts meetings held under the UN Convention on Certain Conventional Weapons ("CCW GGE") has, in the past, focused on defining LAWS. In 2014, it concluded that while the elaboration of a definition was premature, "*autonomy should be measurable and should be based on objective criteria such as capacity of perception of the environment, and ability to perform pre-programmed tasks without further human action*."[21] As it appeared to be challenging to define the concept of autonomy, in 2016 the focus shifted on the functions of a system, which should provide a better understanding of autonomy in LAWS.[22] In 2019, it was suggested to view autonomy instead as a spectrum and the term as covering a wide range of technical capabilities. It was concluded that the "*role and impacts of autonomous functions in the identification, selection or engagement of a target are among the essential characteristics of weapons systems.*" [23]

The CCW GGE underlines that weapon systems are made up of subsystems, which may themselves be used during targeting. This makes it difficult not only to characterise LAWS precisely but also to fully understand how autonomy may impact the ability of parties to a conflict to apply IHL and comply with its rules.[24] With minimal changes to the definition of LAWS, certain types of weapons may be excluded.

[20] Taddeo and Blanchard. *A Comparative Analysis of the Definitions of Autonomous Weapons,* supra note 6, p. 6.

[21] CCW. Report of the 2014 informal Meeting of Experts on Lethal Autonomous Weapons Systems. CCW/MSP/2014/3. 11 June 2014, p. 4. At: https://meetings.unoda.org/section/ccw-gge-2014-documents/ (last accessed 13 April 2022).

[22] CCW. Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS). Submitted by the Chairperson of the Informal Meeting of Experts. 2016, para. 33. At: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Informal_Meeting_of_Experts_(2016)/ReportLAWS_2016_AdvancedVersion.pdf (last accessed 13 April 2022).

[23] CCW. 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/Add.1. 8 November 2019, para. 5. At: https://documents.unoda.org/wp-content/uploads/2020/09/1919338E.pdf (last accessed 13 April 2022).

[24] Ibid.

1.2 Autonomy and human control

When it comes to human control over weapon systems, there are diverging views on its role in determining whether the system is autonomous. Waxman and Anderson argue that "*whether a system is merely highly automated or genuinely autonomous might well depend less on the machine's design than on the anticipated role for the human operators. If they cannot reasonably perform that role* [...]*, a system believed to be merely automated to a limited point might turn out to be effectively autonomous.*"[25] This approach considers that a lack of human control is one of the defining features of autonomy. The German definition, for example, mentions machines that "completely exclude" humans from the decision-making process.

Others argue that autonomy is not defined with respect to human control but rather to the intervention of another agent in the functioning of AWS. Human control is seen as a mode of deploying AWS and not as one of their defining characteristics.[26] According to Taddeo and Blanchard, "[a]*n artificial system can be fully autonomous, insofar as it can operate independently from a human or of another artificial agent, and yet be deployed under some form of human control.*"[27] Proponents of this approach argue that the distinction between autonomy and control is essential for three reasons. First, it avoids considering automation/autonomy and human control as mutually exclusive concepts. Human intervention may be unnecessary in (at least) automated systems, but automation does not make human control impossible.[28]

Second, it secures that any conclusions reached will also be applicable in the future. Much of the debate focuses on the desirable level of control over LAWS rather than the desirable level of their autonomy. Separating the notion of human control from the definition of autonomy may enable us to focus the normative efforts on the level of human control required, irrespective of technological progress.[29]

[25] Anderson, K. and Waxman, M. C. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can.* Stanford University, The Hoover Institution (Jean Perkins Task Force on National Security and Law Essay Series). 10 April 2013, p. 22. At: https://ssrn.com/abstract=2250126 (last accessed 13 April 2022).

[26] Taddeo and Blanchard. *A Comparative Analysis of the Definitions of Autonomous Weapons,* supra note 6, p. 21.

[27] Ibid, p. 13.

[28] Ibid.

[29] Ibid.

The third advantage of this distinction is that it "*pre-empts approaches that leverage the lack of existing examples of fully autonomous AWS to avoid discussing their regulation*."[30] An argument can be added that it lowers how much the debate on LAWS is future-oriented and enables to reconsider even the use of current weapon systems that are not fully autonomous.

Irrespective of the approach one takes, it is beneficial to focus on how human control should be exercised rather than on defining autonomy and describing its levels. The US DoD Defense Science Board's task force has reviewed many studies on "levels of autonomy" and concluded that they are not particularly helpful to the autonomy design process. They recommended that the DoD abandon the use of "levels of autonomy" because *"they focus too much attention on the computer rather than on the* collaboration *between the computer and its operator/supervisor to achieve the desired capabilities and effects*".[31]

1.3 Types of weapon systems

To allow for a better picture of the weapons in question, it is helpful to introduce various types of weapons that have been discussed in the context of LAWS. However, it should be noted that even if some weapon developers may provide the information on possible autonomous functions, information is lacking on the degree of autonomy in which they were deployed in real-world situations by States.

1. 3. 1. Air weapon systems

In the air, it is foreseen that unmanned air systems may be used for air-to-air combat, electronic warfare, and suppression of air defences, in addition to their current use for targeted strikes.[32]

The attention has been focused on autonomous drones. Apart from drones, examples of an application of complex autonomous technology in military systems are the American X47-B and the comparable British system called Taranis, the Russian MiG Skat, the European nEUROn, and the Chinese Anjian. These unmanned combat air systems can autonomously

[30] Ibid.

[31] US DoD Defense Science Board. The role of autonomy in DoD systems, Task Force Report. July 2012, p. 4. At: www.fas.org/irp/agency/dod/dsb/autonomy.pdf (last accessed 13 April 2022).

[32] US DoD. Unmanned Systems Integrated Roadmap 2013, FY2013-2038, p. 24. At: https://www.hsdl.org/?abstract&did=747559 (last accessed 13 April 2022).

perform complex tasks, such as taking off from and landing on an aircraft carrier, conducting mid-flight refuelling, and taking evasive manoeuvres.[33]

Other examples include "homing" munitions that, once launched to a particular target location, search for and attack pre-programmed categories of targets (e.g., tanks) within the area. These weapons engage a specific target pre-selected by a human operator.[34] Depending on how broad a definition is used, such weapons may not fall under the regulations of LAWS because of their "lack" of autonomy.

An example of a fully autonomous weapon system can be certain "loitering" munitions that, once launched, search for and attack their intended targets over a specified area and without any further human intervention.[35] Here, the lines between unmanned combat air vehicles and missiles become increasingly blurred. Certain loitering munitions are essentially "*unmanned air systems that integrate a weapon as part of their construction*."[36]

1. 3. 2. Fixed and mobile ground weapon systems

Popular candidates for being considered autonomous (at least in some of their functions) are defensive weapon systems used to attack incoming missile or rocket attacks. These have pre-programmed categories of targets, among which they independently select and attack. However, a human retains supervision of the weapon operation and can override the system within a limited time period (so-called "veto" power).[37] These include weapons systems that Sense and React to Military Objects (SARMO) for protection against fast incoming munitions such as mortar shells and missiles (e.g., C-RAM, Phalanx, Mantis). None of these is fully autonomous yet; they are programmed to perform a small set of defined actions on repeat mechanically. As far as is known, they are used in highly structured and predictable

[33] Ekelhof, M. A. C. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting. *Naval War College Review.* 2018 71(3), para. 14. At: https://digital-commons.usnwc.edu/nwc-review/vol71/iss3/6/ (last accessed 13 April 2022); Slijper, F. *Where to Draw the Line: Increasing Autonomy in Weapon Systems - Technology and Trends*. Utrecht, Neth.: PAX, 2017, p. 10. At: www.paxvoorvrede.nl/ (last accessed 14 April 2022).

[34] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*. Geneva: March 2014, p. 6. At: https://shop.icrc.org/expert-meeting-autonomous-weapon-systems-technical-military-legal-and-humanitarian-aspects.html?___store=en (last accessed 13 April 2022).

[35] Ibid.

[36] Davison, N., Weizmann, N. and Robinson, I. Background Paper by the International Committee of the Red Cross. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014, p. 58. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

[37] Ibid.

environments with a very low risk of incidental harm. They do not move around, and there is constant human evaluation and monitoring for a rapid shutdown.[38]

Current weapon systems with the highest degree of autonomy are fixed in stationary roles than mobile.[39] Various unmanned ground systems have been equipped with weapons to enable remote operation and potentially a certain level of autonomy. The main potential military uses are to access areas dangerous to humans, as well as bomb disposal or other violent use.[40] The main problem the developers are facing is creating versatile machines that can adapt to arbitrary environments. Currently, there are "*quadruped robots that are able to walk on complicated terrain and recover from strong, unexpected external pushes*."[41] However, it is still very difficult to conceive algorithms that can "*compute the necessary motions for a robot to cross arbitrary obstacles or react to any type of external variation*."[42]

1. 3. 3. Maritime weapon systems

At sea, unmanned underwater and surface vehicles may be used to lay and destroy mines, reconnaissance, and other armed operations.[43]

The use of autonomous functions for unmanned underwater vehicles is very appealing due to the difficulties of communication underwater and the vast areas these vehicles may be operating in. These weaponised vehicles can conduct their tasks without human interaction for many days.[44] At the same time, autonomy in their functions may not be seen as problematic due to the nature of the environment these vehicles operate in, as will be discussed later.[45]

[38] Sharkey, N. Autonomous weapons and human supervisory control. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014, p. 29. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).
Ibid.

[39] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 65.

[40] Ibid, p. 66.

[41] Righetti, L. Civilian robotics and developments in autonomous systems. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014, p. 26. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

[42] Ibid.

[43] US DoD. Unmanned Systems Integrated Roadmap 2013, supra note 32, p. 24.

[44] US DoD, Defense Science Board. Task Force Report: The Role of Autonomy in DoD Systems. 19 July 2012, p. 86. At: https://irp.fas.org/agency/dod/dsb/autonomy.pdf (last accessed 13 April 2022).

[45] See Chapter VI section 3.2.3 below.

2. Categorisation of autonomous weapon systems

When considering the requirement of meaningful human control in the context of LAWS, several categorisations are of relevance. They help form a clearer picture of which categories of weapons should be subject to the requirement, or they enable to draw a line between weapons that some consider unacceptable.

2.1 Automated and autonomous weapons

The first categorisation distinguishes between "automated" and "autonomous" weapons. It serves to distinguish weapons that may rely on computer software to a certain extent but do not possess the quality of being "autonomous"; therefore, the current debate and possible legal regulation do not cover them. Nevertheless, it is not a clear-cut distinction, and certain weapons may be difficult to categorise.

Automated systems operate on pre-programmed instructions to carry out a specific task; they act based on deterministic (rule-based) instructions. On the other hand, autonomous systems act dynamically to decide if, when, and how to carry out a task. They act on probability-based reasoning, which necessarily introduces uncertainty.[46]

In other words, automation means "*running through a fixed pre-programmed sequence of action*".[47] In contrast, autonomy means that "*actions are determined by its sensory inputs, rather than where it is in a pre-programed sequence*".[48] It follows that automated systems are capable of less sophisticated acts than autonomous systems since they cannot diverge from their program.[49]

The notion of "autonomous weapon systems" still presents a whole spectrum, rather than a single type of weapons. It includes both mobile weapon systems able to adapt to changing circumstances and freely determine their targets on the one side, and fixed weapon systems that have pre-defined limitations on their operation and potential targets on the other side.[50]

[46] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*, supra note 34, p. 5.

[47] Winfield, A. F. T. *Robotics: A Very Short Introduction.* Very Short Introductions 330. Oxford: Oxford University Press, 2012, p. 12.

[48] Ibid.

[49] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 15.

[50] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 64.

Taddeo and Blanchard, however, argue that "[w]*hile one may agree that the distinction between automation and autonomy is blurred, this is not because the assessment of autonomy of artificial agents is subjective or context dependent. Within the field of computer science, and particularly of Agent Theory* […], *there is quite a clear understanding of the differences between these concepts.*"[51]

Based upon their characteristics, AWS can be divided into several categories: "*AWS execute tasks to achieve goals (teleological agents), they can adjust their actions based on the feedback that they receive from the environment (automated artificial agents), may also be able to define plans (heuristic artificial agents) to achieve their goals, and may be able to refine their behaviour in response to the changes in the environment (adapting artificial agent).*"[52]

Worth noting is that the literature on military robotic systems and autonomy creates a separate category of remote-controlled or teleoperated systems, which are controlled directly by a remote operator.[53] This distinction is correct as long as the human operator exercises control over all the critical functions of the weapon. For example, current armed unmanned air systems (e.g., "drones") are operated remotely since targeting and firing are carried out by a human operator from a distance, which excludes them from the category of autonomous or automated weapons. Arguably, even if autonomy were present in some non-critical functions, such as flight control or landing, they still would not be considered LAWS. However, once a drone is "free" to select from several targets based on a pre-programmed target profile, it seems like the line of autonomy is crossed.[54]

2.2 The loop scheme

The second categorisation addresses the spectrum of autonomous weapons and distinguishes between different levels of human involvement in the deployment of LAWS.

The US Department of Defence policy divides autonomous weapons into three types according to the level of autonomy and the level of human control in the following way:

[51] Taddeo and Blanchard. *A Comparative Analysis of the Definitions of Autonomous Weapons,* supra note 6, p. 19.

[52] Ibid.

[53] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 62.

[54] Ibid, p. 63.

Semi-autonomous weapon system (also referred to as human "in-the-loop"): "*A weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator.*"[55]

Human in-the-loop implies a human operator behind every decision to authorise and attack targets. According to some opinions, these weapons are not considered fully autonomous and should thus not be subject to the proposed regulation of LAWS. However, coming back to the issue of defining LAWS, even these semi-autonomous weapon systems possess the ability to carry out some of their functions independently on humans after having been deployed. The difference lies in the scale of their autonomous functions. Therefore, it would be wrong to exclude these types of weapon systems from the debate.

Supervised autonomous weapon system (also referred to as human "on-the-loop"): "*An autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur.*"[56]

The human operator on-the-loop supervises the functioning of the weapon system and has the possibility to intervene if a failure is observed or additional verification of a target is needed. However, the weapon system does not require authorisation prior to every engagement. The human operator has the possibility to veto the use of force. With this category, it is the most difficult to determine to what extent a weapon system is reliant on its human operator and in which functions it is "fully" autonomous.

Autonomous weapon system (also referred to as human "out-of-the-the-loop"): "*A weapon system that, once activated, can select and engage targets without further intervention by a human operator.*"[57]

Depending on the approach one takes, only this category of weapon systems can be seen as genuinely autonomous. It is because the weapon system does not need the authorisation to engage a target, and human operators also do not intervene. The system functions on its own after activation.

[55] US DoD. Directive 3000.09 on Autonomy in Weapon Systems, supra note 2, Glossary, Part II Definitions.

[56] Ibid.

[57] Ibid.

2.3 AI and machine learning

While existing weapons might incorporate autonomy in their critical functions, they tend to use relatively simple, rule-based control software to select and attack targets. However, AI and machine-learning software could form the basis of future autonomous weapon systems. These technologies could specifically be developed for "automatic target recognition".[58]

According to the ICRC[59], AI is the "*use of computer systems to carry out tasks previously requiring human intelligence, cognition or reasoning;[60] and machine learning involves AI systems that use large amounts of data to develop their functioning and "learn" from experience*.[61]"

The ICRC is the most concerned with the use of AI and machine learning for decision-making. It could enable widespread collection and analysis of data sources to identify people or objects, assess patterns of life or behaviour, make recommendations for military strategy or operations, or make predictions about future actions or situations.[62] As computers are generally better at collecting, sorting, and analysing large amounts of data, AI and machine learning-based decision-support systems may be beneficial in enabling better compliance with international humanitarian law. The other side of the coin is that the current technology is limited. These algorithmically generated predictions suffer from unpredictability and possible bias induced by their programming.[63]

3. Shifting the focus to the level of control required

As has been argued above, the task of defining LAWS seems rather Herculean. At the same time, the scope of regulation is bound to be determined by the definitions used. Thus, many

[58] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach. Geneva: 6 June 2019, p. 3. At: https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach (last accessed 13 April 2022); ICRC. Statement to the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts on Lethal Autonomous Weapons Systems under agenda item 6(b). Geneva: 27-31 August 2018. At: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2018)/2018_GGE%2BLAWS%2B2_6b_ICRC.pdf (last accessed 13 April 2022).

[59] Ibid, ICRC, Artificial intelligence and machine learning in armed conflict: A human-centred approach, p. 1.

[60] Oxford English Dictionary. "Artificial intelligence". At: https://www.lexico.com/definition/artificial_intelligence (last accessed 13 April 2022).

[61] Oxford English Dictionary. "Machine learning". At: https://www.lexico.com/definition/machine_learning (last accessed 13 April 2022).

[62] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 4.

[63] Ibid, p. 5.

delegations in the CCW Group of Experts believed that "*a technology-neutral approach, focusing on the human element in the use of force, would be more fruitful than taking forward detailed discussions on technical characteristics.*"[64]

Although having a clear-cut definition of LAWS would undoubtedly be desirable, it could prove to be inflexible in the long run. The debate on LAWS is very future-oriented and aims to address weapon systems that have not even been developed yet.[65] Technological experts can perceive in which direction the development might be heading but still, it cannot be determined with certainty which weapons will be developed and what will be their technological parameters. Focusing on human-machine has its benefits in being flexible and perhaps more fitting to address various types of LAWS that may be developed in the future. By requiring a certain level of human control over all weapon systems featuring autonomy in their functions, we would simply determine the boundaries of what is deemed still acceptable.

4. Regulation of autonomous functions of LAWS

Another problem that could arise from formulating a definition of LAWS too narrowly is that certain functions or categories of systems could be left out of the regulation. However, autonomous features should be regulated no matter whether as part of physical or cyber-weapon systems or in decision-support systems.

These "decision-support" or "automated decision-making" systems are effectively an expansion of intelligence, surveillance, and reconnaissance tools. They might be using AI and machine learning to automate the analysis of large data sets to provide "advice" to human operators in making particular decisions.[66]

The range of possible uses of these systems is extensive, from targeting decisions, decisions about whom to detain and for how long, to decisions about military strategy and specific operations, including attempts to predict or pre-empt adversary operations.[67] Even though this

[64] Report of the 2019 session of the GGE, supra note 23, para. 14.

[65] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches. A Primer. Geneva: UNIDIR, 2017, p. 10. At: https://www.unidir.org/publication/weaponization-increasingly-autonomous-technologies-concerns-characteristics-and (last accessed 13 April 2022); Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 28.

[66] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 4-5.

[67] Ibid, p. 5.

"advise" does not equal a targeting decision to use lethal force, it nonetheless forms a part of the same decision-making process. Some claim that such technologies are often underappreciated in the discourse on LAWS because they are not weaponised. These technologies may be closely connected to target selection since the intelligence produced by the human-machine collaboration may result in targets being selected for engagement on the battlefield.[68]

To the ICRC, an extensive range of different AI-influenced decisions by conflict parties can be relevant, especially where they pose risks of injury or death to persons or destruction of objects and where specific rules of international humanitarian law govern the decisions.[69] In the ongoing discussion, it is primarily the types of autonomous weapons that directly pose risks of injury or death to persons or destruction of objects that are considered the object of any future regulation. However, it has been mostly agreed that also weapons with autonomy in their "critical functions" would fall under the definition of LAWS. The notion of critical functions is rather unhelpful as its definition is equally challenging to formulate. Consent has been reached to include stages of target selection, targeting, and deployment. Autonomy in these functions can be understood as the capacity of a machine, following activation, to operate without any external control in some or all areas of its operation for extended periods.[70]

[68] Ekelhof, M. A.C. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting, supra note 33, p. 20.

[69] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 5.

[70] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 62.

5. Current use and drivers for autonomy

As of now, States claim that unmanned systems are used to deliver weapons, but the decision to fire at a specific target is taken by a person and not a machine.[71] Even though drones with the possibility of functioning in an autonomous mode have already been deployed, there is a lack of information on whether they were manually operated or steered themselves using machine vision.[72]

There are several reasons why militaries explore increasing autonomy in the weapon systems they have available. Unmanned systems offer several advantages: force multiplication, reduced risk to military personnel, increased capability over a wider area and deeper in the adversaries' territory, increased persistence on the battlefield, and all this at a potentially lower cost.[73]

Increasing autonomy in some features of weapon systems could yield benefits not only to the military but also to the protection of civilians, especially when it comes to tasks that computers perform better than humans. The cooperation of computers and humans offers multiple opportunities in terms of performance enhancement while retaining human control over the weapon system. On the other hand, there are also specific incentives for increasing autonomy in order to reduce the level of human control.[74] These include: decreasing the necessary number of humans operating unmanned systems; reducing the reliance of the weapon systems on communications links; and increasing their performance and speed of decision-making, which outperforms human-machine cooperation.[75] The following analysis will address the possible advantages of retaining meaningful human control over autonomous weapons systems.

[71] US DoD. Directive 3000.09 on Autonomy in Weapon Systems, supra note 2, p. 24.

[72] UN Security Council. Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011). S/2021/229. 8 March 2021, para. 63. At: https://undocs.org/Home/Mobile?FinalSymbol=S%2F2021%2F229&Language=E&DeviceType=Desktop (last accessed 13 April 2022); Vincent, J. Have autonomous robots started killing in war? The Verge. 3 June 2021, At: https://www.theverge.com/2021/6/3/22462840/killer-robot-autonomous-drone-attack-libya-un-report-context (last accessed 13 April 2022).

[73] Marchant, G., Allenby, B., Arkin, R., Barrett, E., Borenstein, J., Gaudet, L., Kittrie, O., Lin, P., Lucas, G., O'Meara, R. and Silbermann, J. International Governance of Autonomous Military Robots. *Columbia Science and Technology Law Review.* 2011. XII. 272-315. At: https://academiccommons.columbia.edu/doi/10.7916/D8TB1HDW (last accessed 13 April 2022).

[74] US DoD. Directive 3000.09 on Autonomy in Weapon Systems, supra note 2, p. 3.

[75] Ibid; UK Ministry of Defence, Joint Doctrine Note 2/11, The UK Approach to Unmanned Aircraft Systems, supra note 17, pp. 5-10.

II. Why do we need MHC?

Before defining the parameters of MHC, it is essential to have a clear idea of which issues do we seek to solve by requiring that a certain defined level of human control over LAWS is retained. The following Chapter will present several such reasons, from technological limitations, compliance with IHL rules on the conduct of hostilities, ethical and moral reasons, to operationalising the framework of individual criminal responsibility.

1. Technological limitations

The preliminary issue that must be discussed to understand why we need MHC are the limits inherent in current technology.

Several of such limitations are particularly relevant for military applications. Firstly, current autonomous systems are "brittle" (not adaptable and easily break down), which makes them unreliable. Secondly, existing autonomous systems need human input for many functions to correct mistakes. Thirdly, there is no standard review process or methodologies to test and validate autonomous systems. Finally, there is the limited ability of autonomous robotic systems to perceive the environment in which they operate.[76]

Given sufficient time and finances to boost research, some technological limitations are likely to be overcome, such as computational power, actuation and sensor quality and density.[77] However, certain challenges seem to have no solution in the foreseeable future, e.g., creating algorithms that make sense of the real world in a way similar to humans and make reliable decisions consistently, or developing machines that can move around in and adapt to changing and unpredicted environments.[78] Today's systems can only demonstrate reliable, consistent, trusted performance when placed in known environments which are predictable and well understood.[79]

[76] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*, supra note 34, p. 5.

[77] Righetti. Civilian robotics and developments in autonomous systems, supra note 41, p. 27.

[78] Ibid.

[79] Ansell, D. Research and Development of Autonomous 'Decision Making' Systems. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014, p. 40. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

1.1 Object recognition

The first limitation lies in the fact that a system must be able to use sensors to sense, receive, and perceive information about its environment. In some cases, sensors' raw performance for capturing electromagnetic data outperforms human beings' ability, such as digital cameras with enormous resolution and focus.[80] However, despite their capabilities, even these are not flawless. They are dependent on processing the images they receive and are susceptible to manipulation. External interference can add "artefacts" to imagery which causes the system to fail in its task; it cannot recognise the object it "sees". Deliberate external means of disrupting or spoofing the sensing action can also take place.[81] That is of particular importance in the context of deploying autonomous weapons in armed conflicts. An adversary can simply spoof the system by adding artefacts to military objectives so that they are not recognised as such.

In the domain of image recognition, the deep neural network-based approach has outperformed traditional image processing techniques, achieving even human-competitive results.[82] However, studies have revealed that even the slightest disturbance introduced to natural images (such samples are called "adversarial images") can make the deep neural network misclassify the objects.[83] Even one-pixel attacks created with an algorithm successfully fooled three types of deep neural networks.[84]

Even outside of the domain of LAWS, there is still a common misconception that one day, machines will be developed which overcome humans in all their capabilities. This assumes that computational processes could work by analogy with the human brain. Nevertheless, scientists have yet to discover the way how the brain processes information, and it may not be

[80] Ibid, p. 39.

[81] Ibid.

[82] Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. Deepface: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014. 1701-1708. At: https://ieeexplore.ieee.org/document/6909616 (last accessed 13 April 2022).

[83] Goodfellow, I. J., Shlens, J. and Szegedy, C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572.* 2014. At: https://arxiv.org/abs/1412.6572 (last accessed 13 April 2022); Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B. and Swami A. The limitations of deep learning in adversarial settings. *Security and Privacy (EuroS&P), 2016 IEEE European Symposium*. 2016. 372–387. At: https://arxiv.org/abs/1511.07528 (last accessed 13 April 2022).

[84] Su, J., Vasconcellos Vargas, D. and Sakurai, K. One Pixel Attack for Fooling Deep Neural Networks. *IEEE Transactions on Evolutionary Computation.* 2019. 23(5). 828-841.

information processing in abstract computational terms at all.[85] It is thus necessary to differentiate between computation and information processing by the human brain. Generally speaking, "*information processing can be any physical process which transforms an input into an output*."[86] Computation is understood as "*syntactic and symbolic manipulation of information*."[87] In this sense, computation is an algorithmic and deterministic type of information processing.[88] Signorelli explains that the human brain processes information in a more complicated manner. It not only performs computation but can also give interpretations and meaning to its own high-level information processing.[89] This semantic gap shows that humans and machines carry out tasks very differently.[90] An algorithm has no understanding of the meaning or concept of a subject, which means it can make mistakes such as classifying an object as something completely different and unrelated.[91]

1.2 Classification of objects and creating plans

The second limitation flows from the way how a system creates and executes plans of action. After having obtained the visual and other data, the system must then extract information from the data. The whole process bears little similarity to the way how human brain classifies objects. It is very challenging to write a program that fulfils similar functions. Often image-processing techniques are used to classify an object by matching it against a database of similar images (looking for correlation). Usually, the output from these software programs is a classification accompanied by some form of confidence or error rating (e.g., a tank, with 70% certainty).[92]

The next step is to assemble the individual classifications in order to make a diagnosis of the actual situation. The machine infers knowledge, which is often described in the computer

[85] Epstein, R. *The empty brain*. Aeon. 2016. At: https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer (last accessed 13 April 2022).

[86] Signorelli, C. M. Can Computers Become Conscious and Overcome Humans? *Frontiers in robotics and AI*. 2018/5, p. 5. At: https://doi.org/10.3389/frobt.2018.00121 (last accessed 13 April 2022).

[87] Searle, J. R. Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association.* 1990. 64(3). 21-37. At: https://philosophy.as.uky.edu/sites/default/files/Is%20the%20Brain%20a%20Digital%20Computer%20-%20John%20R.%20Searle.pdf (last accessed 13 April 2022).

[88] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 5.

[89] Ibid.

[90] Smeulders, A. et al. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 22(12). 2000. 1349–1380. At: https://ieeexplore.ieee.org/document/895972 (last accessed 2022).

[91] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 11.

[92] Ansell. Research and Development of Autonomous 'Decision Making' Systems, supra note 79, p. 39.

science world as “creating beliefs”. By combining smaller pieces of information, the machine will create many beliefs.[93] These beliefs can only be as certain as was the original classification of the objects sensed.

Further, the robotic system must create a plan based on its previous classifications and beliefs. A plan may be straightforward (look right) but also highly complicated (follow a missile’s trajectory and shoot it down at the moment when it will cause the least damage to the city underneath). The more complex data in the beginning and the task to be carried out, the more complicated the planning process is. Moreover, specific goals and aims have to be integrated into the software to which the system must be able to compare the proposed plan.[94]

In some systems, the planning functions generate more alternative plans, and the system must choose between them according to its goals. This selection process is a form of machine decision-making, where the software must be able to form predictions for several plans and select one. The selection process can be a source of error as the machine may make its selection without full knowledge of, for example, laws and regulations. It is common practice for the machine to interact with a human being at this stage in order to seek authorisation.[95]

1.3 The Precision-Recall Trade-Off

Further limitations are caused by the so-called “precision-recall trade-off”, which is a known phenomenon in the context of machine learning. Precision and recall are metrics of performance for classification algorithms.[96] In other words, they represent two factors that a programmer considers in determining how successful the algorithm is in classifying objects. Precision is how many times an accurate prediction of a particular class occurs per a false prediction of that class. Recall is the percentage of the data belonging to a particular class that the model correctly predicts as belonging to that class.[97]

The traditional way to think of this is first to define true positives, false positives, and false negatives. A true positive is a correct prediction, and the data point belongs to the positive class. A false negative is an incorrect prediction where the actual value is positive, and the

[93] Ibid.

[94] Ibid, p. 40.

[95] Ibid, p. 40.

[96] Bennett, G. The Precision-Recall Trade-Off. 21 June 2020. At: https://datascience-george.medium.com/the-precision-recall-trade-off-aa295faba140 (last accessed 13 April 2022).

[97] Ibid.

predicted value was negative. A false positive is a prediction that was also incorrect because the prediction was positive, but the actual value was negative.[98]

Precision can then be defined as the number of true positives divided by the sum of true positives and false positives. False positives should not be in the category of positives; therefore, the lower their number is, the more precise classification is. In the context of targeting carried out by LAWS, a true positive would be, for example, a combatant. A false positive could be a law enforcement officer carrying his weapon openly.

Recall can be defined as the number of true positives divided by the sum of true positives and false negatives. This time we consider false negatives because they were predicted as negatives by the algorithm; however, their actual values are positive. Hence, they should be included in the category of positives. Looking at an example from a situation of an armed conflict, a false negative could be civilians directly participating in hostilities. LAWS would classify them as "civilian" while, in reality, they fall under the category of a lawful target (positives in our scenario).

Unfortunately, it is not possible to have both precision and recall high. Once precision is increased, recall will reduce and vice versa. This is called the precision-recall trade-off.[99] In other words, lowering the tolerance of false positives induces a higher number of false negatives. The trend is the following: for precision to be 100%, recall will be roughly around 40%. The most balanced option is a trade-off point where precision is nearly 87% and recall is about 70%.[100]

In the context of an armed conflict, LAWS can be programmed to be highly cautious about false positives (not targeting civilians or persons *hors de combat*). This will, however, increase the number of false negatives (classifying lawful targets, such as civilians directly participating in hostilities, as unlawful targets). This effect presents a great difficulty for militaries. To be able to use LAWS in IHL-compliant mode (programming the systems to be exact and avoid targeting unlawful objects), the weapon system necessarily becomes overly "cautious". It renders a higher number of false negatives, which means it will not target the number of military objectives representing the category of false negatives.

[98] Ibid.

[99] Lendave, V. Python Guide to Precision-Recall Tradeoff. Developers Corner. June 10, 2021. At: https://analyticsindiamag.com/python-guide-to-precision-recall-tradeoff/ (last accessed 13 April 2022).

[100] Ibid.

1.4 Bias

On top of all that, bias in the algorithm presents an additional issue. It can reinforce existing human bias or introduce a new one through the design or the use of the system.[101] A common form of bias stems from training data. AI systems learn to make decisions based on training data, including biased human decisions, or reflecting historical or social inequities, even if sensitive variables such as gender, race, or sexual orientation are removed.[102] An algorithm depends on the quantity, quality, and nature of available data to train it for a specific task. This matters even more in armed conflict, where data high in quality and quantity for particular tasks can be challenging to gather. However, bias can also derive from allocating importance to different data elements by the system or its interaction with the environment during a task.[103]

For example, an investigative news site has found that a criminal justice algorithm used in Florida mislabelled African American defendants as "high risk" at nearly twice the rate it mislabelled white defendants. Other research has found that training natural language processing models on news articles can lead them to exhibit gender stereotypes.[104]

This does not mean that it is only AI displaying bias. Gender inequality and discrimination pre-existing an armed conflict create particular challenges for women and girls even with no involvement of LAWS. Many endure extreme hardships, including increased insecurity, restricted mobility, sexual exploitation and abuse, and gender-based violence.[105] However, it is vital to bear in mind that AI may likely mirror existing human bias if particular attention is not paid to this issue.

[101] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 11.

[102] Manyika, J., Silberg, J. and Presten, B. What Do We Do About the Biases in AI? Harvard Business Review. 25 October 2019. At: https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai (last accessed 13 April 2022).

[103] UNIDIR. Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies. A Primer. August 2018. At: https://unidir.org/publication/algorithmic-bias-and-weaponization-increasingly-autonomous-technologies (last accessed 13 April 2022).

[104] Manyika, Silberg, and Presten. What Do We Do About the Biases in AI?, supra note 102.

[105] ICRC. Addressing Internal Displacement in Times of Armed Conflict and Other Violence, 2018. At: https://shop.icrc.org/addressing-internal-displacement-in-times-of-armed-conflict-and-other-violence-pdf-en.html (last accessed 13 April 2022).

1.5 Unpredictability

Since autonomous systems are adaptable (within their pre-programmed boundaries), they are necessarily unpredictable. As argued above, numerous errors may appear during object recognition, planning, and taking action. Their unpredictability is only reinforced by the issues of false positives and any possible bias. Due to the number of possible situations LAWS might face, it is virtually impossible to test a system for all possible scenarios. Moreover, autonomous weapon systems might have to rely on artificial neural networks to conduct the whole process of targeting.[106] These networks evolve in a non-deterministic way thanks to self-learning and training from some given rules. In Signorelli's opinion, "*it is not always possible to ensure what the net is learning, nor control the dynamic evolution of its learning process, even if deterministic learning rules have been given*."[107] One can simply never know what the network has learned until it is tested. And even after testing, it is never possible to be sure which layer of the programme encodes which statistical property of the data. Hence, it is not possible to predict how the net will behave.[108]

This presents a challenge for complying with IHL, pre-eminently the obligation to conduct a legal review of weapons.[109] Under this obligation, States must ensure that any new weapon or means and methods of warfare they develop or acquire will not violate their legal obligations when used.[110] When it comes to LAWS, States possibly cannot review the weapon systems for every possible situation, nor are they required to do so. According to the ICRC Commentary, States need only to determine "*whether the employment of a weapon for its* ***normal or expected use*** *would be prohibited under some or all circumstances. A State is not required to foresee or analyse all possible misuses of a weapon, for almost any weapon can be misused in a way that would be prohibited.*"[111]

[106] See section 2.2.2 below.

[107] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 5

[108] Ibid.

[109] Enshrined in Art. 36 of AP I, which states: "*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.*" There is disagreement on whether this rule is a part of customary law.

[110] ICRC. Guide to the Legal Review of New Weapons, Means and Methods of Warfare. Geneva: 2006, p. 4. At: http://www.icrc.org/eng/assets/files/other/icrc_002_0902.pdf (last accessed 13 April 2022).

[111] ICRC. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*. The Netherlands: Martinus Nijhoff Publishers, 1987, para. 1469 (emphasis added).

Therefore, in assessing the legality of LAWS, States should look at the normal or expected circumstances of their use since any weapon could potentially be used indiscriminately or disproportionately. To discharge this obligation faithfully, they must foresee which tasks the weapon system will perform, in which environment, and how will it function. It may as well be the case that the review shows that a particular autonomous weapon system could be used lawfully only in limited circumstances. These limits then must be incorporated into the instructions and rules of engagement applying to the weapon.[112] This, however, presupposes that the behaviour of the weapon system must be predictable to the extent that its "normal or expected use" can be determined. If the review concludes that it cannot be foreseen how the weapon will perform in the environment in which it is intended to be deployed, then the deployment of such a weapon is unlawful. The question remains, what degree of certainty is required for LAWS, whether the standard should be human behaviour (which is imperfect) or different. It is unclear how both hi-tech and low-tech nations could guarantee the quality of Article 36 weapon reviews.

1.6 Morality

Yet another limitation appears if morality is to be implemented into systems. One of the most distinctive features of human intelligence can be considered "*the capability to integrate rational and emotional thinking to take moral decisions which are adapted to the context*."[113] It is one of the arguments put forward by proponents of a ban on LAWS that these weapon systems are incapable of moral judgement required by IHL. A Canadian robotics manufacturer, Clearpath Robotics, became the first company publicly to refuse to manufacture "weaponised robots that remove humans from the loop."[114] In a letter to the public, the company phrased its views in the following way: "[W]*ould a robot have the morality, sense, or emotional understanding to intervene against orders that are wrong or inhumane? No. Would computers be able to make the kinds of subjective decisions required for checking the legitimacy of targets and ensuring the proportionate use of force in the foreseeable future? No.*"[115]

[112] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 76.

[113] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 5.

[114] Hennessy, M. Clearpath Robotics Takes Stance Against 'Killer Robots. Clearpath Robotics press release. 13 August 2014. At: https://www.clearpathrobotics.com/2014/08/clearpath-takes-stance-against-killer-robots/ (last accessed 13 April 2022).

[115] Ibid.

Whether targeting decisions necessarily have to be subjective will be discussed later.[116] However, it is true that morality can play a role in an armed conflict, particularly in the proportionality assessment. Hypothetically, LAWS could indeed function in a fully autonomous mode and conduct all the steps of the targeting process, which to some necessarily imply applying moral judgements. A question thus needs to be asked, whether it is possible to program morality.

According to Signorelli, "[M]*orality requires many previous processes usually considered as high-level cognition, starting with decision-making to self-reflection, to be able to detect mistakes on these decisions; sense of confidence, to estimate how correct a decision or action is; mental imagery, to create new probable scenarios of action; empathy, to equilibrate individual and social requirements; understanding of context, to adapt moral decisions to the context, among others.*"[117]

To implement morality in computer systems, the first step is to develop a theory that would explain (both biologically and physically) consciousness in the human brain, dynamics of possible mutual relationships of consciousness, and conscious behaviour. The second step would be to define corresponding mechanisms which can be replicated in machines.[118] In other words, the solution could lie in "*moving away from the traditional computational architectures found in deep neural networks,* […] *and towards a neural network that operates similar to the brain.*"[119] Recently, spiking neural networks are being explored that operate in a manner analogous to a drum to carry out brain-like functions. The information processing of drums is a dynamical reaction from external/internal stimuli more than a formal calculation process.[120] This development could lead to computers operating on a basis more similar to human brain.

However, even if programming conscious machines proved possible, several significant problems still arise. According to empirical evidence from psychology and neuroscience, it is impossible to expect an algorithm to control the process of emergence of consciousness in machines. Each machine, even with the same starting rules, would be different. What is an

[116] See section 2.2.1 below.

[117] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 4.

[118] Ibid, p. 15.

[119] Sharp, C. Cognitive Lethal Autonomous Weapons Systems (CLAWS). Articles of War. 5 November 2021. At: https://lieber.westpoint.edu/cognitive-lethal-autonomous-weapons-systems/ (last accessed 13 April 2022).

[120] Ibid.

even more severe issue, the programmers would not be able to control the process and subsequently control the machine.[121] Furthermore, it is likely that a conscious machine would lose the advantages of actually being a computer: to solve problems with accuracy, speed and obedience (exactly those benefits for which militaries want to deploy LAWS in armed conflicts). Signorelli argues that "[a]*ny system with subjective capabilities is not accurate anymore, because if* [it] *replicate*[s] *high-level cognitions of human, it is also expected that they will replicate the experience of colour or even pain, in a way that it will also interfere with rational and optimal calculations, as well as in humans*."[122]

Additionally, as it is challenging to predict how a conscious machine would operate, it is also possible that the machine would develop a new kind of morality based on non-anthropocentric views and even new possible answers to many moral dilemmas.[123] Indeed, the incorporation of computer systems modelling the human brain could enable LAWS to exercise a human-like discretion in targeting decisions.[124] But to be able to harness their potential and their technological capabilities contributing to more exacting results, programmers must ensure that the conscious machines would not lose the computational and algorithmic advantages. Moreover, the possibility of losing human control over the process of machine cognition learning is more than worrying.

[121] Haladjian, H. H. and Montemayor, C. Artificial consciousness and the consciousness-attention dissociation. *Consciousness and cognition.* 2016(45). 210–225, At: https://pubmed.ncbi.nlm.nih.gov/27656787/ (last accessed 13 April 2022); Signorelli, C. M. Types of cognition and its implications for future high-level cognitive machines. *AAAI Spring Symposium Series* (Berkeley, CA). 2017. At: http://aaai.org/ocs/index.php/SSS/SSS17/paper/view/\penalty-\@M15310 (last accessed 13 April 2022).

[122] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 16.

[123] Ibid, p. 4.

[124] Sharp. Cognitive Lethal Autonomous Weapons Systems (CLAWS), supra note 119.

2. Compliance with IHL rules on the conduct of hostilities

Once one understands the limits of current and foreseeable technology, it becomes clear that the first and foremost reason why we need MHC over autonomous weapons is to ensure their compliance with IHL. While the debate remains deeply polarised as to whether the use of AWS is ethically acceptable and legally sound, widespread consent was expressed that IHL rules apply to the deployment of LAWS.[125] The central issue debated is whether these weapon systems can comply with the fundamental rules on the conduct of hostilities, namely the rules of distinction, proportionality, and precautions in attacks. Each of these rules poses its own challenge to LAWS.

2.1 Principle of distinction

The fundamental principle of distinction obliges the Parties to the conflict to distinguish at all times between (a) the civilian population and combatants; and (b) between civilian objects and military objectives. Accordingly, operations can only be directed against military objectives.[126] Attacks on civilians and civilian objects are prohibited.[127] In addition, in case of doubt over the status of a person or an object, they shall be considered civilian.[128]

Respecting the principle of distinction is problematic for LAWS in their current state of development, as autonomous artificial agents cannot analyse the context in which they operate with the necessary precision to distinguish a legitimate target.[129]

[125] See for example: CCW. Meeting of the High Contracting Parties to the CCW, Final report, Annex III, CCW/MSP/2019/9, 13 December 2019, Guiding principle 1. At: https://undocs.org/Home/Mobile?FinalSymbol=CCW%2FMSP%2F2019%2F9&Language=E&DeviceType=Desktop&LangRequested=False (last accessed 13 April 2022).

[126] Art. 48 of Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflict, adopted on 8 June 1977, entered into force on 7 December 1978 (Additional Protocol I or AP I). According to the ICRC, this rule belongs to customary IHL in both international and non-international armed conflicts. See: ICRC. *Customary International Humanitarian Law.* Volume I: Rules. Cambridge: Cambridge University Press, 2005, Rule 1, p. 3.

[127] Arts 51(2) and 52(1) AP I.

[128] Arts 50(1) and 52(3) AP I.

[129] Sharkey N. Staying in the Loop: Human Supervisory Control of Weapons. In: Bhuta, N., Beck, S., Geiss R., Kress, C. and Liu, Hin Yan. *Autonomous Weapons Systems: Law, Ethics, Policy.* Cambridge: Cambridge University Press, 2016, 23-38; Amoroso, D. and Tamburrini, G. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. *Current Robotics Reports*. 2020. 1(4). 187–94. At: https://doi.org/10.1007/s43154-020-00024-3 (last accessed 13 April 2022).

2. 1. 1. Determining military objectives

The principle of distinction remains, without a doubt, applicable to attacks conducted with the use of LAWS. Therefore, if used in an armed conflict, it is of crucial importance that autonomous weapon systems can distinguish between lawful and unlawful targets. Current object recognition uses sensors to detect an object, compare it to images uploaded in its database, and, with some degree of probability, determine whether the object falls under one of the pre-determined categories of military equipment, such as ammunition, tanks, armoured personnel vehicles, etc.[130] Today, several weapon systems are capable of determining the military nature of specific categories of simple targets, based on quantitative data.[131] However, they are used in narrow, specifically determined roles and operating in relatively static, low clutter environments. Abiding by the rule of distinction in more complex environments would require a qualitative assessment, thus making it far more challenging for autonomous weapon systems to comply with IHL. The objects that current systems are able to recognise will meet the definition of a military objective in virtually any armed conflict (e.g., tanks, fighter jets).

On the other hand, objects which are *a priori* civilian (e.g., hospitals, schools, apartment buildings) may become military objectives if the criteria are met, and it is considerably more difficult to programme weapon systems to recognise these objects correctly. Military objectives by their "*nature, location, purpose or use*" must make an "*effective contribution to military action*", and their capture or destruction, in the circumstances ruling at the time, "*offers a definite military advantage*".[132] The criterion of "purpose" is concerned with the intended future use of an object, while that of "use" is concerned with its present function. Thus, a school or a hotel is a civilian object, but if used to accommodate troops or headquarters staff, they become military objectives.[133] The definition of a military objective is thus context-dependent and must be judged on a case-by-case basis. Determining both an "effective contribution to military action" and a "definite military advantage" requires assessing contextual elements that vary according to multiple factors. As such, this exercise

[130] Boothby, W. How Far Will the Law Allow Unmanned Targeting to Go? In: Saxon, D. *International Humanitarian Law and the Changing Technology of* War. The Netherlands: Martinus Nijhoff Publishers, 2013, p. 55.

[131] Wagner, M. Autonomy in the Battlespace. In: Saxon, D. *International Humanitarian Law and the Changing Technology of* War. The Netherlands: Martinus Nijhoff Publishers, 2013, p. 113.

[132] Art. 52(1) AP I.

[133] ICRC. *Commentary on the Additional Protocols*, supra note 111, para. 2022.

would involve a qualitative or subjective judgment.[134] Moreover, even though a computer system may more easily recognise the "nature" of the object, the "use" or the "purpose" of an object seems to suggest that some form of reasoning is necessary. It seems difficult to imagine that an algorithm would be capable of determining the "*intended use of an object*".

Akerson argues that the definition of military objective is "*expressed in general, subjective terms for precisely the reason that it cannot be articulated with any more precision without reference to the context in which the commander must apply it. The paradigm is thus unsuitable for a computer algorithm for two reasons: it cannot be expressed with precision and its value can only be determined in the context of application*."[135] It thus appears that determining military objectives is an exercise where humans need to work closely with LAWS and update the system regularly on the characteristics of certain objects.

2. 1. 2. Distinguishing combatants

Furthermore, under the rule of distinction, civilians are protected from deliberate attack unless and for such time as they are directly participating in hostilities.[136]

Distinguishing regular members of armed forces would pose particular challenges for the programmer of an autonomous weapon system. Even in the "simplest" armed conflict involving only uniformed combatants, LAWS would need to be capable of differentiating between an armed and uniformed soldier and an armed and uniformed civilian such as a police officer, as law enforcement officials *a priori* do not fall under the category of combatants.

Even more difficulties arise in contemporary armed conflicts, typically non-international, where fighters often do not wear distinctive uniforms or any other distinctive signs. LAWS must accurately distinguish a civilian directly participating in hostilities from one who is not.[137] The notion of "direct participation in hostilities" is notoriously difficult to define. The

[134] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 79.

[135] Akerson, D. The Illegality of Offensive Lethal Autonomy. In: Saxon, D. *International Humanitarian Law and the Changing Technology of* War. The Netherlands: Martinus Nijhoff Publishers, 2013, p. 79.

[136] Art. 3 common to Geneva Conventions of 12 August 1949, adopted on 12 August 1949, entered into force on 21 October 1950; Arts 51(2) and (3) AP I; Art. 13, Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), adopted on 8 June 1977, entered into force on 7 December 1978 (Additional Protocol II or AP II).

[137] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 80.

ICRC developed guidance that proposes three cumulative criteria: "*1) the act must be* ***likely to adversely affect*** *the military operations or military capacity of a party to an armed conflict or, alternatively, to inflict death, injury or destruction on persons or objects protected against direct attack; 2) there must be a* ***direct causal link*** *between the act and the harm likely to result either from that act or from a coordinated military operation of which that act constitutes an integral part; and 3) the act must be specifically designed to directly cause the required* ***threshold of harm*** *in support of a party to the conflict and to the detriment of another.*"[138] Each criterion necessarily requires a qualitative analysis of the situation at hand, and all are often challenging even for humans to apply. Such a task appears impossible to translate into a computer programme, at least with any foreseeable technology.

Moreover, LAWS would need to be capable of distinguishing active combatants from persons *hors de combat* (persons in the power of an adverse Party, clearly expressing an intention to surrender, or rendered unconscious or is otherwise incapacitated by wounds or sickness).[139] The ICRC Commentary indicates that a defining feature of persons *hors de combat* is the fact that they are "defenceless", whether or not they have laid down arms.[140] Therefore, the weapon system needs to detect and recognise a person's willingness to surrender. Unless there is a universally recognised sign or movement indicating surrender, such an assessment relies heavily on information reasonably available to commanders at the time they take their action and their ability to deduce a person's intention from their behaviour. Cognitive tasks like that have proven the most difficult to translate into machines. The first significant problem with the supervised machine learning approach would be providing the correct labelling.[141] While providing supervised labelling of objects or animals appearing in an image is a relatively straightforward task, the same cannot be done for labelling emotions in faces, voices or behaviours and assigning them a dimensional quantification.[142] Labelling of emotions is significantly affected by biases introduced by supervising experts and because it does not

[138] See: ICRC. Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law. Geneva: 2009. At: http://www.icrc.org/eng/assets/files/other/irrc-872-reports-documents.pdf (last accessed 13 April 2022).

[139] Art. 41 AP I.

[140] ICRC. *Commentary on the Additional Protocols*, supra note 111, para. 1630.

[141] Franzoni, V., Milani, A., Nardi, D. and Vallverdú, J. Emotional machines: The next revolution. *Web Intelligence*. 2019. 17. 1–7, p. 2. At: https://content.iospress.com/articles/web-intelligence/web190395 (last accessed 13 April 2022).

[142] Franzoni, V., Milani, A., Pallottelli, S., Leung C.H.C. and Li Y. Context-based image semantic similarity. *12th International Conference on Fuzzy Systems and Knowledge Discovery.* FSKD 2015. 2016. At: https://ieeexplore.ieee.org/document/7382127 (last accessed 14 April 2022).

adequately capture the contribution to emotions given by the context.[143] According to Signorelli, "*current implementations of emotions in machines are based on a logical, computable and deterministic approaches, leaving out essential characteristics of emotions such as that emotions interfere with rational processes and optimal decisions*."[144]

Regarding machines' ability to interpret the intent to surrender, Boothby argues that considering how difficult distinguishing persons representing lawful targets is for humans, it appears unlikely bordering on impossible for autonomous weapons.[145] In his opinion, to employ a weapon system that renders it impossible to comply with the rule protecting persons *hors de combat* would not be lawful unless the rule is not relevant for the mission planned.[146] This scenario is unlikely in general but one can imagine operations where no humans could be targeted, such as in deep seas or space. Nevertheless, even if the rule of distinction is not relevant in a particular case, attention still must be paid to the rule of proportionality.

2.2 Rule of proportionality

According to the rule of proportionality applicable to the conduct of hostilities, incidental civilian casualties and damages can be lawful if they are not excessive in relation to the concrete and direct military advantage anticipated, provided other rules are respected.[147] This rule is considered customary law in all types of armed conflict.[148]

2. 2. 1. Excessive civilian casualties

Proportionality arguably belongs to the most complex rules to interpret and apply under IHL, as it necessarily requires a case-by-case assessment in often rapidly changing circumstances.[149] Moreover, assessing its two core elements (excessive civilian casualties and military advantage anticipated) is always context-specific and based upon qualitative judgement. For example, a different number of incidental civilian casualties may be

[143] Franzoni V. and Poggioni V. Emotional book classification from book blurbs. *Proceedings – 2017 IEEE/WIC/ACM International Conference on Web Intelligence.* WI 2017. At: https://dl.acm.org/doi/10.1145/3106426.3109422 (last accessed 14 April 2022).

[144] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 2.

[145] Boothby. How Far Will the Law Allow Unmanned Targeting to Go?, supra note 130, p. 59.

[146] Ibid, p. 60.

[147] Queguiner, J. Precautions under the Law Governing the Conduct of Hostilities. *International Review of the Red Cross.* 2006. 88(864), p. 794. At: https://international-review.icrc.org/articles/precautions-under-law-governing-conduct-hostilities (last accessed 14 April 2022).

[148] ICRC. *Customary International Humanitarian Law*, supra note 126, Rule 14.

[149] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 82.

considered proportionate in an attack on an abandoned airfield and in an attack on an airport from which enemy aircrafts are about to take off. The International Criminal Tribunal for the Former Yugoslavia (ICTY) has held that "*in determining whether an attack was proportionate it is necessary to examine whether a* ***reasonably well-informed person*** *in the circumstances of the actual perpetrator, making* ***reasonable use of the information available*** *to him or her, could have expected excessive civilian casualties to result from the attack.*"[150] The formulation tries to bring an objective element into what appears to rely heavily on subjective judgement. The question is, how can this standard of reasonability be translated into an algorithm. In the case of LAWS, the system would have to be constantly updated on all information that could be relevant. Moreover, the programme would have to be designed to enable it to connect all pieces of information, even if seemingly unrelated.

In some parts of the process of assessing excessive civilian casualties, a computer programme may play a role. Schmitt points to the "collateral damage estimate methodology" or CDEM used by the US military in planning attacks to assess factors such as a weapon's precision, its blast effect, attack tactics, the likelihood of civilian presence, and the composition of buildings. The CDEM itself "*does not resolve whether a particular attack complies with the rule of proportionality", rather it is described as "a policy-related instrument used to determine the level of command at which an attack causing collateral damage must be authorized*."[151] A similar programme could certainly be used in LAWS to calculate possible civilian casualties. Sassòli goes a step further and argues that it might be possible to identify indicators and criteria to evaluate proportionality and make the implied judgment slightly more objective.[152] The problem lies in attributing values to objects and persons to analyse what is excessive. In Thurnher's opinion, "*it is conceivable that AWS could lawfully operate upon a framework of pre-programmed values. The military operator setting these values would, in essence, pre-determine what constitutes excessive collateral damage for a particular target. (...) these values would invariably need to be set at extremely conservative*

[150] ICTY, Prosecutor v. Stanislav Galić, Case No. IT-98-29-T, Judgment, Trial Chamber (5 December 2003), para. 58 (emphasis added).

[151] Schmitt, M. Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics. *Harvard National Security Journal: Features Online*. 2013, p. 19. At: https://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/ (last accessed 14 April 2022).

[152] Sassòli, M. Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified. *International Law Studies / Naval War College*. 2014(90). 308-340, p. 331. At: https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1017&context=ils (last accessed 14 April 2022).

ends to comply with the rule [of proportionality]."[153] However, as argued above, particular objects may present a different "concrete and direct military advantage anticipated" in different situations. Again, the same problem arises that it is not conceivable that a programme could be trained for every possible situation. Along the lines of Thurner's conclusion, LAWS could be programmed to put more emphasis on civilian protection, which would make them compliant with IHL. Nevertheless, it would also diminish their usefulness for the military. If a weapon system is highly conservative, it may have unwanted consequences for military operations.

2. 2. 2. Military advantage and its relevance in the targeting process

Given the complexity and fluidity of the modern battlespace, it would also be complicated to programme an autonomous weapon system to assess the military advantage anticipated. As Sassòli points out, the military advantage "*constantly changes according to the plans of the commander and the development of military operations on both sides. Except where no, or clearly negligible, effects upon civilians can be anticipated, a machine, even if perfectly programmed, could, therefore, not be left to apply the proportionality principle unless constantly updated about military operations and plans*."[154]

Military operations and plans enter the targeting process on different levels of command and in different targeting phases. NATO publication Allied Joint Doctrine for Joint Targeting, AJP-3.9 defines the targeting cycle as the process that "*links strategic-level direction and guidance with tactical targeting activities through the operational-level targeting cycle in a focused and systemic manner to create specific physical and psychological effects to reach military objectives and the desired end state*."[155] This concerns mainly preplanned (or deliberate) targeting, as opposed to dynamic. According to Ekelhof, "[t]*he difference between dynamic and deliberate targeting is, simply put, time. When targets are identified and developed in time, they can be included in the deliberate targeting cycle and actions against them can be scheduled. Dynamic targeting consists largely of the same steps but is more responsive than deliberate targeting since the process is used to prosecute targets that are either unexpected or known to exist in the area of operations, but were not yet detected or*

[153] Thurnher, J. Examining Autonomous Weapon Systems form a Law of Armed Conflict Perspective. In: Nasu, H. and McLaughlin, R. *New Technologies and the Law of Armed Conflict*. The Netherlands: T.M.C Asser Press, 2014, p. 222.

[154] Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 332.

[155] NATO. Allied Joint Doctrine for Joint Targeting. Brussels: NATO Standardization Office, 2016, p. 1-1.

selected for action in sufficient time to be included in the deliberate process." Many of the points addressed in relation to MHC are valid for both types of targeting. However, in dynamic targeting, the role of the pilot in ensuring legal compliance may be very different. For the sake of clarity and conciseness, the following analysis will mainly deal with preplanned targeting. It will demonstrate how determining military advantage anticipated from an attack is linked to multiple stages of the targeting process on the example of the NATO Allied Joint Doctrine for Joint Targeting.

2. 2. 2. 1. Phase 0: Political and strategic objectives and guidance

Prior to the beginning of the whole targeting process, the North Atlantic Council will provide the Military Committee (the senior military authority in NATO) with "*political guidance, overarching military objectives, and the desired end state for a campaign, including any constraints and restraints it wishes to impose.*"[156] This guidance is passed down to the joint force commander (JFC), who is responsible for executing the campaign.[157]

2. 2. 2. 2. Phase 1: Implementing the objectives and guidance

The impact of the political and strategic objectives and guidance will show in the first phase of targeting. The JFC must identify clearly what is to be accomplished and under what circumstances. "*Once the military campaign objectives are defined, the first activity of the joint targeting process is to take these objectives, guidance (including restrictions with regard to collateral damage), and intent and further translate them into a number of discrete operational tasks.*"[158] We see both elements of the proportionality analysis present already at the top of the chain of the targeting process. The senior military authorities will formulate overarching objectives, which directly influence the military advantage to be achieved. This guidance can also formulate restrictions on collateral damage (i.e., incidental civilian casualties and civilian damage that may be considered disproportionate). This all happens at a very abstract and general level. Should LAWS be fully autonomous from the beginning of the targeting process, their computer programme would need to be updated on and informed about this guidance. More importantly, it would need to take this abstract and generally

[156] Ekelhof. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting, supra note 33, pp. 66-67.

[157] NATO. Allied Joint Doctrine for Joint Targeting, supra note 155, p. 3-1.

[158] Ekelhof. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting, supra note 33, p. 66.

formulated objectives and guidance into account when engaging targets and conducting the proportionality analysis.

2. 2. 2. 3. Phase 2: Target development

The second phase covers a number of activities around target development, which, simply put, has five functions: target analysis, target vetting, target validation, target nomination, and target prioritisation. Eligible targets are identified that have the potential to achieve the JFC's objectives, and the principal output is a joint prioritised target list.[159] During target analysis, the most relevant targets linked to strategic and operational objectives are identified together.[160] Commanders look "*beyond the characteristics of a single target; a target's real importance may lie in its relationship to other targets within a particular operational system*."[161] Here again, the whole context of the particular situation plays an important role. LAWS would have to be able to conduct their target analysis in a way that takes into account all realities of a given situation.

Target validation as another sub-step ensures that the selected targets comply with the JFC's objectives, do not violate international law rules, and the analysis used to develop the targets is accurate and credible. Targets also are coordinated and deconflicted with other operations.[162] A computer system would likely be able to have a database of targets selected for different operations and could verify whether there is not a conflict present. However, this process again requires updating the weapon system on all information possibly relevant not only to the particular operation but also any other conducted by the same party to the armed conflict.

During target prioritisation, targets on the joint target list are cleared against the rules of engagement, NATO caveats, and relevant international law (such as the principle of distinction).[163] To demonstrate the complex procedure: "*Targets are developed and reviewed multiple times by many different staff and different commands in the Joint Targeting Working Group. Once fully developed, these targets are presented to the Joint Targeting Coordination Board, which typically consists of functional advisers (e.g., legal, political, information-*

[159] Ibid, p. 67.
[160] NATO. Allied Joint Doctrine for Joint Targeting, supra note 155, p. 2-3.
[161] Ekelhof. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting, supra note 33, p. 70.
[162] Ibid, p. 72.
[163] NATO. Allied Joint Doctrine for Joint Targeting, supra note 155, p. 4-7.

operations, and electronic-warfare advisers, as required), representatives of the different components (land, maritime, air, and special operations), national representatives, and the commander. Different military representatives (e.g., the chief targeteer, legal adviser, director of operations) will provide the commander with the relevant information, In the end, the commander will decide whether to approve the presented targets and place them on the joint prioritized target list, or disapprove or suspend them (e.g., owing to a lack of intelligence)."[164] A weapon system with full autonomy in the whole targeting process would have to gather the same information and conduct the same in-depth analysis that requires many experts.

2. 2. 2. 4. Phase 3: Capabilities analysis

In the third phase of targeting, the capabilities analysis takes place. It is the process of analysing the prioritised targets and matching the most appropriate capabilities, lethal and nonlethal, to generate the desired physical or psychological effects.[165] The output of its first element, weaponeering, is a recommendation of the quantity, type, and mix of lethal and nonlethal weapons needed to achieve the desired effects while avoiding unacceptable collateral damage.[166] This is where the obligation to take all feasible precautions in the choice of means and methods of warfare to spare civilian population plays out.[167] Precautions as such will be discussed in more detail,[168] however, it is important to have a clear picture of when they enter the targeting process. The second element of the capabilities analysis is called a "collateral damage estimation", which estimates the unintentional physical damage to civilians, civilian objects, or the environment resulting from an attack.[169] It is the crucial step where the main proportionality assessment takes place.

It is clear that targeting is a very elaborated and complex process. Autonomy in weapon systems may appear at various points of the process and may have a direct link to the proportionality assessment. As Ekelhof argues, "*autonomous technologies used for target development have an effect on which specific targets end up on the approved target list by*

[164] Ekelhof. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting, supra note 33, p. 73.

[165] NATO. Allied Joint Doctrine for Joint Targeting, supra note 155, p. 2-4.

[166] NATO. Allied Joint Doctrine for Joint Targeting, supra note 155, p. 1-10.

[167] Art. 57(2)(a)(ii) AP I; Art. 7 Second Protocol to The Hague Convention of 1954 for the Protection of Cultural Property in the Event of Armed Conflict, adopted on 26 March 1999, entered into force 9 March 2004; ICRC. *Customary International Humanitarian Law*, supra note 126, Rule 17.

[168] See section 2.3 below.

[169] NATO. Allied Joint Doctrine for Joint Targeting, supra note 155, p. 1-10.

determining what data humans see and how they should conceive the battlefield. The fact that these technologies are not weaponized is irrelevant, as their tasks are potentially even more critical for targeting than those of their weaponized cousins."[170] Focusing only on the assessment of collateral damage and military advantage anticipated, it has been argued that even though the core analysis takes place during the third phase of targeting, the previous phases directly influence both elements. It seems difficult to believe that the current of foreseeable technology would enable LAWS to conduct the whole complex process of targeting autonomously, absorbing all the information possibly relevant and translate general guidance and objectives into specific and context-dependent proportionality analysis.

Another option would be to imagine weapon systems which have autonomy only in the phase of capabilities abilities analysis. Even if a computer system was able to predict likely civilian casualties and attribute value to them, their "proportionality" depends on the concrete and direct military advantage anticipated, which is influenced by a vast number of factors, including the results of all previous phases of targeting. Many argue that the application of the proportionality rule involves a subjective determination.[171] Sassòli, on the other hand, argues that it would be desirable not only for the development of LAWS but also for human operators to quantify how the risk of losing one civilian life compares with the potential of gaining a specific military advantage and what relation between the risk and the advantage would be excessive, together with indicators of the elements that should (not) be taken into account.[172] It might serve as an incentive for militaries to pursue more transparency in their proportionality evaluation, should they want to deploy LAWS with autonomy in functions contributing to or executing the proportionality analysis.

2.3 Precautionary measures

Third, in the conduct of hostilities, IHL requires the parties to armed conflicts to spare the civilians and civilian objects. Apart from the principle of distinction and the rule of proportionality, there is an obligation to take precautions to achieve that aim.[173]

[170] Ekelhof. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting, supra note 33, p. 87.

[171] Human Rights Watch. *Losing Humanity: The Case Against Killer Robots*, supra note 3, p. 32; Heyns. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, supra note 16, para. 70.

[172] Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 334.

[173] Art. 57 AP I.

2. 3. 1. Feasibility of precautions

The bottom-line is that those who plan or decide upon an attack shall do everything feasible to verify that the objectives to be attacked are lawful target[174] and choose the means and methods of attack to minimise civilian casualties.[175] This rule undeniably applies to attacks conducted with the use of LAWS.

AP I does not define "feasible precautions." The word feasible can be defined as "*possible to do easily or conveniently.*"[176] The standard is thus not perfection, rather what is practicable. However, determining whether a certain precautionary measure is feasible has to be measured against the human standard and not against the possibility for a machine to take a particular action.[177] If another method of attack (than using LAWS) would permit certain precautions, then such precautions would be considered feasible and thus required.[178]

On the other hand, there can be situations where precautions unavailable to humans could be possible for LAWS because the human life of the pilot or weapons operator is not at risk.[179] Both of these options need to be taken into account when considering compliance of deploying LAWS with the obligation to take feasible precautions.

2. 3. 2. Verifying the nature of the objective

Some argue that the obligation to verify the nature of the objective could be complied with in a situation where the target to be engaged is susceptible to "mechanical target recognition". Pre-determined categories of military equipment (tanks, armoured vehicles, missiles) could be detected and recognised by sensors and verified prior to the attack.[180]

However, with respect to attacks, the precautions shall be taken by "*those who plan or decide upon an attack*." Only humans are addressees of IHL rules. Here, the category of humans obliged to comply with the obligation to take feasible precautions in attack is narrowed down. We can see two possible approaches to this rule regarding LAWS. First, the weapon system

[174] Art. 57(2)(a)(i) AP I
[175] Art. 57(2)(a)(ii) AP I
[176] Oxford English Dictionary. "Feasible". At: https://www.lexico.com/definition/feasible (last accessed 13 April 2022).
[177] Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 336.
[178] Boothby. How Far Will the Law Allow Unmanned Targeting to Go?, supra note 130, p. 61.
[179] Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 336.
[180] Boothby. How Far Will the Law Allow Unmanned Targeting to Go?, supra note 130, p. 55.

will be programmed to verify targets before engaging them. However, as explained above,[181] the technological limitations would constrain the operations of such LAWS only to predictable, un-cluttered environments and to clearly recognisable targets. In this case, the human operator could discharge his obligation to verify the nature of the objective by assessing the lawfulness of a target in the programming stage. The reliability of such assessment would be doubtful where LAWS would be deployed in a dynamic environment.[182]

The other option is to retain a human operator in control of verifying the nature of the objective to be targeted. Suppose an autonomous weapon system would pre-select targets and require approval (verification) from its operator. In that case, this type of human-machine interface could fall under the human-in-the-loop scheme. If the operator had mere veto powers, this would belong in the human-on-the-loop category of LAWS.

2. 3. 3. Choosing means and methods

Moreover, IHL obliges Parties to the conflict to choose the methods and means of warfare likely to cause the least danger to civilian lives and civilian objects.[183] As explained above, the choice of means and methods of warfare forms a part of the targeting process, namely the capabilities assessment. If we consider LAWS with autonomy in their targeting function, these thus have to be able to comply with this requirement.

This obligation could apply to autonomous weapon systems in two distinct ways. Firstly, in terms of the decision of a commander to deploy LAWS; and secondly, regarding the specific means the autonomous weapon system selects when it engages a target.[184]

Concerning deployment of LAWS, autonomous weapon systems can lawfully be used to achieve military objectives if other available systems would cause more or comparable collateral damage.[185] On the other hand, if it is clear or predictable that the deployment of LAWS would cause fewer incidental civilian casualties and/or less incidental damage to civilian objects compared to the use of conventional weapons, it is certainly preferable (and in

[181] See Chapter III, Section 1 above.

[182] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 85.

[183] Article 57(2)(a)(ii) AP I; the ICRC considers the obligation as a customary rule of international law, see ICRC. *Customary International Humanitarian Law*, supra note 126, Rule 17.

[184] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 86.

[185] See for example: Schmitt. Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics, supra note 151, p. 24.

line with Article 57 (1) AP I) to avoid unnecessary incidental effects. Some claim that the rule on precautions may therefore require a commander to consider using the autonomous weapon, subject to other considerations, such as the need to reserve their use for other militarily more important tasks or tasks involving higher risks for civilians.[186] However, this should not be interpreted as imposing a duty on States to acquire LAWS.[187] IHL does not oblige States to invest in modern weapons. Though if a State obtains them, it gains military advantages (such as the possibility of a more precise attack or an attack with less casualties), and IHL rules on the conduct of hostilities might implicitly require that these weapons are used.

Concerning the second aspect of this rule, the specific means the autonomous weapon system selects when it engages a target, there are particular challenges in programming LAWS to be capable of respecting this rule. The obligation to minimise the incidental negative effects on civilians covers both the choice of a particular weapon to be used but also imposes restrictions on the timing, location, or even angle of an attack.[188] Again, the point must be made that it is human operators of LAWS who indeed have to comply with this rule. As Sassòli argues, human planners may be temporally and geographically removed from a particular attack as long as they ensure that LAWS comply with the pre-defined parameters and have the necessary information to apply them.[189]

Taking the issue of timing and location as an example, different scenarios may arise regarding the deployment of LAWS. On one side of the spectrum, the weapon system may be autonomously operating for an undefined period of time and capable of moving across a vast area. In this scenario, it is simply not viable for human planners to reliably predict all circumstances that may arise. This renders it impossible to define the parameters of each attack beforehand. Therefore, in order to comply with the obligation to take precautions in the choice of methods of warfare, a higher level of human control must be retained over the operation of the weapon system.

[186] Davison, Weizmann, and Robinson. Background Paper by the International Committee of the Red Cross, supra note 36, p. 86, referring to Kellenberger, J. International Humanitarian Law and New Weapon Technologies. ICRC, Keynote address at 34th Round Table on Current Issues of International Humanitarian Law, San Remo. 8-10 September 2011. At: https://international-review.icrc.org/articles/international-humanitarian-law-and-new-weapon-technologies-34th-round-table-current-issues (last accessed 15 April 2022); Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 320.

[187] Sassòli, M. *International Humanitarian Law Rules, Controversies, and Solutions to Problems Arising in Warfare*. Cheltenham: Edward Elgar Publishing, 2019, p. 366, para. 8.331.

[188] Queguiner. Precautions under the Law Governing the Conduct of Hostilities, supra note 147.

[189] Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 336.

In another scenario, LAWS may be deployed in a simple environment with imposed time restrictions. The environment presents few challenges, and human planners can predict a high percentage of possible scenarios. To ensure that, LAWS would be deployed only for a limited period. In such a case, it seems imaginable that an autonomous weapon system could comply with the rule on precautions, in a way ensuring legal compliance by design. Even in this scenario, human control over the choice of means and methods of warfare is retained, as the relevant decision was responsibly taken before the deployment of the weapon system.

To sum up, the analysis of the rules of distinction, proportionality and precautions shows that compliance with IHL rules on targeting requires a certain level of human control, which may vary in different circumstances, but its existence is still essential.

3. Consequences for people's lives and livelihoods

Another reason why meaningful human control is needed has been suggested by the ICRC, which believes that it is "*essential to preserve human control over tasks and human judgement in decisions that may have serious consequences for people's lives in armed conflict, especially where they pose risks to life, and where the tasks or decisions are governed by specific rules of international humanitarian law.*"[190] It goes on to argue that LAWS (possibly employing AI and machine-learning) are being developed to perform tasks that would ordinarily be carried out by humans and that there is an inherent tension between this pursuit and the centrality of the human being in armed conflict.

Anderson and Waxman argue that the development of automated (if not autonomous) systems is inevitable, in part because it is not merely a feature of weapons technology but of technology generally.[191] The role of humans in an armed conflict has constantly been changing and is becoming increasingly remote. However, even if remote, it remains central to the aspects of targeting and other critical functions. The challenge is to adopt a regulation that sufficiently reflects the emerging technologies and allows the military to benefit from their use while retaining respect for not only legal but also ethical and moral requirements. That is one of the reasons why the emphasis in the debate around LAWS has been put on so-called "critical functions". In the ICRC's view, human control and judgment are crucial for tasks and

[190] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 7.

[191] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 4.

decisions that can lead to injury or loss of life or damage to civilian infrastructure.[192] Interestingly, it does not focus solely on the loss of lives of civilians but injury or loss of life in general. This aligns with the current prevailing use of autonomous weapon systems, which are limited in the types of targets to primarily vehicles or objects rather than personnel. However, some existing anti-personnel weapon systems have autonomous modes, such as the so-called "sentry weapons".[193]

On the other hand, flight functions do not raise a similar level of concerns. Some functions, such as "autopilot" in military and civilian aircraft, have been autonomous for many years.[194] Today's unmanned aerial vehicles are not yet fully autonomous; they are, however, increasingly automated in their flight functions, such as self-landing capabilities. A single pilot can operate several unmanned aircraft simultaneously, increasing efficiency considerably.[195] Given that speed is particularly important, the future design will emphasise automating as many of these functions as possible.[196] However, all aircraft safety-critical software in operation today is entirely predictable, i.e., given a particular set of inputs, it will always produce the same output.[197] Given the utmost significance and impact of decisions on the use of force (determining who and what is targeted and attacked in armed conflict), it is imperative that meaningful human control is retained. As already argued above, the rules of international humanitarian law are addressed to humans, who have to comply with and implement the law. It is humans who will be held accountable for violations. As the ICRC points out, "[s]*ince humans are the legal – and moral – agents in armed conflict, the technologies and tools they use to conduct warfare must be designed and used in a way that enables combatants to fulfil their legal and ethical obligations and responsibilities*."[198]

[192] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 7.

[193] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*, supra note 34, pp. 6-7.

[194] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*, supra note 34, p. 5.

[195] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 4.

[196] Ibid, p. 5, referring to Osinga, F. P. B. *Science, Strategy and War: The Strategic Theory of John* Boyd. London and New York: Routledge, 2006.

[197] Ansell. Research and Development of Autonomous 'Decision Making' Systems, supra note 79, p. 40.

[198] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 7.

4. Ethical reasons

In every debate on LAWS, an argument regarding ethical considerations is made. When considering human dignity, the focus is on the process through which the decisions to injure or kill are made. Some believe that human dignity is violated if a machine makes a potentially lethal decision.[199] Others claim that to preserve a measure of humanity in armed conflict, we cannot delegate decisions to kill humans to a machine. For Asaro, "*the very nature of IHL (...) presupposes that combatants will be human agents"* in the same way that judges, prosecutors, defenders, witnesses and juries all assess *"the match between an abstract set of rules and any given concrete situation*."[200] While the latter is true in the sense that sometimes abstract IHL rules must be applied to a particular situation, there is an essential distinction between judges and combatants. As Sassòli points out, "[t]*o target a person is* […] *definitely not to render justice or more precisely, it is not a determination that the person deserves the death penalty, but involves exclusively a categorization of the person (as a combatant) or their conduct (direct participation in hostilities) without any determination of fault or culpability.*"[201] From an IHL perspective, the question thus is not how ethical it is for LAWS to render judgements but whether it is ethical or moral to allow machines to make targeting decisions that may cause loss of human life.

Although ethical considerations are not specified in the law of armed conflict and cannot be interchanged with legal rules, they often serve as a basis for the latter. They can influence the interpretation of the law. For example, it was argued that moral judgment underlies the determination of whether a weapon is of a nature to cause superfluous injury.[202] According to Heyns, it is an underlying assumption of most of the laws that humans should make decisions with potentially lethal consequences.[203] In his opinion, it is implied by IHL treaties, the rules

[199] Taddeo and Blanchard. *A Comparative Analysis of the Definitions of Autonomous Weapons,* supra note 6, referring to a number of authors.

[200] Asaro, P. On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanisation of Lethal Decision-Making. *International Review of the Red Cross*. 2013. 94(886), p. 700. At: https://international-review.icrc.org/articles/banning-autonomous-weapon-systems-human-rights-automation-and-dehumanization-lethal (last accessed 14 April 2022).

[201] Sassòli. *International Humanitarian Law Rules, Controversies, and Solutions to Problems Arising in Warfare*, supra note 187, p. 520, para. 10.78.

[202] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*, supra note 34, p. 16.

[203] Heyns. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, supra note 16, para. 89.

of which assume the conduct of human soldiers or commanders rather than machines.[204] However, it is unlikely that the possibility of a non-human decision making in armed conflicts was considered at the time of the creation of the traditional IHL rules.

Waxman and Anderson have raised the issue of defining "*the tipping point into impermissible autonomy, given that the automation of weapons functions is likely to occur in incremental steps.*"[205] Since that argument was made, discussions in the CCW Group of Experts have clarified that the concerns relate to critical functions of LAWS and the design of a human-machine interface that impacts the actual involvement of humans in the decision-making process.

If one concludes that it would be unethical for weapon systems to make potentially lethal decisions autonomously, the question arises about the sufficient level of human supervision. Asaro claims that "*including a human in the lethal decision process is a necessary, but not a sufficient requirement. A legitimate lethal decision process must also meet requirements that the human decision-maker involved in verifying legitimate targets and initiating lethal force against them be allowed sufficient time to be deliberative, be suitably trained and well informed, and be held accountable and responsible*." In other words, the process should allow meaningful human control. While the elements and requirements of the "meaningfulness" of human control will be discussed below,[206] it presents a plausible solution to the ethical concerns expressed. It would preserve a certain measure of humanity in warfare. However, the ethical argument should not be stretched to argue in favour of a complete ban on automation or autonomy in weapon systems. Recognising the incremental evolution of these technologies is key to addressing the ethical dilemmas associated with their inevitability.[207] Waxman and Anderson compare the debates over LAWS to those that arose with respect to technologies that emerged with the industrial era, such as submarines and military aviation. A core objection was that of "remoteness" of humans from the battlefield, that it is unethical to attack from a safe distance. However, weapons superiority is lawful and assumed as part of

[204] Ibid, citing Hague Convention II with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land, adopted on 29 July 1899, entered into force on 4 September 1900; Hague Convention IV; Art.1(2) AP I.

[205] Waxman, M. and Anderson, K. Law and Ethics for Robot Soldiers. *Policy Review*. 2012 (176), p. 7. At: https://scholarship.law.columbia.edu/faculty_scholarship/1742/ (last accessed 14 April 2022).

[206] See Chapter VI section 3 below.

[207] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 3.

military necessity.[208] One also should not forget that new weapons may bring more precision in attack and reduce incidental civilian casualties.

5. MHC in connection to responsibility

The last but not least reason in favour of the need for meaningful human control is connected to the framework of individual criminal responsibility. Violations of IHL primarily give rise to the responsibility of the parties to the conflict, usually States and non-state armed groups. However, certain unlawful conduct in warfare may also trigger the international criminal responsibility of an individual, under a different legal regime. These two forms of responsibility coexist and can apply in parallel to the same conduct.[209] A question needs to be raised, whether and how these traditional modes of responsibility address possible violations committed through the acts of LAWS, and what is the role of meaningful human control in that context.

The problem of human accountability for acts carried out by LAWS appears in almost every debate over their compatibility with international law. The majority opinion is that MHC requires structures of accountability.[210] Amoroso and Tamburrini suggest a threefold role for human control: a fail-safe actor, an accountability attractor, and a moral agency enactor.[211] They consider that "*in order to avoid accountability gaps, human control is required to function as accountability attractor, i.e., to secure the legal conditions for responsibility ascription in case a weapon follows a course of action that is in breach of international law.*"[212] The question is, how to secure the legal conditions for responsibility ascription in cases where a weapon follows a course of action that would otherwise give rise to individual

[208] Ibid, p. 8.

[209] Gaeta, P. and Jain, A. G. Individualisation of IHL rules through criminal responsibility for war crimes and some (un)intended consequences, p. 1. In: Akande, D. and Welsh, J. *The Individualisation of War*. Oxford: Oxford University Press, 2021. At: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3853333 (last accessed 14 April 2022)

[210] Article 36. Key elements of meaningful human control. Paper presented at Technical Report Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, Switzerland, April 11–15 2016; Geiß, R. and Lahmann, H. Autonomous weapons systems: A paradigm shift for the law of armed conflict? In: Ohlin, J. D. *Research Handbook on Remote Warfare*. Cheltenham: Edward Elgar Publishing, 2017, 371–404.; Roff, H. M. Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons. Paper presented at Technical Report Briefing Paper for the Delegates at the Review Conference on the Convention on Certain Conventional Weapons. Geneva, Switzerland: December 12–16, 2016. At: https://article36.org/wp-content/uploads/2016/12/Control-or-Judgment_-Understanding-the-Scope.pdf (last accessed 13 April 2022).

[211] Amoroso and Tamburrini. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues, supra note 129, p. 189.

[212] Ibid.

responsibility? If it is legally permissible to use lethal autonomous weapons, how should these weapons be integrated into the existing command-and-control structure so that responsibility remains associated with specific human actors?[213]

5.1 The "responsibility gap"

Importantly, State responsibility of the relevant party to the conflict may arise for the violation of any rule contained in the Geneva Conventions. By contrast, the Geneva Conventions and Additional Protocol I identify a limited set of violations, so-called "grave breaches", which are particularly serious violations that give rise to specific obligations of repression for States. Grave breaches must be prosecuted by States on the basis of the principle of universal jurisdiction.[214] Together with other serious violations of IHL (established by customary international law and by international criminal law treaties), grave breaches constitute war crimes.[215] Criminal responsibility of individuals is therein expressly provided for only with respect to a specific set of unlawful behaviours. To put it differently, under relevant IHL treaties, State responsibility is the primary way of addressing violations, while individual criminal responsibility solely for grave breaches supplements the former.[216]

Waxman and Anderson indeed argue that "[e]*xcessive devotion to individual criminal liability as the presumptive mechanism of accountability risks blocking the development of machine systems that might, if successful, reduce actual harms to soldiers as well as to civilians on or near the battlefield. Effective adherence to the law of armed conflict traditionally has been through mechanisms of state (or armed party) responsibility*."[217] While one may agree that State responsibility plays its part in promoting compliance with IHL (especially when a violation of IHL does not constitute a grave breach for which an individual could be held responsible anyway), it would be wrong to conclude that this mechanism alone is sufficient. Arguably, collective responsibility is typical of rudimentary legal systems, whereas individual criminal responsibility is seen as contributing to the increasing sophistication of the

[213] Russell, S., Dewey, D. and Tegmark, M. Research Priorities for Robust and Beneficial Artificial Intelligence. *AI Magazine.* 2015. 36(4), 105-114, p. 107. At: https://doi.org/10.1609/aimag.v36i4.2577 (last accessed 14 April 2022).

[214] Art. 49 GCI; Art. 50 GCII; Art. 129 GCIII; Art. 146 GCIV; Art. 85(1) AP I.

[215] ICRC. *Customary International Humanitarian Law*, supra note 126, Rule 156.

[216] Gaeta and Jain. Individualisation of IHL rules through criminal responsibility for war crimes and some (un)intended consequences, supra note 209, p. 2.

[217] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 17.

international legal systems in line with the developments in national legal orders.[218] Recently, the emergence of the so-called "fight against impunity" paradigm has emphasised the importance of criminal repression of violations of IHL, which has been considered more suitable than collective forms of responsibility.[219]

In the LAWS debate, most authors agree that retaining human responsibility is a *conditio sine qua non* for the lawful deployment of autonomous weapon systems.[220] If finding a person responsible for grave breaches is not a practical possibility, then it is feared that there could well be a "responsibility gap" that would enable impunity for the use of autonomous weapon systems.[221] Whilst a possible responsibility gap is problematic in all the categories of use of AI within the defence and security domain, the gap would be particularly worrying when considering the adversarial and kinetic uses of AI (typically conduct of hostilities), given the high stakes involved.[222] Some claim that the two traditional modes of responsibility (individual criminal responsibility and State responsibility) are not suitable for dealing with LAWS, since they are situated somewhere between weapons and combatants.[223] While it is true that autonomy in the weapon systems presents particular challenges to the attribution of responsibility, these can be solved through adapting the law as it is, instead of creating a new mode of responsibility. As Sassòli argues, weapons and humans are not situated at a sliding scale but on different levels as objects and subjects of legal rules.[224]

Even in the (admittedly unlikely) scenario where LAWS would act upon legal rules and be unable to divert from them, it would never be the weapon that would be responsible for any possible violation of IHL, or indeed any other legal rule. The machine would make the decision to engage a specific target, but there would still be a human who has decided to

[218] Sassòli, M. Humanitarian Law and International Criminal Law. In Cassese, A. *The Oxford Companion to International Criminal Justice.* Oxford: Oxford University Press, 2009, p. 113.

[219] Ibid.

[220] Heyns, C. Increasingly Autonomous Weapon Systems: Accountability and Responsibility. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014, p. 47. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

[221] Heyns. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, supra note 16, para. 77.

[222] Taddeo, M. and Taylor, I. *Ethical Principles for Artificial Intelligence in the Defence and Security Domain - Part 1 of 2*. The Alan Turing Institute. 2021; Sparrow, R. Killer Robots. *Journal of Applied Philosophy*. 2007. 24 (1). 62–77.

[223] Liu, Hin-Yan. Categorisation and Legality of Autonomous and Remote Weapon Systems. *International Review of the Red Cross*. 2012. 627(94), p. 629. At: https://international-review.icrc.org/sites/default/files/irrc-886-liu.pdf (last accessed 14 April 2022).

[224] Sassòli. *International Humanitarian Law Rules, Controversies, and Solutions to Problems Arising in Warfare*, supra note 187, p. 526, para. 10.91.

deploy the machine in the first place. Even if, hypothetically, machines were able to create and programme new weapons, there is still a human who programmed the original machine to be able to do that. As the ICRC puts it, "[t]*he rules of international humanitarian law are addressed to humans. It is humans that comply with and implement the law, and it is humans who will be held accountable for violations*."[225] That is one of the reasons why criminal responsibility should not be put aside that quickly.

Individual criminal responsibility can be dated back to the Lieber Code and is a long-standing rule of customary international law. It also has its place in IHL.[226] According to the ICRC, "[i]*t is a basic principle of criminal law that individual criminal responsibility for a crime includes attempting to commit such crime, as well as assisting in, facilitating, aiding or abetting, the commission of a crime.*"[227] The question remains, who is the person responsible for grave breaches of IHL committed through the acts of LAWS? The UK has expressed that legal responsibility for any military activity remains with the last person to issue the command authorising a specific action.[228] Others claim that persons who could be considered responsible include programmers, manufacturers, officers who deploy the autonomous weapon systems, military commanders, and political leaders.[229]

Heyns has proposed that "[s]*ince a commander can be held accountable for an autonomous human subordinate, holding a commander accountable for an autonomous robot subordinate may appear analogous*."[230] On the other hand, Sassòli argues that a commander's responsibility for deploying LAWS would rather be a case of direct responsibility than command responsibility under international law, just as that of a soldier firing a mortar. If the weapon system is unpredictable, a commander would be responsible for the mere fact of having deployed them.[231] However, that presupposes that the commander in question had

[225] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 7.

[226] ICRC. *Customary International Humanitarian Law*, supra note 126, Rule 102; Art. 33(1) GCIV; Art. 75(4)(b) AP I; Art. 6(2)(b) AP II; Art. 25(2) Rome Statute of the International Criminal Court, adopted on 17 July 1998, entered into force on 1 July 2002 ("ICC Statute"); Art. 5(3) American Convention on Human Rights, adopted on 22 November 1969, entered into force on 18 July 1978, etc.

[227] ICRC. *Customary International Humanitarian Law*, supra note 126, Rule 102, under "Interpretation", available at: https://ihl-databases.icrc.org/customary-ihl/eng/docindex/v1_rul_rule102

[228] UK Ministry of Defence, Joint Doctrine Note 2/11, The UK Approach to Unmanned Aircraft Systems, supra note 17, para. 510.

[229] Heyns. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, supra note 16, para. 77.

[230] Ibid, para. 78.

[231] Sassòli. *International Humanitarian Law Rules, Controversies, and Solutions to Problems Arising in Warfare*, supra note 187, p. 527, para. 10.94.

sufficient information about the weapon system itself and was aware of the likelihood that the consequences may take place. In other words, the requisite mental element has to be established. The following part will consider the challenges posed by the deployment of LAWS to the interpretation of the relevant international criminal law ("ICL") concepts, including *mens rea*.

5.2 The relevance of ICL

Preliminarily, it must be noted that ICL does not prohibit deploying certain types of weapons, *per se*. The lawfulness of weapons and their use in armed conflict is to be determined according to the rules of IHL. However, when it comes to criminal responsibility for violations of IHL, the two branches of international law are intertwined.

When humans interact with LAWS in the targeting process, the attribution of criminal responsibility for targeting-related war crimes raises specific legal challenges. For example, an issue arises when the mental element must be established. According to the ICRC, under IHL the exact mental element varies depending on the crime concerned.[232] For example, under Additional Protocol I, certain acts shall be regarded as grave breaches, when committed **wilfully**, and causing death or serious injury to body or health.[233] With regards to the proportionality rule, AP I considers it a grave breach to "[launch] *an attack* ***in the knowledge*** *that such attack will cause incidental loss of civilian life, injury to civilians or damage to civilian objects which would be clearly excessive in relation to the concrete and direct military advantage anticipated*."[234] Since the words "in the knowledge" are added, there only is a grave breach if the person committing the act knew that the described results would ensue and this does not cover recklessness.[235] However, when it comes to attacks directed at civilians, simply "*making the civilian population or individual civilians the object of attack*" is prohibited, without specifying the necessary *mens rea*. This raises the question: Which standard should be required? To answer that, one should look at the relevant practice.

[232] See ICRC. Paper prepared by the International Committee of the Red Cross relating to the crimes listed in article 8, paragraph 2 (e) (i), (ii), (iii), (iv), (ix) and (x), of the Rome Statute of the International Criminal Court. Doc. PCNICC/1999/WGEC/INF.2/Add.4. 15 December 1999. At: https://www.legal-tools.org/doc/dc889c/pdf (last accessed 15 April 2022).

[233] Art. 85(3) AP I.

[234] Art. 85(3)(b) AP I (emphasis added).

[235] ICRC. *Commentary on the Additional Protocols*, supra note 111, para. 3479.

ICL generally requires intent (*dolus*) in the sense of excluding risk-taking forms of criminal behaviour. National legal systems and the case law of international tribunals consider *dolus eventualis* the lowest sufficient standard of *mens rea*, and so does customary international law. Since the Geneva Conventions and Additional Protocol I require States to prosecute grave breaches of IHL, but oftentimes do not provide for the mental element necessary, the standard applicable in ICL should be taken into account. In the context of LAWS, if commanders deploy an autonomous weapon system and its actions lead to a grave breach, IHL obliges States to hold them responsible. Any prosecution will be carried out under the relevant rules of ICL. If those rules require a certain standard of *mens rea*, like *dolus eventualis*, that is the standard that needs to be considered in the deployment of LAWS. The parameters of the human-machine interaction therefore must be set up in a way that ensures that operators exercise a sufficient level of control. The requirement of meaningful human control ensures that the lack of *mens rea* under ICL will not prevent persecution of grave breaches.

It must be noted that IHL poses realistic standards on the conduct of hostilities, its rules are mostly obligations of conduct, rather than results. Not every civilian death constitutes a violation of IHL. And not every violation of IHL gives rise to specific obligations of repression for States, such as to prosecute individuals. There are, however, situations which illustrate the difficulties of applying the current framework of international criminal law to the modalities and consequences of the use of LAWS. Can a commander who has reason short of certainty to doubt the deployment of or reliance on LAWS, that is, the soldier who acts negligently or recklessly or with *dolus eventualis* be held responsible?[236] How can this situation be reconciled with the mental element required?

[236] Jain, A. G. Autonomous Cyber Capabilities and Individual Criminal Responsibility for War Crimes. In: Liivoja, R. and Väljataga, A. *Autonomous Cyber Capabilities Under International Law*. NATO Cooperative Cyber Defence Centre, 2021, p. 300.

5.3 The need for MHC

Since LAWS may make targeting decisions autonomously, and given the very specific *mens rea* requirements for criminal responsibility for war crimes, is control over LAWS required for the purposes of attracting responsibility? Key concepts of ICL such as intent, foreseeability, voluntary act, and causality indeed entail some control conditions. Most importantly, individual criminal responsibility is causal responsibility: it requires human causal control over events.[237] Control is thus a concept already embedded in criminal law. Bo even argues that "*an obligation to ensure meaningful human control is functional to a fair ascription of criminal responsibility.*"[238] The CCW GGE considers that a focus on characteristics related to the human element in the use of force and its interface with machines is necessary for addressing accountability and responsibility.[239] Indeed, if one accepts that individual responsibility offers viable solutions to attributions of acts of LAWS to their human operators, it is still subject to stringent conditions, particularly concerning the intent of the operators and their knowledge about the weapon systems used.

Some suggest that *dolus eventualis* is a sufficient *mens rea* in national legal systems and in the case law of other international tribunals and must be considered customary international law as well as a general principle of law.[240] Bo argues that, for example, the crucial provision for determining standards of *mens rea* for committing the war crime of attacking civilians under the ICC Statute is Article 30(2)(b), pursuant to which a person has intent in relation to consequences if he "*means to cause that consequence*" (first alternative) or if he is at least "*aware that* [the consequence] *will occur in the ordinary course of events*" (second alternative).[241] In the context of LAWS, the second alternative would be particularly relevant. Although commanders may mean to deploy LAWS in order to commit war crimes, the more problematic scenario would be where the human operator deploying autonomous weapon

[237] Bo, M. Meaningful Human Control over Autonomous Weapon Systems: An (International) Criminal Law Account. *Opinion Juris*. 18 December 2020. At: http://opiniojuris.org/2020/12/18/meaningful-human-control-over-autonomous-weapon-systems-an-international-criminal-law-account/ (last accessed 27 April 2022).

[238] Ibid.

[239] CCW. Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. CCW/GGE.1/2018/3, p. 5. At: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/323/29/PDF/G1832329.pdf?OpenElement (last accessed 15 April 2022).

[240] Werle, G. and Jeßberger, F. *Principles of International Criminal Law*. Oxford: Oxford University Press, 2020, para. 576.

[241] Bo, M. Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute. *Journal of International Criminal Justice*. 2021. 19(2). 275–299, pp 286-287. At: https://doi.org/10.1093/jicj/mqab005 (last accessed 2 May 2022).

systems envisages and accepts the risk of civilians being attacked. In Bo's opinion, situations like this would fall under indiscriminate attacks that "*are characterized by the lack of certain knowledge of the civilian status of the targets and by the* awareness of the possibility *that some of the targeted persons might have civilian status or by the perpetrator's 'awareness of his lack of awareness' as to whether some civilians might be hit*".[242] However, it is unclear from the ICC case law whether the perpetrator must be certain or whether awareness of the high probability or of the possibility that some of the targets are civilians is sufficient.[243] Bo argues that the latter should be the case because this interpretation would allow for the attribution of individual responsibility for indiscriminate attacks stemming from the use of LAWS.[244]

It must be mentioned that even though the Pre-Trial Chamber of the ICC has once ruled that *dolus eventualis* is covered by the scope of Article 30(2)(b),[245] the Appeals Chamber has so far ruled out the applicability of that standard under Art. 30 of the ICC Statute.[246] While it is true that the doctrine of *stare decisis* should not apply before the ICC a change in its jurisprudence regarding *dolus eventualis* should therefore be possible,[247] the Appeals Chamber has repeatedly confirmed its approach. Importantly, the ICC is not the only Court dealing with war crimes. Those can equally be prosecuted before other international tribunals or even domestic courts based on the principle of universal jurisdiction. Should the arguments above regarding the interpretation of the ICC Statute concerning the applicability of *dolus eventualis* be rejected, there is little doubt that this *mens rea* standard is considered customary law. Thus, it can be applied by other courts and tribunals than the ICC.

But even if we consider *dolus eventualis* the applicable standard, how can its rather strict conditions be fulfilled, for example, when it comes to indiscriminate attacks or attacks directed at civilians? Determining the level of human control we consider sufficient can be a way how to allow for attribution of responsibility. If meaningful human control over LAWS is

[242] Ibid, p. 291 (references omitted).

[243] Ibid, p. 290.

[244] Ibid, p. 295.

[245] See for example: Lipovský, M. Mental Element (Mens Rea) of the Crime of Aggression and Related Issues. In: Šturma, P. *The Rome Statute of the ICC at Its Twentieth Anniversary*. Leiden, The Netherlands: Brill | Nijhoff, 2018, p. 116. doi: https://doi.org/10.1163/9789004387553_008; ICC, Decision on the Confirmation of Charges, Lubanga (ICC-01/04-01/06), Pre-Trial Chamber I (29 January 2007), § 352 (let. ii).

[246] Bo. Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute, supra note 241, p. 280.

[247] Art. 21(2) ICC Statute.

exercised, it makes it possible to prove the knowledge and intent of the operator. Suppose the actions of LAWS fulfil the *actus reus* of a particular war crime. The operators may be held responsible if they were sufficiently aware of the likelihood that this result would occur. The requirement of meaningful human helps to ensure that this is possible. On the other hand, it also prevents unfair attribution of responsibility to commanders who deployed LAWS and had no reason to doubt their efficiency and precision. In this case, IHL sees no violation, even if civilian deaths may occur. However, the trust the commander puts in the operation of the weapon system should be reasonable and justified - and this is precisely what meaningful human control helps to ensure. Therefore, the requirement of MHC can enable fair attribution of individual criminal responsibility.in cases where prosecution of violations is required by IHL.

III. Control exercised over weapons currently in use

Having explored the reasons why meaningful human control is not only desirable but also necessary, then following Chapter will focus on the nature and quality of human control as it is exercised over weapobns currently used in warfare. While the requirement of meaningful human control has been introduced with future weapon systems in mind, some highlight the importance of how, in reality, practices shape norms.[248] Non-verbalised practices also influence the understanding of appropriateness, which may be the case also when it comes to LAWS. While States in the CCW GGE meetings primarily focus on weapon systems that may be developed in the future, some claim that, for example, most air defence systems already have significant autonomy in the targeting process and military aircraft have highly automatised features.[249] Below, an argument will be explored whether current practices have set a precedent and created an understanding of meaningful human control.

1. Systems with automated/autonomous functions setting a precedent

Bode and Watts argue that the focus of the CCW GGE discussion directed toward emerging technologies is problematic because it "*risks missing the important precedents for what counts as meaningful human control set by existing technologies*."[250] In their opinion, automated and autonomous features have been integrated into the critical functions of air defence systems for decades. From these examples, one can deduce how "meaningful human control" is currently exercised. Their study of human-machine interaction in air defence systems has led them to believe that while human operators formally retain the final say in specific targeting decisions, the "meaningfulness" of this decision is debatable. They identify three main problems: (1) the complexities of human-machine interaction do not allow for situational awareness; (2) human operators are not equipped with the expertise necessary to understand the system; and (3) the setting of the system does not give them the time to engage in deliberation.[251]

[248] Bode, I. and Huelss, H. The Future of Remote Warfare? Artificial Intelligence, Weapons Systems and Human Control. In: McKay, A., Watson, A. and Karlshøj-Pedersen, M. *Remote Warfare: Interdisciplinary Perspectives.* E-International Relations, 2021, 218-233, p. 224.

[249] Boulanin and Verbruggen. *Mapping the Development of Autonomy in Weapons Systems*, supra note 15.

[250] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 28.

[251] Ibid, p. 61.

1.1 Case study of an air defence system downing a fighter jet

One of the examples Bode and Watts provide is the US MIM-104 Patriot that was involved in two fratricidal engagements in 2003 in Iraq, killing three US crewmembers.[252] In simple terms, the Patriot's radar tracks objects in the air and its engagement algorithm "*identifies those objects, and then displays them as symbols on a screen*".[253] How the next steps look like depends on the mode of deployment. In semi-automatic mode, the human operators receive computer-based engagement support, but they make all the critical targeting decisions.[254] The Patriot thus functions as a "human-in-the-loop" system. In automatic mode, however, Patriot becomes a "human-on-the-loop" system, being "[...] *nearly autonomous, with only the final launch decision requiring human interaction.*"[255] When an incoming threat is detected, the system is put into a "ready" state. Afterwards, it can fire without further human engagement.[256] Human operators continue to monitor the command module, but the Patriot system is "*capable of applying lethal force with little or minimal direct human oversight.*"[257] Independent of how the human-machine cooperation may be described, the decisive factor is how the system operates in reality. This mode of engagement effectively reduces the human operator's role to veto power in targeting decisions. Moreover, a problematic fact is that Patriot operators only have a few seconds to exercise their veto.[258] While the categorisation of air defence systems as "autonomous weapon systems" is not universally agreed upon, a closer look at the functioning of the Patriot shows that they can indeed fulfil the requirements of autonomy, depending on the mode they are deployed in.

[252] US DoD. Report of the Defense Science Board Task Force on Patriot System Performance. Report Summary. Washington, DC: Office of the Under-Secretary of Defense for Acquisition, Technology, and Logistics, January 2005, p. 2. At: https://dsb.cto.mil/reports/2000s/ADA435837.pdf (last accessed 15 April 2022).

[253] Leung, R. The Patriot Flawed? CBS News. 19 February 2004. At: https://www.cbsnews.com/news/the-patriot-flawed-19-02-2004/ (last accessed 15 April 2022).

[254] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 53.

[255] Missile Defense Project. "Patriot," Missile Threat. 2018. At: https://missilethreat.csis.org/system/patriot/ (last accessed 15 April 2022).

[256] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 53.

[257] Hawley, J. K. Patriot Wars: Automation and the Patriot Air and Missile Defense System. Center for a New American Security Project on Ethical Autonomy Working Paper. 2017, p. 4. At: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf?mtime=20170106135013&focal=none (last accessed 2 May 2022).

[258] Leung. The Patriot Flawed?, supra note 229.

1.2 Classification problems

The British Board of Inquiry conducted a report of the fratricidal engagements in 2003, which lists multiple interrelated factors that have contributed to the incidents, among which were: (1) the system's missile classification criteria; (2) the firing doctrine and crew training; (3) autonomous operation; as well as (4) orders and instructions.[259] The list represented here is certainly not exhaustive. It features only those factors relevant to the nature of human-machine interaction, as it is tied with the analysis of meaningful human control.

Concerning the system's missile classification criteria, Bode and Watts conclude that *track classification* problems in the Patriot system were a major factor leading to both incidents. The Patriot system classifies "tracks" and targets as aircraft, different kinds of missiles, or other categories based on "*flight profiles and other track characteristics such as point of origin and compliance with Airspace Control Orders*."[260] The system thus does not operate on object recognition based on specific characteristics but instead focuses on flight profiles. Patriot is programmed to defend against an envelope of possible target profiles.[261] The reason for that is precisely the precision-recall trade-off discussed earlier. Patriot's system was designed to avoid the high number of false negatives that inevitably appear when the target's parameters are defined too precisely. Apparently, the system's track classification suffered from occasional misclassifications, which had been known prior to the incidents. Rather than communicating those deficiencies to the system's operators, the US Army framed these as a software problem, a fix of which did not present difficulties.[262] This rhetoric contributed to an over-trust in the system.[263] This underlines the importance of proper instructions and informing the operating crew. The misclassification issue is itself a severe problem, which is, however, reinforced by human mistakes in communicating this issue.

[259] UK Ministry of Defence. Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710. 2004, pp 2-3. At: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/82817/maas03_02_tornado_zg710_22mar03.pdf (last accessed 15 April 2022).

[260] Hawley, J.K. and Mares, A.L. Human Performance Challenges for the Future Force: Lessons from Patriot after the Second Gulf War. In: Savage-Knepshield, P. *Designing Soldier Systems: Current Issues in Human* Factors. Burlington, VT: Ashgate, 2012, 3-34, pp 6-7.

[261] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 54.

[262] Hawley and Mares. Human Performance Challenges for the Future Force, supra note 260, p. 7.

[263] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 55.

1.3 Over-trust in the system

Furthermore, in the British Ministry of Defence's assessment, "*Patriot crews are trained to react quickly, engage early and to trust the Patriot system.* [...] *The crew had about one minute to decide whether to engage.*"[264] Not only the operators lacked the necessary information and understanding of the system's functioning and deficiencies, but they also lacked the time needed for a meaningful deliberation on whether to overrule the targeting decisions made by the system. The natural consequence was an unwarranted trust in the system. The Defence Science Board's review of the Patriot noted: "*The operating protocol was largely automatic, and the operators were trained to trust the system's software; a design that would be needed for heavy missile attacks.*"[265] Certainly, some level of trust in the system is necessary. Otherwise, it could not even be deployed, and it would not enable its human operators to benefit from the capabilities of an automated weapon system, most notably its speed. On the other hand, this trust should come from a place of being fully informed and aware of the advantages and imperfections of the weapon system. Even if there is a human operator overlooking the operation of an automated (or indeed autonomous) weapon system, no meaningful control is retained when unfounded trust in the system is deeply ingrained into the training.

1.4 Precedent for the role of the human operator?

The analysis of the malfunctioning of the Patriot system reveals the challenges human operators face while remaining "on the loop". Hawley refers to this as the "*humans' residual role in system control*"[266] and emphasises how difficult the role is to perform. To be able to distinguish when to trust the system and when to question its decision, human operators must understand how the system works, what its weaknesses are, and retain situational awareness.[267] For human operators, the Patriot, is, therefore, "*knowledge-intensive in terms of the amount of information required to characterise and comprehend the system*"[268] while at the same time, they can be underloaded with tasks vis-à-vis those delegated to the system.[269]

[264] UK Ministry of Defence. Aircraft Accident to Royal Air Force Tornado, supra note 259, p. 3.

[265] US DoD. Report on Patriot System Performance, supra note 252, p. 2.

[266] Hawley. Patriot Wars: Automation and the Patriot Air and Missile Defense System, supra note 257, p. 2.

[267] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 56.

[268] Hawley, J. K. Looking Back at 20 Years on MANPRINT on Patriot. *Army Research Laboratory*. 2007, p. 1. At: https://apps.dtic.mil/sti/pdfs/ADA472740.pdf (last accessed 15 April 2022).

[269] Kantowitz, B. H. and Sorkin, R. D. Allocation of Functions. In: Salvendy, G. *Handbook of Human Factors*. New York: Wiley, 1987, 355–69.

This leads to what Bode and Watts call "*a minimal but impossibly complex role*".[270] "Minimal" in the sense that the human operators often simply monitor the system's operation and may exercise their veto powers. "Complex" in that what appears to be a straightforward task on paper is much more challenging to execute in reality. There is a lack of knowledge about how the system reaches its decisions, accompanied by the need to be constantly vigilant and aware of all possible circumstances that might have escaped the system's attention due to its inability to contextualise.

While the exact requirements for and elements of meaningful human control will be elaborated upon below,[271] the purpose of this real-world example was to demonstrate how weapon systems are being deployed and the consequences the lack of human control can have. The case of the Patriot shows how the operation of air defence systems has contributed toward setting a trend of how human-machine interaction looks like and what its acceptable quality is.[272] Bode and Watts use this observation as an argument for the inclusion of air defence systems in the debate on LAWS, as well as to draw attention to the fact that the way these systems are currently used is eroding what we may consider meaningful. These emerging, silent norms on the quality of human engagement with weapon systems are potentially alarming when considering how practice may shape norms. For example, since States have started using drone technology, they have adopted novel interpretations of principles governing the use of force, such as attribution and imminence. While those principles concern *jus ad bellum*, it shows how new technology used in warfare can impact even longstanding concepts of international law, not to mention emerging concepts such as meaningful human control. This presents an argument on why the requirement of meaningful human control should be further explored, defined, and incorporated in deploying all weapon systems with autonomous functions.

[270] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 62.

[271] See Chapter VI section 3 below.

[272] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 58.

IV. Legal Basis for Human Control

The previous section was devoted to analysing the level of human control over weapons currently used. On that background, the following part will explore the legal basis for the requirement of retaining a certain level of human control over weapons in general and LAWS in particular. Relevant IHL rules will be analysed to determine whether meaningful human control is an existing binding rule or whether it is rather a moral obligation. Finally, the question of meaningful human control as a customary rule of international law will be considered.

1. Relevant rules of IHL

During an ICRC Expert Meeting in 2014, it was suggested that "[…] *human control and human decision making are implicitly and explicitly required by international human rights law and international humanitarian law. As such, it was argued that there is a need to develop a legal norm requiring, and defining, 'meaningful human control' of weapon systems, and that further discussions on this issue are vital*."[273]

This subchapter aims to examine whether human control is implicitly or explicitly required by IHL and, if so, to explore the content of such a rule further.

1.1 Rules on targeting

The ICRC argues that for conflict parties, human control over LAWS (particularly those using AI and machine-learning) employed as means and methods of warfare is required to ensure compliance with international law, specifically IHL.[274] Boutin and Woodcock suggest that the notion of a responsible agent is also reflected in IHL norms, in some instances, obligations are even explicitly directed towards individuals (such as to those who launch attacks with regard to taking feasible precautions).[275] Indeed, as has been argued above, rules on the conduct of hostilities were formulated with human judgement in mind, relying on the human ability to recognise which targets are important for the adversary, to weigh incidental casualties against

[273] ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*, supra note 34, p. 16.

[274] ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach, supra note 58, p. 7.

[275] Boutin, B. and Woodcock, T., Aspects of Realizing (Meaningful) Human Control: A Legal Perspective. Forthcoming in: Geiß, R. and Lahmann, H., *Research Handbook on Warfare and Artificial Intelligence.* Cheltenham: Edward Elgar Publishing, 2022, p. 12.

a possible military gain, and to take various steps to gain more certainty that excessive civilian losses will be avoided. In other words, probability-based and moral reasoning is at the core of the rules on targeting. Those rules are certainly context-specific but that does not necessarily mean that a human must be in control of the whole decision-making process. They were drafted in the most objective way possible, aiming to avoid arbitrary targeting decisions. As Sassòli points out, the fact that the application of the proportionality rule involves a subjective determination can either be a description of the unfortunate reality, or a normative proposition that the determination should be subjective.[276] Considering that the current doctrine cannot find agreement on this subject, it is difficult to argue that the rules on the conduct of hostilities explicitly require human judgement and human control over targeting decisions.

However, the ICRC is certainly right when it suggests that "[w]*here AI systems are used in attacks – whether as part of physical or cyber-weapon systems, or in decision-support systems – their design and use must enable combatants to make these* [targeting] *judgements*."[277] Does this mean that human control is therefore implicitly required to ensure that the very detailed rules on distinction, proportionality, and precautions can be complied with?

Looking at the state of current technology, the answer seems to be affirmative.[278] Unless it is possible to program LAWS to be able to apply the rules on targeting in a way comparable to humans, it would be unlawful to deploy an autonomous weapon system without the oversight and control of its human operator. However, this conclusion is based upon the current reality, taking into consideration what is technologically possible and what is not. Nevertheless, to determine whether IHL requires human control over targeting decisions, the conclusion must be general. The rules should be applicable not only to weapons past and present but also to weapons that may come in the future, which could possibly replicate the qualities of human reasoning. The rules on the conduct of hostilities would be embedded in the computer algorithm and its decisions would be guided by these rules in the same way as when humans act upon them. Should that be the case, it would no longer be the combatants making the

[276] Sassòli. Autonomous Weapons and International Humanitarian Law, supra note 152, p. 334.

[277] ICRC. Statements to the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts on Lethal Autonomous Weapons Systems. Geneva: 25–29 March 2019. At: https://www.unog.ch/__80256ee600585943.nsf/(httpPages)/5c00ff8e35b6466dc125839b003b62a1?OpenDocument&ExpandSection=7#_Section7 (last accessed 15 April 2022).

[278] See for a similar view: Asaro. Jus nascendi, robotic weapons and the Martens Clause, supra note 1.

targeting decisions but the autonomous weapon system itself. There seems to be no rule on the conduct of hostilities prohibiting this hypothetical situation *per se*. The rules are driven by their objective, they regulate the conduct based on its outcome, rather than focusing on the process of decision making itself. However, in such a case human control would still be present, only exercised at a prior stage, through the weapon design. Some argue, drawing on general principles of the law of armed conflict, that there is already an implicit requirement for meaningful human judgment in every individual decision to use lethal force.[279] A conclusion can be made that IHL rules on targeting are built on context-based judgement and complex evaluations, which implicitly require human control to be exercised at a certain stage of the targeting process.

1.2 The Martens Clause

One of the principles that is being pointed out in the debate on the legal basis of MHC is the Martens Clause. The clause was initially proposed by F.F. de Martens, a Russian delegate to the Hague Convention. Its introduction was motivated by concerns over extending humanitarian law to armed partisans in occupied territories.[280] Versions thereof can be found in all four Geneva Conventions of 1949 and Additional Protocols I and II.[281] The API I version of the Martens Clause states: "*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 from the dictates of public conscience*."

The clause is invoked generally in disarmament and new technology contexts because it refers explicitly to the public conscience, providing a role for public opinion and civil society representatives in the moral assessment of IHL.[282] In its advisory opinion on the Legality of the Threat or Use of Nuclear Weapons, the ICJ stated that the Martens Clause had "*proved to be an effective means of addressing rapid evolution of military technology*."[283] The Court also found that the Martens clause represents customary international law. As the ICRC stresses, "*ethical decisions by States, and by society at large, have preceded and motivated the*

[279] Asaro. On Banning Autonomous Weapon Systems, supra note 200.
[280] Mero, T. The Martens Clause, Principles of Humanity, and Dictates of Public Conscience. *American Journal of International Law*. 2000. 94(1), 78-89. At: doi:10.2307/2555232.
[281] Art. 63 GC I, Art. 62 GC II, Art. 142 GC III, Art. 158 GC IV, and Art. 1(2) AP I, preamble to AP II.
[282] Asaro. Jus nascendi, robotic weapons and the Martens Clause, supra note 1.
[283] ICJ, Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, ICJ Reports 1996, para. 78.

development of new international legal constraints in warfare, including constraints on weapons that cause unacceptable harm. In international humanitarian law, notions of humanity and public conscience are drawn from the Martens Clause."[284]

Its interpretation is, inevitably, subject to considerable debate. The rather minimalist interpretation of the clause is that acts are not necessarily legal or permissible simply because they are not explicitly prohibited by humanitarian law. This approach modifies the well-known Lotus principle[285] when it comes to international humanitarian law. Others argue that the dictates of the public conscience drive the evolution of custom, and perhaps of the law as a whole, by inspiring treaty negotiators.[286] On the contrary, Schmitt has suggested that the Martens clause "*applies only in the absence of treaty law. In other words, it is a failsafe mechanism meant to address lacunae in the law; it does not act as an overarching principle that must be considered in every case*."[287] Be it as it may, even the strictest interpretation would still subscribe to some relevance of the Martens Clause in the context of human control over weapon systems. Their use raises moral and legal concerns that are indeed not addressed explicitly by the rules of IHL.

The question thus is whether "the principles of humanity and the dictates of public conscience" require that humans remain in control of targeting decisions. Many claim so. Asaro argues that "[i]*f any new principle might be convincingly derived from the "principles of humanity" as expressed in the Martens Clause, surely it would be a principle that ensures human control over the violence of war*".[288] Others suggest that a broad interpretation of the Martens clause may render the use of LAWS illegal under existing IHL, to the extent that it

[284] ICRC. Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?, p. 1. At: https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control (last accessed 15 April 2022).

[285] Named after The Case of the S.S. Lotus, 1927 PCIJ Series A, No. 10. According to the classical formulation of this principle, "whatever is not explicitly prohibited by international law is permitted". See for example: Weil, P. The Court Cannot Conclude Definitively... Non Liquet Revisited. *Columbia Journal of Transnational Law*. 1998. 109(36), p. 112. At: https://heinonline.org/HOL/LandingPage?handle=hein.journals/cjtl36&div=14&id=&page= (last accessed 15 April 2022); Roth, B.R. The Enduring Significance of State Sovereignty. *Florida Law Review*. 2004. 1017(56), p. 1029. At: https://digitalcommons.wayne.edu/lawfrp/188/ (last accessed 15 April 2022).

[286] W Boothby, W. *Weapons and the Law of Armed Conflict.* Oxford: Oxford University Press, 2009, p. 14. See also Dinstein, Y. *The Conduct of Hostilities under the Law of International Armed Conflict*. Cambridge: Cambridge University Press, 2004, p. 57.

[287] M. Schmitt, Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics, supra note 151, p. 32.

[288] Asaro. Jus nascendi, robotic weapons and the Martens Clause, supra note 1, referring to: Roff, H. M. The Strategic Robot Problem: Lethal Autonomous Weapons in War. *Journal of Military Ethics*. 2014. 13(3). 211-227. At: https://doi.org/10.1080/15027570.2014.975010 (last accessed 15 April 2022).

can be shown that the dictates of public conscience and the principles of humanity abjure their use.[289]

However, determining what the "public" deems inhuman is a difficult task. In any survey research, the methodology can distort the expression of opinion by choice of words as well as the order of questions.[290] Sparrow suggests that public debate conducted in an open society could generate more reliable conclusions, provided that the participants in public debate state the reasons for their views.[291] This, in turn, may influence others and create not just a collection of opinions of individual members of the public but rather a shared belief that resulted from the free exchange of thoughts. As much as this approach may yield reliable conclusions, this proves to be very problematic when it comes to the regulation of LAWS, considering again how much the debate is future-oriented. It may as well be the case that the use of lethal force by autonomous weapon systems goes against the dictates of public conscience. Still, one cannot imagine how such a public debate could be realistically led before introducing LAWS into world warfare. Hence, it appears unlikely that the Martens Clause would, on its own, prohibit the use of LAWS without human control over them.

2. MHC as a rule of customary international law

The requirement of meaningful human control could also have its legal basis as an independent rule of customary international law. In the highly polarised and inconclusive debate over LAWS, already in 2015, there seemed to be widespread consensus on requiring a certain level of human control.[292] No real opposition and immediate positive reactions have led some to conclude that it is either a newly developed customary norm or a pre-existing, recently exposed rule of customary international law.[293] This contention would align with the

[289] Sparrow, R. *Ethics as a source of law: The Martens clause and autonomous weapons*. 2017. At: https://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/ (last accessed 15 April 2022).

[290] Rose, N. and Osborne, T. Do the Social Sciences Create Phenomena: The Case of Public Opinion Research. 1999. *BRIT. J. SOC*. 367(50). At: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-4446.1999.00367.x (last accessed 15 April 2022).

[291] Sparrow. *Ethics as a source of law: The Martens clause and autonomous weapons*, supra note 289.

[292] Biontino, M. (Chairperson of the Informal Meeting of Experts). Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS). Geneva: 2015, p. 11. At: http://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2015/Draft Report.pdf (last accessed 15 April 2022).

[293] Crootof, R. A Meaningful Floor for Meaningful Human Control. *Temple International & Comparative Law Journal*. 2016 (30). At: https://ssrn.com/abstract=2705560 last accessed 15 April 2022); referring to: Asaro. Jus nascendi, robotic weapons and the Martens Clause, supra note 1. See also Horowitz, M. and Scharre, P. Meaningful Human Control in Weapon Systems: A Primer. Center for a New American Security Project on

previously reached conclusion that IHL implicitly requires human control. Nevertheless, there is disagreement in scholarship as to whether MHC is an existing or emerging international legal standard.[294]

Crootof argues that the downside of this broad support "*comes at a familiar legislative cost; there is no consensus as to what 'meaningful human control' actually requires*."[295] However, the abstract character of the rule would neither be an exception in international law nor an obstacle. Many of the core customary rules of IHL are of a somewhat abstract nature, and their content has been clarified and specified through practice and interpretation. Relying on the doctrine of two elements[296] when identifying whether a rule is of customary character, States have been voicing their support for the binding nature of the rule of meaningful human control, thus perhaps demonstrating their *opinio juris*. The real problem lies in proving enough relevant State practice in this regard. Can a customary rule addressing the mode of use of LAWS arise when these weapon systems are not yet being deployed, at least not in their fully autonomous mode? Can we look at practice regarding existing automated weapons to make conclusions about any customary rule on the use of LAWS?

In the social sciences, social norms are conceived as reliably observable patterns of social behaviour that manifest an underlying set of shared beliefs about the acceptability of certain conduct.[297] Customary rules of international law are indeed social norms on a bigger scale, worldwide, regional, or even bilateral. In theory, a custom requires consistent actions of a significant number of (specially affected) states based on beliefs that there is an obligation to act so. When it comes to new technology being introduced, the patterns of States' behaviour usually change in response, and new norms (including customary rules) may be created. Asaro points out an issue connected to the emergence of customary rules prior to developing new technology: "*If we have not yet implemented a new technology, we cannot observe what the new norms are (if we limit norms to already-recognized and accepted*

Ethical Autonomy Working Paper. 2015, p. 6. At: www.cnas.org/sites/default/files/publications-pdf/Ethical_Autonomy_Working_Paper_031315.pdf (last accessed 15 April 2022).

[294] Boutin and Woodcock. Aspects of Realizing (Meaningful) Human Control, supra note 275, p. 10.

[295] Crootof, R. A Meaningful Floor for Meaningful Human Control, supra note 293, p. 54.

[296] The ICJ has developed an approach to determining a rule of customary international law which highlighted long-standing practice and usage, see for example: ICJ, Right of Passage over Indian Territory (Portugal v. India), Merits, Judgment, ICJ Reports 1960, p. 40.

[297] Pospisil, L. The Attributes of Law. In: Bohannon, P. *Law and Warfare: Studies in the Anthropology of Conflict.* New York: American Museum of Natural History, 1967, 25-41.

behavior). We can examine existing norms and try to determine if the use of a new technology would challenge or violate those norms. If so, we might try to regulate that technology and try to ensure that the norm remains in effect."[298]

On the other hand, it may as well be the case that the capabilities of new technology will create situations that prompt us to explicitly recognise norms that had always been tacitly assumed, taken for granted. But obeyance with these norms has never been articulated because it simply was not necessary. Therefore, it is difficult to find examples of State practise or *opinio juris*. Asaro argues that the requirement of meaningful human control is such a case, given the broad-based agreement of States at the CCW GGE meetings. In his view, it constitutes an "emerging principle".[299] However, he recognises that it is less clear whether States really believe that they already have the obligation to subject targeting decisions to meaningful human control, in other words, whether there is *opinio juris*.

States have indeed expressed their views on the obligation to maintain meaningful human control over targeting decisions and the nature of this obligation. The US DoD, for instance, published a policy that directs that "*autonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force*."[300] And according to the Netherlands, "*meaningful human control is required in the deployment of autonomous weapon systems*".[301]

Austria has expressed that "[…] *the enhanced technological capabilities,* […] *which potentially include the notion of transferring control over (lethal) weapon systems to machines, make the question of meaningful human control all the more important. In our view, these questions* [of accountability, responsibility and ultimately political responsibility] *cannot be fully answered in the context of existing norms, but require further clarity to prevent unintended consequences in the long run*."[302] Given the phrasing here, it appears as if MHC was a requirement that has yet to be established.

[298] Asaro, P. Jus nascendi, robotic weapons and the Martens Clause, supra note 1.

[299] Ibid.

[300] US DoD. Directive 3000.09 on Autonomy in Weapon Systems, supra note 2, p. 2.

[301] Government of the Netherlands. Government Response to AIV/CAVV Advisory Report No. 97, 'Autonomous Weapon Systems: The Need for Meaningful Human Control'. 2 March 2016. At: http://aiv-advies.nl/ (last accessed 15 April 2022).

[302] Chairperson's Summary of the CCW GGE Meeting of 19 April 2021. CCW/GGE.1/2020/WP.7, Annex III, Commentaries on the 11 guiding principles, p. 28. At: https://documents.unoda.org/wp-content/uploads/2020/07/CCW_GGE1_2020_WP_7-ADVANCE.pdf (last accessed 15 April 2022).

According to Japan, "[…] *it is indispensable that a lethal weapon system be accompanied with meaningful human control by securing proper operation and be operated by persons with sufficient information on such weapons systems. It would be necessary to deepen discussion on where and how much meaningful human control is necessary in the life-cycle of weapons systems*."[303] In a similar vein, Panama considers that "*LAWS would not have the ability to make these assessments due to their mechanical intelligence.* […] *They should not replace human beings in the work of discernment and take decisions about their own use.*"[304] In both statements, the States stress the importance of MHC but suggest that this principle and its characteristics should be agreed upon in the future.

Portugal's view on the nature of the obligation to maintain MHC over lethal autonomous weapon systems is relatively clear when it states that "[b]*y reaffirming the need for a meaningful human control over LAWS at all stages of their life cycle, the GGE-LAWS is not merely translating into a principle several legal imperatives stated by IHL and other areas of international law. In addition to concerns on accountability, ethical and moral considerations preside over this and other Guiding Principles, addressing a fundamental problem where the use of lethal force by machines is involved: human life and human safety cannot be left to the autonomous decision/choice of a machine/algorithm*."[305] In other words, according to Portugal, meaningful human control does not represent an overarching principle deduced from the fundamental IHL rules on the conduct of hostilities but rather a principle emerging together with the possibility of machines making decisions affecting human life.

A conclusion can be made that the majority of States recognise the crucial importance of the requirement of human control and wish to see it put into practice. However, they do not consider it a customary rule of IHL that has been dormant. Instead, they see it as a rule to be applied *pro futuro*.

Furthermore, another issue lies in proving relevant State practice. Considering the previous analysis of the Patriot system downing a friendly fighter jet introduced above, it seems that while States are determined to require the relevant standard of human control (or indeed "appropriate levels of human judgment" when it comes to certain countries) in LAWS deployed in the future, there is certainly lack of rules and clarity over current practice. Even

[303] Ibid, p. 57.
[304] Ibid, p. 65.
[305] Ibid, p. 74.

though some argue that weapon systems have already been deployed in what should be considered an "autonomous mode", States reject this contention and claim that their weapon systems currently in use are at most highly automated. This creates undesirable confusion and makes it even more challenging to see how the principle of human control could fulfil the requirements of a customary rule of international law.

On the other hand, in the North Sea Continental Shelf Judgment, the ICJ clarified that even without the passage of any considerable period, very widespread and representative participation in a multilateral convention might suffice in itself to generate customary rules, provided that it included that of States whose interests were especially affected.[306] It is also argued that resolutions adopted by the General Assembly could potentially be relied on to establish both State practice and *opinio juris*. Through resolutions, an *opinio juris* recognising the existence of a rule of international law could be expressed even before the emergence of a corresponding practice.[307] To some scholars, *opinio juris* is the essence of customary law and that state practice is either entirely unnecessary or at least unessential to prove.[308] Bin Cheng argued that customary law could be created instantaneously, among all or some Members of the UN.[309]

Therefore, there seem to be paths that could establish a customary rule requiring meaningful human control to be exercised over targeting decisions even prior to their development and deployment. A multilateral convention emerging from the CCW GGE meetings with widespread and representative participation could potentially be one of them. One must, however, consider the differences between formally concluding a multilateral treaty and mere expressions of States' opinions during the meetings. Those do not bind States, and it is dubious that they could spark the same effects. The second path could be a General Assembly resolution. Even though formally non-binding, it could, in certain circumstances, provide important evidence of the existence of a rule requiring MHC or the emergence of an *opinio juris*.[310] However, although some States have brought up the topic of autonomous weapons in

[306] ICJ, North Sea Continental Shelf, Judgment, ICJ Reports 1969, p. 42, para. 73.
[307] Yusuf, A. A. Statement of the President of the International Court of Justice before the Sixth Committee of the General Assembly. New York: 1 November 2019, pp 4-5. At: https://www.icj-cij.org/public/files/press-releases/0/000-20191101-STA-01-00-EN.pdf (last accessed 15 April 2022).
[308] See, e.g.: Bin Cheng, Cheng, B. United Nations Resolutions on Outer Space: 'Instant' International Customary Law? In: Cheng, B. *International Law: Teaching and Practice*, London: Stevens, 1982; Guzman, A. T. *How International Law Works: A Rational Choice Theory*. Oxford: Oxford University Press, 2008.
[309] Ibid, Cheng. 'Instant' International Customary Law?, p. 252.
[310] ICJ, Legality of the Threat or Use of Nuclear Weapons, supra note 283, para. 70.

the General Assembly forum, there has been no resolution up to date. The lack of relevant resolutions on LAWS does not allow to draw any conclusions on potential *opinio juris* in favour of the customary nature of the MHC requirement.

Interestingly, in the case of Prosecutor v. Kupreskić, the ICTY developed a creative interpretation of the above-mentioned Martens Clause, linking it to customary law: "*In the light of the way states and courts have implemented it, this Clause clearly shows that principles of international humanitarian law may emerge through a customary process under the pressure of the demands of humanity or the dictates of public conscience, even where state practice is scant or inconsistent.*"[311] This understanding of the Clause seems to suggest that even though State practice may be lacking, the principles embodied in the Clause can serve as an element of a customary rule of international law.

In conclusion, there are possible ways how the obligation to maintain meaningful human control over LAWS could reach the status of a customary rule of international law, even prior to the development or deployment of the weapon systems. However, it has been argued that even widespread agreement of States expressed during the CCW GGE meeting would likely not suffice, all the more because States themselves seem to consider that the requirement and its characteristics should be agreed upon in the future.

[311] ICTY, Prosecutor v. Kupreškić et al, IT-95-16-T, Judgment, Trial Chamber (14 January 2000), para. 527.

V. Requirement of meaningful human control over LAWS

In the previous chapter, it has been argued that human control is implicitly required by IHL rules on targeting and maybe even an emerging rule of customary international law. However, the concept is very abstract and has not been analysed in great detail.[312] The following chapter will thus focus on selected elements and factors influencing the quality of meaningful human control and analyse them in more depth. Additionally, it will be demonstrated how various factors are intertwined and influence each other.

The concept of human control as such has been linked to the debate over autonomous weapon systems since the very beginning. The 2012 US Directive on Autonomy in Weapon Systems uses the phrase "human control" in its definition of semi-autonomous weapon systems.[313] In 2010, The International Committee for Robot Arms Control ("ICRAC") convened a meeting of a group of experts who warned of "*the loss of human control over the maintenance of security, the use of lethal force and the conduct of war* [...]".[314]

However, the phrase "*meaningful* human control" was coined by Richard Moyes and was first articulated by the non-governmental organisation Article 36 in a 2013 report on how the United Kingdom is approaching autonomous weapon systems.[315] The proposition sparked lively debate and attracted considerable attention at the CCW GGE meetings. While some prefer the term "appropriate levels of human judgement", this paper will use and focus on the term "meaningful human control", also referred to as "MHC" for the sake of continuity with the ongoing debate. As noted already in the very beginning of the debate, the specific term "meaningful" is less important than the concept behind it.[316]

Notwithstanding the terminology one uses, all concepts target the same issue, which will be the object of the following chapter. First, various definitions will be presented. Second, the

[312] For a recent analysis, see for example: Boutin and Woodcock. Aspects of Realizing (Meaningful) Human Control, supra note 275, p. 12.

[313] US DoD. Directive 3000.09 on Autonomy in Weapon Systems, supra note 2, p. 14.

[314] See ICRAC. Berlin Statement. At: http://icrac.net/statements/ (last accessed 15 April 2022).

[315] See Article 36. Killer Robots: UK Government Policy on Fully Autonomous Weapons. April 2013. At: http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf (last accessed 15 April 2022); Article 36. Structuring Debate on Autonomous Weapons Systems. November 2013. At: http://www.article36.org/wp- content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf (last accessed 15 April 2022).

[316] As an Article 36 briefing paper notes: [T]*here are other terms that refer to the same or similar concepts. These include 'significant', 'appropriate', 'proper', or 'necessary' 'human judgement' or 'human involvement'*. See: Article 36. Key Areas for Debate on Autonomous Weapons Systems. May 2014, p. 2. At: http://www.article36.org/wp- content/uploads/2014/05/A36-CCW-May-2014.pdf (last accessed 15 April 2022).

questions of what is "control" and what should it be exercised over will be explored. Third, aspects determining the "meaningfulness" of how a weapon is controlled will be analysed.

1. Diverging views on MHC

It has been suggested above that there is widespread consensus that a sufficient level of human control must be retained over LAWS. Meaningful human control may be seen as a useful tool enabling to move forward the discussion about what is problematic about autonomy in weapon systems, also because the concept can be considered before any large-scale deployment of LAWS as well as prior to their potential absolute ban.[317] However, even if all States agree on the requirement, an obstacle remains. As the Czech Republic noted, "*the decision to end somebody's life must remain under meaningful human control.* […] *The challenging part is to establish what precisely 'meaningful human control' would entail.*"[318] Without a reasonable definition, the concept risks merely shifting the debate to "what is meaningful?" Indeed, one common criticism of MHC is that it is vague and imprecise.[319] Individual positions differ significantly, authors use different terms and conditions, ultimately making it unclear how MHC is to be operationalised or applied in practice.[320] Without a clear idea how a desirable standard of control should be defined and achieved, it is difficult to make any informed policy decisions with MHC as a basis, whether during weapon development or during targeting procedures.[321] On the other hand, it is perhaps not necessary that the MHC concept be precisely defined in great detail. Many core concepts contained in international humanitarian law instruments are not defined in themselves, such as "unnecessary suffering" and "indiscriminate effects".[322]

[317] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward. 2014, p. 3. At: https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-considering-how-meaningful-human (last accessed 15 April 2022).

[318] Czech Republic. Statement at the CCW Meeting of Experts on Lethal Autonomous Weapons Systems. 2015. At: https://www.mzv.cz/public/29/e/7d/1448252_1299062_CZ_statement_general_debate_LAWS_ver2_1.pdf (last accessed 15 April 2022).

[319] van den Boogaard, J. C., and Roorda, M. P. Autonomous Weapons and Human Control. In: Bartels, R., van den Boogaard, J. C., Ducheine, P. A. L., Pouw, E. and Voetelink, J. *Military Operations and the Notion of Control Under International Law*. Berlin: Springer, 2021, 421-39.

[320] Jensen, E. T. The (Erroneous) Requirement for Human Judgment (and Error) in the Law of Armed Conflict. SSRN Electronic Journal. 2020. 96. 26–57. At: https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2916&context=ils (last accessed 15 April 2022).

[321] Kwik, J. A Practicable Operationalisation of Meaningful Human Control. *Laws.* 2022. 11(43), p. 3. At: https://doi.org/10.3390/laws11030043 (last accessed 15 April 2022).

[322] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, p. 4.

The concept of MHC must take into account the variety of current weapon systems with autonomous features and the ways those are used; however, it must also address the potential and gradually greater autonomy in weapon systems.[323] The scope of weapon systems that should be subjected to the requirement of MHC is a topic of heated debate. Some argue that any definition of meaningful human control that rules out the use of uncontroversial weapon systems, including systems that make civilian casualties less likely, is not practical or likely to be adopted.[324] The current use of weapon systems with automated or autonomous features should be duly considered when defining what meaningful human control is. However, if deficiencies or shortcomings flowing from the human-machine interface are revealed, which impede the meaningfulness of human control being exercised, why should not these be addressed? Why not use the opportunity for reflection that autonomy provides us with to analyse what level of control over weapons we consider sufficient and indeed explore which mistakes might have been made in the past? After all, ruling out a weapon system is the most extreme measure. In case of a disharmony with what is to be considered meaningful human control, the simplest solution is to change the parameters of the human-machine interface.

When attempting to define MHC, the most common approach is to focus on its elements, each of which has the potential to change the level of control exercised over LAWS. The ICRC has been urging States to identify practical elements of human control as a basis for internationally agreed limits on autonomy in weapon systems with a focus on the following:

"*What **level of human supervision, intervention and ability to deactivate** is required during the operation of a weapon that selects and attacks targets without human intervention?*

*What **level of predictability** – in terms of its functioning and the consequences of its use – and reliability – in terms of the likelihood of failure or malfunction – is required?*

*What other **operational constraints** are required for the weapon, in particular on the tasks, targets (e.g. materiel or personnel), environment of use (e.g. unpopulated or populated areas), duration of autonomous operation (i.e. time-constraints) and scope of movement (i.e. constraints in space)?*"[325]

[323] Horowitz and Scharre. Meaningful Human Control in Weapon Systems: A Primer, supra note 293, p. 6.

[324] Scharre and Horowitz. *An Introduction to Autonomy in Weapon Systems*, supra note 14.

[325] ICRC. The Element of Human Control, Working Paper, Convention on Certain Conventional Weapons (CCW) Meeting of High Contracting Parties. CCW/MSP/2018/WP.3. 20 November 2018. At: https://undocs.org/Home/Mobile?FinalSymbol=CCW%2FMSP

This list of criteria is not meant to be exhaustive; it rather directs attention to the most pressing issues that should be addressed when defining meaningful human control. The same or similar issues have been raised by others. Article 36 has argued that meaningful human control requires (1) **predictable**, reliable, and transparent technology; (2) **accurate information** for the user on the outcome sought, the technology, and the context of use; (3) **timely human judgement** and action, and a potential for timely intervention; and (4) **accountability** to a certain standard.[326]

ICRAC suggested three minimum necessary conditions for meaningful human control: (1) a human operator must have **full contextual and situational awareness** of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack; (2) there must be **active cognitive participation** in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of attack, and likely incidental and possible accidental effects of the attack; and (3) there must be a **means for the rapid suspension or abortion** of the attack.[327]

While each set of requirements slightly differs, they highlight the same fundamental issues, and they have several general elements in common when it comes to defining what level of control is "meaningful". A recent study engaged in a thorough literature analysis of papers published between 2013–2021 to determine features common to these proposals. It identifies five core elements: awareness, weaponeering, context control, prediction, and accountability.[328] Here, the elements will be analysed in more detail, categorised into three groups. First, there is a technological element, focusing on predictability, reliability, and the means for suspension. Second, a conditional element highlights the issues of particular tasks, targets, environment, and time restraints. Third, there is a decision-making element, encompassing mainly the nature of the role of the human operator, availability of information, and situational awareness, leading to accountability. However, before focusing on what makes

%2F2018%2FWP.3&Language=E&DeviceType=Desktop&LangRequested=False (last accessed 15 April 2022) (emphasis added).

[326] Article 36. Key elements of meaningful human control, supra note 210, p. 1 (emphasis added).

[327] Sauer, F. ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting, ICRAC INT'L COMM. FOR ROBOT ARMS CONTROL. 14 May 2014. At: https://www.icrac.net/icrac-statement-on-technical-issues-to-the-2014-un-ccw-expert-meeting/ (last accessed 15 April 2022) (emphasis added).

[328] Kwik. A Practicable Operationalisation of Meaningful Human Control, supra note 321.

control meaningful, two preliminary questions should be addressed: over what control should be exercised and at which level it should be done.

2. The requirement of "control"

"Human control" may be understood in a variety of ways. Taking a stringent view, control would mean that a human operator monitors the system and makes all critical decisions. On the opposite side of the spectrum, others may argue that human control can be sufficiently exercised through the design of a system by ensuring that it functions reliably and predictably, even without having a human in the loop (a legal 'compliance by design' approach).[329] Control can also be manifested throughout different stages of weapon development or employment. The question that pertains to the definition of meaningful human control is what should control be exercised over, whether, for example, an individual attack, the weapon system itself, or its critical functions.

2.1 Control over what?

The original idea of Article 36 was to call for MHC over individual attacks. However, various interpretations have emerged since, e.g., meaningful human control over weapon systems, the critical functions of autonomous weapons, the use of force, or the targeting decision-making process.

As shown above,[330] attacks have different aspects and stages in the targeting process. If LAWS take part in some steps, each of these may potentially be subject to varying degrees of human control. The question "why someone or something is targeted" is typically addressed in several stages before the individual attack (at least in preplanned targeting). It is hardly imaginable that LAWS would be carrying out this step without any human engagement. In the first stages of targeting decision-making, abstract objectives and plans are formed. During this stage, the role of LAWS is highly likely to be advisory, which poses no problem to retaining MHC. What is the target of the attack (and who or what will be harmed indirectly) would likely be decided on an abstract level by humans by way of determining a category of possible targets and then by LAWS concretely in each case. The question "how force is used" could technically fall into the scope of decisions already taken by LAWS themselves if they had

[329] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, p. 3.
[330] See Chapter II, section 2.2.2.

multiple methods of conducting an attack at their disposal. The time and place of an attack can be specified by human operators, but if speed is the advantage militaries want to make use of, it is more likely that LAWS would make these decisions.

What then should be subject to meaningful human control? One option is an individual attack. If we considered an attack as the whole targeting process, we would apply the requirement of MHC to a broad scope of individual decisions. Some of those may be taken solely by humans, some entirely by LAWS, and some in cooperation. It would only be the overall process that would be subjected to the scrutiny of meaningfulness of human control. It would probably sufficiently answer the ethical and moral concerns about using lethal force. However, it would not address the issues of ensuring compliance with targeting rules of IHL and, to some extent, the issues of accountability. Human control would be retained, but only in general.

Another option is meaningful human control over the weapon system *per se*. This could mean that every decision the programme makes, and the weapon system executes, is subject to human supervision and control. In this scenario, MHC is undoubtfully retained. However, such an approach would diminish the benefits of the autonomous capabilities. Should LAWS be deployed, it would be because of their efficiency, speed, and other capabilities that overcome humans. Moreover, such a strict level of control is not required even in currently used weapons, not only those with automated features but also traditional weapons and weapon systems. This approach is simply not realistic and would mean a step backwards. Additionally, if AI is employed, it is doubtful that a human could retain control over the decision process of the programme. Even if certain transparency requirements were complied with, it is not realistic that a human operator amid an armed conflict would keep track of and control over every step.

Retaining MHC over the weapon system could also be interpreted as a general requirement of a human operator being aware of the steps of the decision-making process of the programme, coupled with certain restraints that could be introduced (on the tasks, targets, time and place of attacks, etc.). It is undoubtedly a more satisfactory solution for addressing the "accountability gap". A human operator could have overall control over the weapon system, with sufficient knowledge about its functioning. However, this approach does not address the issue of ensuring compliance with the rules on distinction, proportionality, and precautions. Even if the overall control over the weapon system was meaningful, compliance with the IHL rules on targeting requires more attention, as the probability of LAWS being able to act in conformity with these rules is relatively low, at least in the near future.

Another suggestion is to require MHC over the critical functions of autonomous weapons. Obviously, this approach raises the question of which functions are "critical". A possible answer is those which have a direct impact on targeting decisions. While still rather general, this approach has the benefit of being flexible enough to be applied in various situations and with regard to different types of LAWS. It would rule out the weapon systems whose autonomy does not have a direct connection to the possibly lethal decision at the end of the process, such as systems deployed to collect and analyse data or systems with mere advisory functions. On the other hand, it would include systems that create lists of possible targets or conduct the proportionality assessment, both of which directly influence the targeting decisions and the attack's compliance with IHL. This approach is less strict and allows militaries to benefit from the capabilities of computer systems. Nevertheless, it also ensures that a human would monitor the compliance with IHL in every "critical" step and thus could be held responsible if a violation of IHL occurs.

Additionally, when the US DoD issued the first policy document on autonomous weapons, they stated: "*Autonomous and semi-autonomous weapons systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment* ***over the use of force***."[331] An argument can be made that retaining human control over the critical functions of autonomous weapon systems is analogical to control over the use of force since functions are likely to be considered "critical" when they directly impact how force is used.

While there are multiple plausible solutions, it appears that requiring MHC over critical functions of LAWS brings the most benefits. Its downside is undoubtedly the abstract and hard-to-define core term "critical". On the other hand, it allows for flexibility that is needed when all different types of weapons may be designated as autonomous, and their functions would differ substantially. Certain function will be critical in all cases, such as the decision to fire, however, other functions can have a stronger or weaker link to the use of lethal force in particular cases.

[331] US DoD. Directive 3000.09 on Autonomy in Weapon Systems, supra note 2, p. 14.

2.2 Control at which level?

Having explored several possible approaches to what control should be exercised over, the most favourable solution seems to require MHC over critical functions of LAWS. However, given that "control" in military terms includes a variety of processes (such as intelligence collection, context analysis, target identification, a proportionality calculation, the decision to attack),[332] is MHC equally necessary at each stage? The previous analysis shows that the two questions are closely interlinked. If control should be exercised over critical functions, it does not matter during which stage of the whole process this particular function manifests. What matters is the impact the function has on the targeting decision, its compliance with IHL, and the operator's accountability, particularly in cases of grave breaches, where IHL requires prosecution of violations by States.

On the other hand, it must be recognised that the quality and nature of control will differ at each level. A commander determining the rules of engagement at the top of the command chain is exercising a certain kind of control, which, however, does directly impact the final decision. It may also affect the compliance of LAWS with IHL in general, as the rules of engagement will be embodied in the programme. A commander ordering an attack and deploying LAWS in a particular context is exercising another type of control, more closely linked to the obligation to choose appropriate means and methods of warfare. And the individual operator implementing the order is bound to exercise yet another kind of control, more direct in terms of actually supervising the weapon system, ensuring the compliance of the attack with IHL rules on the conduct of hostilities. These differences play out during preplanned targeting. The role and control of a pilot will again be very different during dynamic targeting.

This shows that the requirement of MHC should be flexible enough to cover various types of control over critical functions of LAWS that are exercised throughout the chain of command, irrespective of whether that happens before, during, or after the use of lethal force. Focusing solely on the last link in the chain, the operator of the weapon system, risks overlooking the importance and the impact of the role the prior decisions may have on a weapon system's compliance with IHL, even though those decisions may be more of a tactical nature.

[332] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, p. 4.

3. The qualifier of "meaningful" human control

The previous section has established that the most plausible solution is to require control over critical functions of LAWS and that the way how MHC is to be defined must be flexible enough to encompass various types of control. Further, the qualifier of meaningfulness will be explored. While "meaningful" is the term most often used, some suggested other formulations such as "appropriate", "sufficient", "effective" or "adequate" human control.[333] Other also propose framing the concept as "human judgement" or "human involvement".[334] This paper opts for "meaningful human control" as it is the concept that has gained traction and is used by most authors.

Whether something is "meaningful" is inherently subjective, as individuals give different meanings to the same sets of facts. In the context of LAWS, it might refer to whether there is sufficient time for a human to intervene, exercise judgment, override or terminate an attack.[335] However, the reality is much more complex, and a number of factors influence the quality and nature of control. This paper proposes dividing these factors into three categories. The first category focuses on the technological element, particularly the communication link. The second one reflects the broad range of circumstances and environments in which LAWS can be deployed and how these factors influence the level of control that should be deemed sufficient. The third one is the decision-making element, addressing the human factor in the equation. It focuses on the psychological aspects of humans controlling autonomous weapon systems and how should the human-machine interface be set up.

[333] Chair of the GGE LAWS. Chair's summary of discussion, Agenda item 6(b). 2018. At: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2018)/Summary%2Bof%2Bthe%2Bdiscussions%2Bduring%2BGGE%2Bon%2BLAWS%2BApril%2B2018.pdf (last accessed 15 April 2022); ICRC. Ethics and autonomous weapon systems: An ethical basis for human control? UN Doc CCW/GGE.1/2018/WP.5. 29 March 2018, p. 2. At: https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control (last accessed 15 April 2022).

[334] See e.g.: CCW. Canadian response to the Chair's request for input on potential consensus recommendations. 2021, pp 1-2. At: https://documents.unoda.org/wp-content/uploads/2021/06/Canada_Commentary-on-potential-consensus-recommendations.pdf (last accessed 28 May 2022).

[335] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, p. 3.

3.1 Technological element

The first element that largely influences the quality of human control over LAWS is the technological setting of the weapon system and the design of its programme. It should enable human agents to exercise control through the design of weapon parameters, mainly of the communication interface.

To meaningfully control the operation of a weapon system, the human operator has to be able to understand the information the system is providing and to communicate back the instructions to the system (an upward and downward stream of information). This communication can take place on the battlefield directly but it can also be done remotely from a command centre. The design of the communication interface has to be accommodate depending on which is relevant in the particular environment and context. It has to reflect the training and expertise of the operators, which is indeed a very general requirement pertaining to all weapon systems.

However, there is an issue particular to the context of communication and the deployment of LAWS. The characteristic of autonomy might be highly attractive to the military when communication with a weapon system is not feasible or is lost or jammed.[336] Such could be the case with submarines and space warfare, but also in more common environments, where the connection may be lost due to foreign interference or the lack of signal. The question is, how could MHC be exercised in such contexts?

The first option is to maintain a reliable and secure communication connection throughout the whole LAWS deployment period. This requirement must be tested while carrying out the weapons review. Moreover, fail-safes must be designed in case of failure of the communication link. The programming of the weapon system must ensure that in case of a loss of connection, the weapon systems will either automatically deactivate or, in the case of mobile LAWS, return to the base without carrying on the attack.

The other feasible solution is to rely on the programming of the weapon system when no real-time control can be maintained. Following the analysis of the problems with ensuring compliance with IHL introduced above, this approach carries high risks. If MHC is to be

[336] Ibid.

retained in all situations, reliance on programming can only be envisaged in a very particular set of circumstances.

In all cases, the technological setup must ensure that humans can exercise a sufficient level of control. The following sections will analyse factors influencing the quality of the control exercised, focusing on the design of the weapon system and its programme.

3. 1. 1. MHC and the loop scheme

Preliminarily, it must be mentioned that there are multiple ways how can the system be designed to engage its human operator as well as multiple approaches to describing human-machine cooperation. The debate on LAWS frequently differentiates the levels of autonomy in weapon systems by referring to the loop scheme.[337] The image of the control loop, typically referred to as the OODA loop (orient, observe, decide, act),[338] helps to visualise the engagement of the human operator in specific situations when targets are selected and engaged. Its downside is that it does not focus on earlier phases of the targeting process, e.g., strategic planning.[339]

Meaningful human control is another concept addressing the spectrum of levels of human control, independent of the loop categorisation. The two concepts can be compared to one another to gain more clarity about the nature of human engagement. Some have attempted to translate the loop characteristics into a specific language and provide examples. Sharkey divides levels of human control into five categories, to which the "loop" characteristics can be assigned: (a) humans deliberate about specific targets before initiating an attack (in-the-loop); (b) humans choose from a list of targets suggested by a program (in-the-loop); (c) programs select the calculated targets and needs human approval before the attack (on-the-loop); (d) programs select calculated targets and allocate humans a time-restricted veto before attack (on-the-loop); (e) programs select calculated targets and initiate attacks without human involvement (out-of-the-loop).[340]

[337] For detailed information, see Chapter I, section 2.2 above.

[338] The OODA loop was developed by US Air Force Lieutenant General John Boyd, see Anderson, W. R., Husain, A. and Rosner, M. The OODA Loop: Why Timing Is Everything. Cognitive Times. December 2017. At: https://www.europarl.europa.eu/cmsdata/155280/WendyRAnderson_CognitiveTimes_OODA%20LoopArticle.pdf (last accessed 15 April 2022); Pearson, T. The Ultimate Guide to the OODA Loop. 2017. At: https://taylorpearson.me/ooda-loop/ (last accessed 15 April 2022).

[339] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 19.

[340] Sharkey. Staying in the Loop: Human Supervisory Control of Weapons, supra note 129, pp 34–37.

When comparing the two concepts (MHC and the loop scheme), a suggestion can be made that in-the-loop systems always retain meaningful human control. On the other hand, the out-of-the-loop characteristic could, at first sight, appear as ruling out any possibility of human control. However, this categoric approach would not reflect the realities of even current weapons and would overlook the different circumstances in which LAWS may be used. The middle ground, systems with a human on-the-loop, comprises a variety of possible situations and settings. The breaking down of various situations above illustrates the differences between the loop scheme approach and the meaningful human control approach.

Sharkey's first category introduced above is uncontroversial; human deliberation before an attack ensures that MHC is retained. The second category is unlikely to pose a problem to maintaining meaningful human control either. The third category is more problematic. The positive action required from the human operator ensures a more meaningful role, as requiring approval engages more deliberation. On the other hand, there are issues such as automation bias and over-trust that need to be taken into account.[341]. In this scenario, the meaningfulness of human control will depend on the modalities of the approval process and the settings of the human-machine interface. Therefore, while categorised as on-the-loop, various programming modes can comply with or exclude meaningful human control. The same applies in the fourth category. However, allowing the human operator only to veto an attack is much more likely to fall outside what should be considered meaningful human control (except for very specific circumstances). Keeping a human on-the-loop does not solve the problem of reducing the role of human operators to mere supervision of decisions taken at superhuman speed, while leaving the illusion that the human control requirement is still complied with.[342]

In Sharkey's fifth category, the programme selects and calculates targets and initiates attacks without any human involvement, which places the human operator out of the loop entirely. Even though very unlikely, there can be a limited set of circumstances when this mode of operation could still fulfil the requirement of meaningful human control, as will be explored later.[343]

[341] See section 3.3.2.1. below.
[342] Amoroso and Tamburrini. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues, supra note 129, p. 189.
[343] See section 3.2.3 below.

This brief analysis shows that the two concepts – the loop scheme and MHC – can yield different results when it comes to determining the permissible level of human engagement.

3. 1. 2. Design of the programme

As shown above, the meaningfulness of human control is independent of the loop characteristics. It is, however, certainly influenced by the design of the programme, the technological settings and the way how the system cooperates with its human operators.

A targeting process has multiple stages and the design of the programme may be of relevance in more than one of them, especially concerning the information available to the system. While it may appear that the weapon system is the final link and needs only a part of the data relating to the characteristics of the targets, it has been argued that all the information prior to the last link of the chain influences the targeting decision. It certainly presents a programming challenge, but even the highly abstract data relevant to the mission objectives should be translated into the algorithm. All data possibly connected to the whole operation scheme should be incorporated into the programme in order to prevent mistakes caused by a lack of information that was seemingly unrelated. After all, one of the main benefits of using computer algorithms in weaponry is precisely their ability to process a large amount of data. This point also relates to data from other operations that may be carried out by the same contingent or nearby areas, which may introduce potential clashes and subsequent failures. On the same note, the programming of LAWS should take into account the possibility of enemy/friendly autonomous weapon systems appearing in the area and their interaction.

3. 1. 3. Predictability and arbitrariness

The issue of (un)predictability of the behaviour of LAWS has been already discussed above as one of their technological limitations and reasons why MHC is needed to ensure compliance with IHL.[344] However, it also needs to be addressed as one of the conditions for retaining meaningful human control. A human operator must be able to understand how the system functions and reasonably predict its behaviour to exercise control that fulfils the qualitative requirement of MHC. The advantage of addressing (un)predictability in the framework of MHC is that it allows for desirable flexibility. A weapon system over which a human operator retains control does not have to be perfectly predictable. If LAWS were

[344] See Chapter III, section 1.5.

deployed without human control, there would have to be very high certainty about its behaviour, otherwise the IHL rules on conduct of hostilities cannot be complied with. However, a human operator can monitor and possibly intervene if an unpredictable irregularity appears. This approach is more flexible, as it enables to modify the level of human control depending on the predictability of the behaviour of the weapon system. Should a programme be highly reliable and predictable, a lower degree of oversight and control may be required from human operators.

On the other hand, if a system is using AI and its behaviour demonstrates a high degree of unpredictability and arbitrariness, retaining meaningful human control can make the deployment of such a system still lawful. Though should that be the case, the requirements for the quality of human control will be considerably higher. For example, a simple veto power would not fulfil the criteria of MHC in this case, as this setting of the human-machine interaction would not provide sufficient time for deliberation. This situation would result in a high risk of violating IHL, thanks to the combination of an unsuitable form of human intervention and the unpredictability of the weapon system in question.

In other words, if critical functions of a weapon system should be under meaningful control, the system needs to behave in predictable ways in the environment in which it is deployed. This will, in large, depend on the functions of the weapon and the space and time in which it is used. The role of the environment is crucial as a particular weapon system may be predictable in one environment (for example, in the sea and air environments) but may not necessarily behave in the same manner in another (such as a city with a wide range of objects and stimuli).[345] This effect also needs to be taken into account in the weapons review and testing. A given action may be predictable in one set of circumstances, but in interaction with a different environment, the results of that action may be arbitrary. Should that be the case, every such weapon system should be accompanied by instructions on its use and the environment in which it is suitable to be deployed. That may be a very optimistic idea, as there is no way to ensure that a particular weapon system is not used in a situation for which it was not cleared or that the nature of the environment will not change throughout the deployment period of LAWS. However, this is not a novel problem as the same is true for

[345] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, pp 5-6.

every weapon that may only be used in certain circumstances and States must have procedures on how to deal with this issue.

3.2 Conditional element

The second element influencing the quality of human control are the conditions on the use of LAWS, such as restrictions on tasks, targets (e.g., materiel or personnel), the environment of use (e.g., unpopulated or populated areas), duration of autonomous operation (i.e., time-constraints) and scope of movement.

3. 2. 1. Tasks

The tasks entrusted to the weapon system belong among the conditions on the use of LAWS contributing to the meaningfulness of human control. There are various tasks at which computers already exceed humans, algorithmic calculations being one of them. An example of that can be the recent report of AlphaGo zero which can learn without human intervention and play at a super-human level.[346] Computers' superior data processing capacity is precisely the reason why they are used for target identification via pattern recognition in vast amounts of data, both in targeting and in surveillance more generally.[347]

On the other hand, while a computer is designed to make faster calculus, algorithms, and other kinds of instrumental tasks, it cannot take advantage of anything that it does.[348] Humans show better performance in tasks involving deliberate and qualitative judgement. One option to overcome human abilities might be to design a cognitive computer system. However, even if this proved possible, this kind of computer could reach and overcome some, but not all, human capabilities.[349] Therefore, this division of tasks should be made use of when deploying LAWS. Let us consider the example of a proportionality analysis. Humans should be entrusted with taking decisions requiring deliberate and qualitative judgement, such as determining the military advantage anticipated. On the other hand, computer systems could aid with the analysis of expected casualties and suggest the best circumstances for carrying out a particular attack (e.g., time and location). If approved by a human operator, the weapon

[346] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. Mastering the game of Go without human knowledge. *Nature*. 2017. 550. 354–359. doi: 10.1038/nature24270.

[347] Bode and Huelss. The Future of Remote Warfare?, supra note 248, p. 221.

[348] Signorelli. Can Computers Become Conscious and Overcome Humans?, supra note 86, p. 7.

[349] Ibid, p. 2.

system could autonomously carry out the preplanned attack at the time most convenient, for example, when no civilians are present.

Moreover, should MHC be required only over critical functions of the weapon system, there will be certain functions that might not need human oversight at all. This could be, for example, various kinds of data collection or analysis. On the other hand, certain tasks should not be delegated to LAWS either at all or only with a rigorous human control retained, such as precisely determining military advantage or attributing values to human casualties, as these, at least for the time being, cannot be translated to a computer programme. One can also think of launching an operation that would trigger an international armed conflict. If subscribes to the first-shot theory, this would not hypothetically be an unlikely scenario, but certainly one with grave consequences. These types of decisions should certainly remain under very strict human control. Hence, the division of tasks should consider the kind of skills necessary for its lawful and efficient execution.

3. 2. 2. Timespan of deployment

Limiting the deployment period of a weapon system is an additional way of influencing the level of human control over its critical functions. This is not a newly invented method; it has been used in the past with other types of weapons; precedents include time limitations on the active life of unanchored sea mines contained in the 1907 Hague Convention (VIII) and those on the active life of remotely delivered anti-personnel mines in Amended Protocol 2 of the CCW.[350]

A suggestion might be considered that the requirements on MHC should differ based on the duration of the deployment of LAWS. Prior to any exercise of autonomous functions of a weapon system, a lower degree of human control would likely be sufficient, if it indeed would be required at all. Depending on the expected time frame for mission accomplishment, different levels of human control may be necessary. The reasons behind this approach mostly flow from the technological limitations of the computer system. It is more feasible to programme a reliable and predictable weapon system when its deployment is expected to last only a very short period of time. It also makes it easier to predict future developments in the area and estimate any incidental civilian casualties. Moreover, any possible brittleness of the

[350] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, p. 5.

computer systems or malfunctions are much less likely to occur. On the contrary, should LAWS be deployed for a substantial period, the risks are certainly higher, and stricter human control must be exercised to fulfil the standard of meaningfulness. That is the only way to react to changes in the area that may happen throughout the time.

Once the expected time frame expires, or the mission's objective has been achieved, the system should be required to return to a mode directly controlled by humans or be redeployed. The time limitation parameter is closely linked to other factors. The UK elaborates that control can take the form of "*restricting the type of target and task, temporal and spatial constraints, constraining weapon effects, allowing for deactivation and fail-safe mechanisms where appropriate, and controlling the environment to exclude civilians or civilian objects*".[351] Specific missions may use LAWS for a long time in particular environments where communications are limited, difficult or impossible (the marine areas or the space). More attention in these circumstances has to be paid to ensuring the limitations are adequate. Suppose communication after the end of the mission is impossible. In that case, there is no way to verify that a weapons system has actually ended its autonomous functioning, or to gain direct human control over it again.[352] Some fail-safe mechanism should be embodied in the programme to ensure that LAWS in these circumstances would not carry on exercising their autonomous function with human control being exercised neither in real-time nor through prior programming since the time span of the expected deployment has expired.

3. 2. 3. Environment and its effect on MHC

The environment in which a particular autonomous weapon system is deployed determines what it senses, and thereby how it acts upon it. Simultaneously, decisions made by LAWS have tangible effects on, and alter, the environment. Ideally, from a military perspective, this impact would overlap with the desired effects of that operation.[353] The environment thus plays a crucial role when considering the obligations stemming from IHL. The nature of modern warfare and the unfortunate fact that hostilities often take place in urban areas always have been among the principal arguments of the proponents of a complete ban of LAWS. And one

[351] United Kingdom. Expert Paper: The Human Role in Autonomous Warfare, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System. Agenda Item 5. Technical Report CCW/GGE.1/2020/WP.6. Geneva: 21–25 September 2020 and 2–6 November 2020. At: https://documents.unoda.org/wp-content/uploads/2020/07/CCW_GGE1_2020_WP_7-ADVANCE.pdf (last accessed 28 May 2022).

[352] Ibid.

[353] Kwik. A Practicable Operationalisation of Meaningful Human Control, supra note 321, p. 8.

has to be sympathetic with these arguments, as deployment of autonomous weapon systems in these incredibly complex environments with civilians more often than not outnumbering combatants raises significant concerns. With the current state of technology, it is virtually impossible to programme LAWS that would be able to autonomously move around an urban area, distinguish between legitimate and unlawful targets, take all precautions that would be feasible for human soldiers, and apply the proportionality rule.

A UNIDIR publication suggests that restricting the use of certain autonomous weapon systems to specific environments could prove a reasonable mechanism for ensuring control over them, particularly in relation to concerns of ensuring predictability in diverse environments.[354] Indeed, the issues of predictability, timespan, and environment are closely linked to one another. The quality of human control should be assessed by looking at the overall interplay of these factors, on a case-by-case basis. A particular weapon system can prove predictable and reliable in a specific environment for a certain period of time. However, the assessment of its functions can radically change if one of the elements is altered.

The most crucial issues relating to the deployment of LAWS in the IHL context arise due to their insufficient capabilities to comply with the rules of distinction, proportionality, and precautions. Those must be abided by every party to any armed conflict. However, these rules would not necessarily pose such a challenge for LAWS in certain situations. As Sharkey remarks, currently used "automatic target recognition methods" do work in low cluttered environments and with clearly recognisable military objects such as tanks in the desert and ships at sea.[355] In easily navigable and simple environments, the level of human control over critical functions of LAWS may be lower, depending on the probability that an unpredictable event will occur, which would turn the situation into a very complex one. For example, when it comes to deep sea areas, the likelihood of a civilian submarine appearing out of nowhere is rather low. Of course, it is neither possible nor desirable to draft a list of "simple" areas in which any autonomous weapon system could be deployed. Even a low cluttered environment can change its nature throughout time, which is one of the reasons why a certain level of human control should be retained even in these situations. A civilian plane may still appear unexpectedly, flying over a desert, because it had to change its course. It is, however, not a lawful target and if the weapon system was not programmed to recognise civilian planes, its

[354] Ibid.

[355] Sharkey. Autonomous weapons and human supervisory control, supra note 38, p. 29.

human operator must intervene. While this solution appears practical, one has to ask whether the geographic restrictions would be realistic or effective. Throughout history, weapon systems designed for one type of use have been employed for other uses and in new contexts based on need and innovation.[356] The same point has to be made as above, every autonomous weapon system should be designed together with instructions and restrictions on its use.

An even more problematic question arises, whether retaining meaningful human control over the critical functions of LAWS could enable their deployment even in highly cluttered environments. One can imagine a very wide range of their possible use even in urban warfare, some raising more concerns than others. Apart from situations of offence that so often come to mind, Waxman and Anderson provide an example where LAWS would serve "*efforts to protect peacekeepers facing the threat of snipers or ambush in an urban environment, or infantry teams working to secure a town. Small mobile robots with weapons could act as roving scouts for the human soldiers, with "intermediate" automation - the robot might be pre-programmed to look for certain enemy weapon signatures and to bring the threat to the attention of a human operator, who then decides whether or not to pull the trigger*."[357] Here the possible answer depends on many more factors than just the timespan of deployment and predictability, such as the purpose of their use, targets they might engage, or the level of human oversight, to name just a few. These factors will be further explored.

One further point should be elaborated upon when discussing the role of the environment. It has been suggested to differentiate between stationary versus mobile roles of LAWS.[358] It could be argued that a lower degree of human control would be acceptable for stationary systems (e.g., those defending a particular location against specific types of threats). In contrast, human control over critical functions of mobile LAWS would have to be stricter. This approach again reflects the reality of the environment in which the weapon system would be deployed. It is much easier to achieve the reliability of object recognition and compliance with the rule of distinction when the environment around the weapon system does not change when it is stationary. Therefore, commanders and operators can rely on the weapon's programming with more certainty, as the likelihood of unpredictable events or unrecognisable

[356] UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, supra note 317, p. 5.

[357] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 6.

[358] Ibid.

objects appearing is lower. However, the same requirements of IHL apply to both types of LAWS, and the overall goal is to achieve a carefully balanced human-machine cooperation that would ensure compliance with the rules on distinction and proportionality. The example of the Patriot system analysed above[359] proves that even when it comes to stationary systems, over-reliance on the capabilities of the computer system can lead to failures, caused also by the inappropriate setup of the human-machine interface. Therefore, even though the required level of human control can differ for mobile and stationary LAWS, the requirement of MHC still stands.

3. 2. 4. Targets

Targets engaged represent another factor influencing the quality of human control over critical functions of LAWS. A distinction is usually made between systems intended for anti-material targets and those used against combatants or with foreseeable consequences for civilians.

A suggestion was made that the sufficient degree of human control would differ for systems that target anti-materiel systems and for those designed for targeting persons. Again, as with restrictions on tasks or the area of deployment, is it realistic to expect that weapon systems will only be used for the purpose for which they are designed? That is precisely something that the obligations under Article 36 of Additional Protocol I aim to address. Weapons may be cleared for deployment only in particular circumstances.

In some cases, complying with target restrictions would not pose a problem, as, for example, a stationary defence system possibly cannot start moving around and attack civilians. However, the restriction on targets gains crucial importance if we consider mobile LAWS such as drones. Drones could be deployed in an autonomous mode in very near future (if not already). Claiming that they will be allowed to attack humans based solely on object recognition is farfetched, hopefully. However, some suggest that mini-drones could be deployed to search for a particular person and target them based on facial recognition.[360] Putting the other legal issues pertaining to this idea aside, an autonomous drone like that could possibly comply with the rule of distinction. It would largely depend on the preciseness

[359] See Chapter IV, section 1.1 above.

[360] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 7.

of the facial recognition system, its susceptibility to foreign interference, etc. Those factors will influence the level of human control required, but the need for retaining accountability still needs to be borne in mind.

Finally, possible secondary effects of deployment of LAWS also have to be considered. Even anti-materiel weapon systems can severely impact civilians when deployed in urban areas. This only shows how all the factors influencing the compliance of LAWS with IHL and the level of human control required are closely interlinked. In each case, multiple factors have to be taken into account to determine how to retain human control that will still be meaningful and ensure that IHL is not violated, and that accountability is secured.

3.3 Decision-making element

The third element of control defines acceptable forms of human-machine interaction through ensuring appropriate levels of human supervision, intervention and ability to deactivate, as well as considering various psychological factors inherent to the human mind and behaviour.

3. 3. 1. Ability to deactivate

The definition of meaningful human control proposed by the ICRAC includes a provision that a commander must have "*full contextual and situational awareness of the target area and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack.* […] [T]*here must be a means for the rapid suspension or abortion of the attack.*"[361] This provision has been criticised by Scharre and Horowitz for its lack of reflection on the realities of warfare, articulating an idealized version of human control. According to them, even the catapult could serve as an example of a weapon's employment where human lack perfect, real-time situational awareness of the target area. The essence of a projectile weapon is the inability to suspend and abort the attack after launch and only with some advanced weapons do commanders nowadays have the ability to retarget or abort a projectile in flight.[362] However, there seems to be a misunderstanding of what is meant by the ability to deactivate autonomous weapon systems in the context of defining meaningful human control. It should not be interpreted as requiring redirecting a launched missile (even though one can imagine that being viable with the use of advanced technology), but rather as

[361] Sauer. ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting, supra note 327.

[362] Horowitz and Scharre. Meaningful Human Control in Weapon Systems: A Primer, supra note 293, p. 9.

being able to suspend the operation of LAWS deployed over a more extended period of time or prior to launching an attack against a particular target, if relevant.

When considering the (in)ability to deactivate, there is a close connection to the time limitation of deployment of LAWS. Once a weapon system is activated and operating in an autonomous mode for a longer period, its human operators should retain the ability to deactivate the system, should that be necessary. A problem lies in the fact that the communications link between humans and the weapon system could be jammed or hacked. It has been suggested that the solution might be to "*reduce the vulnerability of the communications link by severing it - making the robot dependent upon executing its own programming, or even rendering it genuinely autonomous*."[363] This is mainly a technological limitation, which directly impacts the weapon system's compliance with IHL. Either it has to be ensured that the communication link remains secured or a possible solution could be to programme LAWS to return to the base or even self-destruct if the return is not possible.

If a weapon system carries out one action and goes back to a mode of direct human control, a different question arises. Should there always be a possibility for a human operator to be able to cancel an action taken by the system? Many have argued that speed is the crucial advantage of computerised weapons, for example, in cyber warfare or in a conflict where both sides employ LAWS. In line with what has been said previously, the answer to this question has to be given on a case-by-case basis. A suggestion can be made to differentiate primarily between the attack in defence and offence. Should LAWS be deployed in defence against attacks conducted in a way that humans are unable to react to, one could argue that the speed of the defensive action has to match the offence, and therefore no ability to suspend a defensive move should be required. On the other hand, the ability to suspend an attack planned by LAWS should be required when it comes to offence, as there is no prior attack the speed of which should be matched.

3. 3. 2. Operator factors

While technological capabilities of LAWS and the environment play a crucial role, there is a third important factor in the equation - the human operator. Control over critical functions of weapon system can only be meaningful if various psychological aspects pertaining to human

[363] Anderson and Waxman. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*, supra note 25, p. 7.

collaboration with machines are taken into account. An analysis has shown that "human error" appears as a prominent attribution of responsibility in the cases where air defence systems have destroyed civilian airplanes, however, the reality is much more complex.[364] Many of the problems that emerge out of the interaction of humans and machines arise from inappropriate expectations and are rooted in misperceptions or incomplete understanding of the human-machine decision-making system.[365]

Humans bear responsibility for LAWS, so it seems logical that the way how the interface is designed has to reflect human needs, make the most of human capabilities, and pursue the most efficient human-machine cooperation. The following sections will introduce several issues or requirements that human operators are facing when controlling LAWS and suggest how the standard of MHC can be achieved.

3. 3. 2. 1. Over-trust or automation bias

The concept of over-trust,[366] also known as automation bias, or "automation complacency",[367] refers to human operators being overly confident about the reliability of automated and autonomous systems and the accuracy of their outputs. This manifests in a "*psychological state characterized by a low level of suspicion*".[368] Automation bias occurs when humans fail to notice problems because a computer system fails to detect them (an omission error) or when they inappropriately follow a system's decision (a commission error).[369] Automation bias was, for example, among the issues identified in the Patriot case as one of the causes of the erroneous shooting down of a friendly aircraft.[370] In that case, over-trust, which is natural to human behaviour, was amplified by instructions and information received, promoting trust

[364] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 41.

[365] Mosier, K. L. and Skitka, L. J. Human decision makers and automated decision aids: Made for each other? In: Parasuraman, R. and Mouloua, M. *Automation and human performance: Theory and applications*, CRC Press: 1996, 201–220, p. 202.

[366] Boulanin, V., Davison, N., Goussac N. and Peldán Carlsson M. *Limits of Autonomy in Weapon Systems: Identifying Practical Elements of Human Control*. Stockholm International Peace Research Institute and International Committee of the Red Cross. 2020, p. 19. At: https://www.icrc.org/en/document/limits-autonomous-weapons (last accessed 15 April 2022).

[367] Parasuraman R. and Manzey, D. H. Complacency and Bias in Human Use of Automation: An Attentional Integration. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2010. 52(3). 381–410. At: https://doi.org/10.1177/0018720810376055 (last accessed 15 April 2022).

[368] Wiener, E. L. *Complacency: Is the Term Useful for Air Safety*. Proceedings of the 26th Corporate Aviation Safety Seminar (Denver, CO: Flight Safety Foundation, Inc.). 1981, p. 117.

[369] Mosier and Skitka. Human decision makers and automated decision aids: Made for each other?, supra note 365, p. 206.

[370] See Chapter IV, section 1.3 above.

in the weapon system's efficiency. A review of the Patriot concluded that the operators were indeed trained to trust the system's software, perhaps uncritically.

Automation bias is not a new concept associated only with LAWS. It occurs commonly when humans collaborate with computers. A study conducted on an interface designed for supervision of an in-flight GPS-guided Tomahawk missile tasked the operators with deciding which missile would be the correct one to redirect to a time-critical emergent target.[371] The impact on the speed and accuracy of decision making was tested for two different methods: (1) the computer provided the operator with ranked recommendations, including the most "optimal" missile given the situation; and (2) the computer filtered all missiles that were not candidates because of physical restraints and the operator had to decide which missile will be fired. When the computer recommendations were correct, Type 1 operators made significantly faster decisions overall, and their accuracy was equal to those of the slower type 2 operators. However, when the computer recommendations were wrong, Type 1 operators had a significantly decreased accuracy.[372] The study proves the existence and impact of automation bias; operators tended to accept the computer recommendations without seeking any disconfirming evidence.

Over-trust is likely inherent to any human-machine interaction. And to a certain extent, human operators need to trust the weapon system they are using. They need to be confident in using it and not doubt and verify every calculation it makes. Otherwise, it would make no sense to use LAWS in the first place. However, this human trait to be biased in assessing the actual efficiency of the system has to be taken into account, as it presents a significant challenge to retaining meaningful human control over critical functions of LAWS. It can be targeted through rhetoric and training. It should inspire trust in the weapon system to the extent that it is realistic and based on information learned through testing. However, as in the Patriot case, it should not happen that information about any possible software deficiencies will be withheld or it will be claimed that a technological fix is just around the corner when it is not the case. The over-trust issue can also be targeted indirectly through other factors that influence how the operator perceives the weapon system they are using, such as training.

[371] Cummings, M. L. *Automation bias in intelligent time critical decisions support systems*. American Institute of Aeronautics and Astronautics Third Intelligent Systems Conference, Chicago. 2004.
[372] Ibid.

3. 3. 2. 2. Expertise and training

Proper training of the operators must enable them to recognise the system's strengths and weaknesses, as well as make them aware of their role in operating the system. It is crucial not only for fighting over-trust in LAWS but also for maintaining meaningful human control in general. It is one of the requirements that has been called for the most often. It also helps to combat the automation bias.

Coming back to the Patriot case, it is claimed that throughout the year between the two Gulf Wars, the US Army had become so confident of the Patriot system's automatic mode that it even de-skilled its operators, "*reduc*[ing] *the experience level of their operating crews* [and] *the amount of training provided to individual operators and crews*".[373] In the end, the experience level of the Patriot crew involved in the shooting down of the friendly jet was very poor. Some suggest that "*the person who made the call* [...] *was a twenty-two-year old second lieutenant fresh out of training*".[374] It is indisputable that the training and expertise of human operators strongly influence how meaningfully they exercise their control over the critical functions of LAWS. Operating these complex systems must be done competently, and this can only be achieved by providing human operators with adequate information.

3. 3. 2. 3. System and situational understanding

Most definitions of MHC call for a certain level of system and situational understanding. Undoubtedly, the operators need to understand the way how the weapon system they are using functions. The question here is, what level of expertise should be sufficient. It would be unrealistic to require an operator to understand every technical detail of the programme. On the other hand, meaningful human control can only be exercised if the operator has a clear idea of how the system carries out the functions that are deemed critical. For example, whether it is programmed to engage a particular category of targets or has more "freedom" in selecting objects it will engage. The operators should be acquainted with the weapon system's review and testing results, particularly its predictability and precision-recall trade-off. In other words, they need to know how reliable and precise the system is to be able to judge how much trust they can put in the system's analysis and decisions.

[373] Hawley. Patriot Wars: Automation and the Patriot Air and Missile Defense System, supra note 257, p. 8.

[374] Scharre, P. Army of None: *Autonomous Weapons and the Future of War*. New York and London: W. W. Norton, 2018, p. 166.

Bode and Huelss argue that the "[t]*he inexplicability of algorithms makes it harder for any human operator, even if provided the power to intervene, to question the data provided by the weapon system's programme as the basis of targeting and engagement decisions*."[375] This conclusion definitely has some merit, all the more when it comes to machine learning and AI, which increases the unpredictability of the system and the complexity of the processes. However, soldiers do not have to know all the technical details behind the weapon they are operating. First, it is essential to understand the principle on which the weapon is operating and how reliable its performance is, including the system's strengths and weaknesses. Expertise is required to recognise areas in which the particular weapon system performs with certain deficiencies, and therefore, be aware that control over those functions should be stricter. But if the system carries out certain tasks reliably and efficiently, human oversight is not as paramount. The overarching goal is to enable trained human operators to have a clear understanding of how the weapon will function in certain environments and its limitations to use it appropriately.[376]

Furthermore, the operators must have a situational understanding to be able to conduct their role meaningfully. Here again the question of applicable standards arises. The ICRAC's definition of meaningful human control requires human commander to have "*full contextual and situational awareness of the target area*".[377] This strict standard has been criticised by many, pointing out that if these minimum requirements applied to all attacks, many currently used weapons would be rendered unlawful, which could have detrimental effects on both soldiers and civilians.[378] Humans have not enjoyed perfect, real-time situational awareness of the target area even when using a catapult.[379] According to these opinions, this minimum standard is not consistent with how weapons are actually being used. Crootof even goes as far as to claim that "[n]*ot only does the ICARC's definition articulate an idealized version of human control divorced from the reality of warfare, it actually threatens to undermine fundamental humanitarian norms governing targeting*."[380]

Horowitz and Scharre argue for Article 36's standard of "adequate" information being more appropriate while recognising that this only shifts the debate to the question of how much

[375] Bode and Huelss. The Future of Remote Warfare?, supra note 248, p. 222.
[376] Horowitz and Scharre. Meaningful Human Control in Weapon Systems: A Primer, supra note 293, p. 13.
[377] Sauer. ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting, supra note 327.
[378] Crootof. A Meaningful Floor for Meaningful Human Control, supra note 293, p. 61.
[379] Horowitz and Scharre. Meaningful Human Control in Weapon Systems: A Primer, supra note 293, p. 9.
[380] Crootof. A Meaningful Floor for Meaningful Human Control, supra note 293, p. 61.

information is adequate. In their opinion, it should be "*enough information about the target, the weapon, and the context for engagement for the person to make an informed decision about the lawfulness of their action*".[381] Sharkey disagrees and argues that "[i]*f one of the reasons for using advanced technology to apply violent force is genuinely to reduce or eliminate harm to civilians, to others hors de combat and to civilian infrastructure, then striving for full contextual and situational awareness at the time of attack is a way forward*".[382] In his opinion, the purpose of defining meaningful human control is not to overwrite rules on the use of existing weapons but rather to use technological developments as a way how to upgrade our sensibility to civilian harm. Modern technology could also be employed to enable commanders and operators to actively participate during attacks rather than simply plan them, for example, thanks to the use of advanced camera systems to view targets and verify their legitimacy. While the high standard of full contextual and situational awareness is undoubtedly something that we should strive for, the goal is to define a concept of human control that is functional. Since it has been argued that meaningful human control is implicit in IHL, it should be applied consistently to all types of weapons in use. This does not mean that a higher standard could not be agreed upon if new rules should be drafted specifically for LAWS. However, it is more practical and reasonable to pose realistic requirements that all States (and possibly non-state armed groups) are willing and able to comply with than to have idealistic standards that do not reflect reality.

The standard of contextual and situational awareness allowing to take informed decisions also relates to the discussion about ensuring responsibility for the actions of LAWS that pose fair requirements on human operators. Meaningful human control is one of the means by which accountability can be secured. Still, it can only be exercised if the human operator has an overall picture of the situation, as well as the functioning of the weapon system. This does not mean that each human operator involved in the decision-making chain needs exhausting information about the target, the weapon, the environment, and the context for engagement. The US emphasised that the commander deploying LAWS must be aware of the "*system performance, informed by extensive weapons testing as well as operational experience*".[383]

[381] Horowitz and Scharre. Meaningful Human Control in Weapon Systems: A Primer, supra note 293, p. 13.

[382] Sharkey. Staying in the Loop: Human Supervisory Control of Weapons, supra note 129, p. 29.

[383] United States of America. Intervention on Appropriate Levels of Human Judgment over the Use of Force delivered by John Cherry. Paper presented at Technical Report Convention on CCW, GGE on LAWS. Geneva, 15 November 2017. At: https://geneva.usmission.gov/2017/11/16/u-s-statement-at-ccw-gge-meeting-intervention-on-appropriate-levels-of-human-judgment-over-the-use-of-force/ (last accessed 28 May 2022).

But it does mean that an overall, reasonable understanding of these factors should be aimed for. This can be ensured through proper training and adapting the role of the human operators so that they genuinely retain control over the functions of LAWS and can be held accountable in a fair way, not just as a "scapegoat". There is hardly anything meaningful about human control that would consist of simply executing decisions based on indications from a computer if those are not accessible to human reasoning due to the "black-boxed" nature of algorithmic processing.[384]

3. 3. 2. 4. The role of the human operator

Another crucial factor is the role of the human operator of LAWS. Even highly trained personnel with enough expertise will struggle if their role is not adjusted to the realities of warfare and human capabilities. When humans fail at human-computer tasks, it can simply mean that they are being asked to perform in a mode of operation that is not well suited to human psychology.[385] The question to be asked here is what is the nature of their engagement that would enable them to carry out their task in controlling LAWS meaningfully.

For example, it has been suggested that "[t]*he S-400 Triumf, a Russian-made air defence system, can reportedly track more than 300 targets and engage with more than 36 targets simultaneously*".[386] Is it possible for a human operator to supervise the operation of such systems meaningfully? Granted, a whole team can cooperate to ensure a sufficient level of human control, but this would diminish the benefit the militaries are seeking by deploying LAWS – the potential for reduced operating costs and personnel requirements. It is indeed considerably difficult to balance striving for higher efficiency and maintaining meaningful human control.

Bode and Watts point out that the recent trend has been delegating a broader range of tasks to machines, which has changed the human operators' role in the operation of weapon systems from active control to passive supervision. Consequently, human agents are relegated to minimal but impossibly complex roles.[387] Notably, this setting of the human-machine interface is not in line with psychological and neuroscientific research on how humans pay attention. An influential theory put forward by Desimone and Duncan, the "biased

[384] Bode and Huelss. The Future of Remote Warfare?, supra note 248, p. 223.

[385] Sharkey. Staying in the Loop: Human Supervisory Control of Weapons, supra note 129, p. 29.

[386] Boulanin and Verbruggen. *Mapping the Development of Autonomy in Weapons Systems*, supra note 15, p. 37.

[387] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 3.

competition" theory,[388] characterizes attention as a signal competition within the brain. "*Signals compete in order to be more deeply processed and ultimately to influence and guide behavior. This signal competition emerges at the earliest stages of processing in the nervous system and is present at every stage*."[389] Different factors can influence or bias the outcome of this competition; especially intense or salient stimuli can grab attention more easily. The outcome of this signal competition can be slanted in a goal-directed manner based on the demands of the current task. Some signals can be boosted, and irrelevant signals can be suppressed.[390] Nevertheless, focusing on salient cues may result in misinterpretation or in a lack of attention to less obvious but equally important information.[391]

Studies have also explored various forms of attention and their behavioural consequences. By giving subjects repetitive tasks that require a level of sustained attention, researchers have observed extended periods of poor performance in drowsy patients.[392] Yet, there are ways in which tasks can be made more engaging that can lead to higher performance, such as increasing the promise of reward for performing the task, adding novelty or irregularity, or introducing stress.[393] Therefore, general attention appears to be limited in the case of a mundane or insufficiently rewarding task but can be called upon for more promising or interesting work.[394]

Applying these conclusions to the problem of the human role in controlling LAWS, we can see that operators are likely to perform poorly when given mundane tasks, such as simply vetoing or approving decisions based on indications from a computer. This effect will only be amplified when the human operator uncritically (over)trusts the system and therefore has no stimuli to engage in deliberate reasoning. After a certain amount of time of monotonous

[388] Desimone, R., and Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 1995. 18. 193–222. doi: 10.1146/annurev.ne.18.030195.001205.

[389] Graziano M. S. A. and Webb T. W. The attention schema theory: a mechanistic account of subjective awareness. *Frontiers in Psychology*. 2015(6), p. 3. At: https://www.frontiersin.org/article/10.3389/fpsyg.2015.00500 (last accessed 15 April 2022).

[390] Ibid, p. 7.

[391] Mosier and Skitka. Human decision makers and automated decision aids: Made for each other?, supra note 365, p. 204.

[392] Grace W. Lindsay, Attention in Psychology, Neuroscience, and Machine Learning, Frontiers in Computational Neuroscience, vol. 14, 2020, p. 2, at: https://www.frontiersin.org/article/10.3389/fncom.2020.00029 (last accessed 15 April 2022).

[393] Oken, B. S., Salinsky, M. C. and Elsas, S. Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin. Neurophysiol. 2006(117). 1885–1901. doi: 10.1016/j.clinph.2006.01.017.

[394] Lindsay, G. W. Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*. 2020(14), p. 2. At: https://www.frontiersin.org/article/10.3389/fncom.2020.00029 (last accessed 15 April 2022).

supervising the operation of a programme, re-appearing signals no longer attract attention. It has been observed already decades ago that highly automated environments, such as glass cockpits or nuclear power plants, create complacency, boredom, and poor monitoring behaviour.[395] Notably, this effect will only be magnified if the operator lacks an understanding of how the machine functions. Consequently, it is impossible for a human to stay focused and exercise their control over LAWS meaningfully. Many studies of vigilance have proven that the ability to monitor a system for the occurrence of infrequent, unpredictable events typically declines over time.[396] Additionally, the human-machine interface must be designed in a way so that all the important pieces of information act as salient cues. It is of crucial importance that the implications of psychological research into attention and awareness are considered when determining human-machine cooperation.

3. 3. 2. 5. Time for deliberation

Last but certainly not least, allowing sufficient time for deliberation is a crucial step towards achieving meaningful human control. As has been previously discussed, computers are indeed more efficient at some tasks than humans, such as calculating numbers, searching large datasets, responding quickly to control tasks, performing repetitive routine tasks, or conducting deductive reasoning. On the other hand, humans overcome computers in deliberative reasoning, perceiving novel patterns, metacognition, applying diverse experiences to novel tasks exercise, meaningful judgment, or reasoning inductively.[397] The control of weapon systems through computer programs requires the human and machine to operate together, and it should be done in a way that optimises the division of the tasks. It is important to realise that even if tasks are adequately divided between humans and systems, the strengths and limitations of each human-machine interface design depend on how much it inhibits the possibility of human reasoning.[398] Designing an interface to a weapons system that does not take human capabilities into account not only infringes upon the quality of human control but also diminishes the benefits of deploying LAWS.

[395] Parasuraman, R. Human-computer monitoring. *Human Factors*. 1987. 29(6), 695-706.; Chambers, A.B. and Nagel, D.C. Pilots of the future: Humans or computer? *Communications of the ACM*. 1985(28). 1187 - 1199.

[396] For examples of such studies, see: Mosier and Skitka. Human decision makers and automated decision aids: Made for each other?, supra note 365, p. 209.

[397] Cummings, M.L. *Automation Bias in Intelligent Time Critical Decisions Support Systems*. American Institute of Aeronautics and Astronautics. AIAA 3rd Intelligent Systems Conference Chicago. 2004. At: https://arc.aiaa.org/doi/10.2514/6.2004-6313 (last accessed 15 April 2022).

[398] Sharkey. Staying in the Loop: Human Supervisory Control of Weapons, supra note 129, p. 27.

Human psychology distinguishes between automatic and controlled processes.[399] Automatic processing refers to fast responses that are always cued automatically, for example, taking cover when gunshots are heard. Controlled processing indicates slower deliberative processes when making a thoughtful decision, such as conducting the proportionality analysis. The deliberative processes always follow the automatic ones and are thus slower. They require attention and free memory space. Automatic processes dominate if a distraction takes human attention away or requires memory resources. The advantage of automatic decision processes is that they can be trained through repetition and used any time routine decisions have to be made rapidly for predictable events.[400] This process can be replicated and embedded in computer programmes, and these tasks can be delegated to LAWS. Nevertheless, Kahneman illustrates certain properties of automatic reasoning that would be problematic if supervision over critical functions of LAWS was executed solely based upon automatic reasoning: it (1) neglects ambiguity and suppresses doubt; (2) infers and invents causes and intentions; (3) is biased to believe and confirm; and (4) focuses on existing evidence and ignores absent evidence.[401] These properties show that if automatic reasoning were used to exercise control over lethal targeting decisions, there would be severe deficiencies in how meaningful the control would be. Situations when there is contradictory information about target legitimacy would be particularly problematic. Contradictory evidence could remain unseen or be disbelieved. Humans would tend to display automation bias even on a bigger scale. Unfortunately, that is frequently the reality even with currently used weapon systems. A recent analysis has concluded that the current engagement window of air defence systems provides their human operators only a few seconds to make decisions, which places impossible demands on any potential critical deliberation.[402]

It is thus vitally important to ensure that deliberative reasoning is enabled in the design of supervisory control for LAWS. For example, a single operator controlling multiple weapons systems with only a short time window to veto their targeting decision is an absolutely unsuitable set-up of a human-machine interface. As Sharkey remarks, "[t]*here must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the*

[399] For example: W Schneider, W. and Chen, J. M. Controlled and automatic processing: behavior, theory and biological mechanisms. *Cognitive Science.* 2003(27). 525–59.; Evans, S. B. T. and Stanovich, K. E. Dual-process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science*. 2013. 8(3). 223–41.

[400] Sharkey. Staying in the Loop: Human Supervisory Control of Weapons, supra note 129, p. 32.

[401] Kahneman, D. *Thinking, Fast and Slow*. London: Penguin, 2011.

[402] Bode and Watts. *Meaning-less Human Control*, supra note 13, p. 59.

target, its significance in terms of the necessity and appropriateness of attack and the likely incidental and possible accidental effects of the attack."[403]

The advantage of computer-human cooperation, if laid out right, could be that the computer makes the automatic decision, as it is more effective at this task, and the human operator has time to conduct deliberative reasoning. Even the mode of operation when a computer program selects the target and a human must approve it before the attack has been experimentally shown to create automation bias, in which human operators disregard, or do not search for, contradictory information.[404] Providing only a short time to approve or veto a targeting decision is bound to reinforce automation bias and leave no room for doubt or deliberation. The time pressure will result in operators failing to fail in exercising deliberative, critical judgement. The Patriot case analysed above is a perfect example of the errors caused by the operators having to resort to fast veto. In a scenario where a computer program selects the targets and initiates attacks without human involvement, weapons systems could not comply with international law except in a very narrowly defined environment. Meaningful human control would be only exercised through the programming of the computer algorithm, which has to rely on what would be automatic reasoning in the case of humans.

More research is necessary to define parameters that would ensure that supervisory interfaces are designed to allow for the level of human reasoning needed to comply with IHL rules. What is certain is that this analysis is one of the factors to be considered when determining the level of human control that is exercised over LAWS in each case. Indeed, the quality of meaningful human control will mainly be influenced by the human-machine interface design, which needs to consider all factors influencing how humans perform in their role. The necessary level of human supervision will then be influenced by the previously analysed technological and conditional elements.

[403] Sharkey. Staying in the Loop: Human Supervisory Control of Weapons, supra note 129, p. 34.

[404] Ibid, p. 35.

VI. Assessing meaningful human control

This final part draws conclusions from the previous chapters and suggests an approach that should be taken when determining what constitutes meaningful human control. However, it follows from what has been argued that no clear-cut answer to all possible situations and every type of weapon system imaginable can be given. Instead, it seems that flexibility will be required when assessing meaningful human control.

The analysis above has elaborated upon a multitude of factors that influence the quality of human control exercised over critical functions of LAWS. These factors can be divided into three categories: the technological element, the conditional element, and the decision-making element. All elements consist of a number of factors that are interdependent, each one is important on its own, but the meaningfulness of human control is primarily determined by the interplay of these factors. For example, there is a close connection between restrictions on targets, timespan of deployment, and the environment. With one of these factors changing in a way that reduces the quality of human control, the other ones can balance it out. That is one of the reasons why the approach to defining a sufficient level of human control should be flexible and comprehensible.

Another reason lies in the fact that a variety of possible critical functions of weapon systems can feature autonomy, which would make it very difficult, if not impossible, to define one set of absolute rules applicable at all times. Rather, the quality of human control should be determined in each case, considering all the factors introduced above, ranging from the design of the programme to the mode of cooperation between the weapon system and its human operator. Certain rigid boundaries relating to particular elements must be respected. These boundaries often flow from requirements of IHL and/or realistic limits of human and computer capabilities. For example, it follows from the prohibition of the use of indiscriminate weapons that LAWS that are not sufficiently precise cannot be deployed in populated areas, as they would not be able to comply with the principle of distinction. Another example can be the training and expertise of the operators, coupled with the way how the human-machine interface is set up. The findings of psychological research on attention and cooperation with computer systems must be considered and operationalised in practice.

The flexibility of this approach also means the greater scope of applicability of the requirement of meaningful human control. Considering this rule implicit in IHL makes it applicable to all types of weapons used in the past, present, and future. It also addresses one of the most significant issues of the debate over LAWS: its orientation on the future. Defining particular elements but leaving space for their interaction enables applying the requirement of MHC even to technologies and weapon systems that are yet to be developed. It also aims to pose realistic limitations to the use of LAWS, allowing the militaries to benefit from the advantages computerised weapon systems provide. On the other hand, it bears in mind the overarching goal of IHL and weaponry development – sparing civilians and raising the bar on humanitarian standards in conflict.

VII. Conclusion

This paper has argued that meaningful human control ought to be exercised over critical functions of lethal autonomous weapon systems. The appropriate level of control should be determined for each particular set of circumstances in a way that ensures compliance of the weapon system with relevant rules of international humanitarian law, as well as the potential responsibility of its operator for all the resulting actions of the weapon system.

The aim was to continue the debate over the lawful use of LAWS, analyse the emerging principle of meaningful human control, and explore its elements and requirements. The concept of MHC was explored mainly through the lenses of IHL to determine its precise requirements. Various reasons why meaningful human control should be required were presented, with the focus on technological limitations and the need to ensure compliance with IHL rules. However, it is important to realise that the concerns and arguments behind these reasons can apply to a varying extent. In some circumstances, technological limitations of object recognition will not pose an obstacle as no other than military objects may be present. That is why the proposal is to adopt a flexible approach and differentiate sufficient levels of human control in various scenarios.

A particular added value of this paper is the more detailed analysis of a concrete list of elements and factors influencing how meaningful human control is. Practical guidance was offered on actions that commanders can take to exercise meaningful control over LAWS, as well as elaboration on how various factors impact each other. This added clarity can aid further legal analysis and policymaking. The main conclusion of this paper is twofold:

First, the required level of human supervision will mainly be influenced by the following:

(1) Predictability and reliability of the algorithm

The defining feature of LAWS is their autonomy which goes hand in hand with unpredictability. However, it may be possible that future development of technology will enable translating IHL rules on the conduct of hostilities into a computer programme and these systems will demonstrate a level of predictability and compliance comparable to humans. This may lead to a lower standard of human control necessary. If a system's review shows low reliability and object recognition failures occur, much stricter human control needs to be exercised.

(2) Complexity of the environment

LAWS may be operating in diverse circumstances, and each environment will pose challenges to their deployment. The prevailing opinion is that autonomous weapon systems should be used in low cluttered, predictable environments. A suggestion is made that if an environment is very simple and the occurrence of civilians or civilian objects very rare, a lower standard of human control may be sufficient.

(3) Targets

Finally, the category of targets that LAWS are programmed to engage influences the required human control standard. Many argue that autonomous weapon systems should only be used for anti-material targets. One of the reasons for this argument is that their technological limitations do not enable them to comply with the rule of distinction. Should LAWS deployment against persons be lawful, a notably higher level of human control would undoubtedly be required to be considered meaningful, as the operator needs to verify the legitimacy of the human target actively.

Second, the quality of meaningful human control will mostly be influenced by the human-machine interface design, which needs to consider all factors influencing how humans perform in their role, mainly:

(1) Expertise and training

The importance of adequate training and expertise cannot be overestimated. An example of air defence system operators shows that a lack of training can contribute to malfunctions of the system operation. Proper training can also target issues such as automation bias.

(2) System and situational understanding

The emphasis is very much put on system and situational understanding. Human operators need to trust the weapon system they are operating but this trust has to be reasonable and justified. They need to be aware of the system's strengths and weaknesses, especially the results of the weapon system's testing and review. Additionally, operators need to be able to understand how the weapon system is going to interact with the environment it is deployed in.

(3) The role of the human operator

Finally, relevant findings of psychological research, especially on attention and human-machine cooperation, need to be taken into account. To enable the exercise of meaningful control, the role of the operators has to be adjusted to human capabilities and behaviour. Particularly important is allowing time for deliberation and exercising qualitative judgement.

The analysis above selected issues relevant to how meaningful human control can be exercised over critical functions of lethal autonomous weapon systems. It was argued that a flexible and holistic approach should be taken in determining the applicable standard of human control on a case-by-case basis, but always with regard to the elements enumerated. Otherwise the compliance with meaningful human control cannot be monitored or predicted in weapon's review. It was stressed that the overall quality of human control will result from an interplay of various factors, some more influential than others. The appropriate level of meaningful human control is to be determined with its aims in mind - ensuring compliance of autonomous weapon systems with relevant rules of international humanitarian law, as well as the potential responsibility of their operators for the acts carried out, should individual criminal responsibility be the appropriate framework for addressing certain violations of IHL.

However, the list of factors presented is by no means exhaustive. The scope of this paper is limited to selected issues, it only points out the essential elements which could influence the level of human control that might be considered sufficient in different scenarios. There can be other elements not included in the analysis here that should be explored, particularly from an operational perspective. On the same note, it is necessary that a deeper analysis of how international criminal law and human rights law interact with the use of LAWS is conducted, which is beyond the scope of this paper.

Throughout this paper, it has been argued that the requirement of meaningful human control is essential to ensuring compliance of lethal autonomous weapon systems with international humanitarian law. It also addresses the moral and ethical concerns connected to their deployment in situations of armed conflict. However, it should be stressed that meaningful human control is not a magic wand that will make LAWS the perfect weapons, sparing civilian lives and reducing human suffering in warfare. The requirement is not all-powerful even if complied with.

Instead, it should be viewed as a tool helping States balance their pursuit of more efficient technologies and the need for retaining humans in the centre. This was the idea behind this paper's approach: to help clarify where the principle stems from and how it should be perceived and integrated. The aim was not to provide the final solution but rather to contribute to the debate concerning LAWS and how we should approach human control over their autonomous functions. This paper inevitably looks into the future and works with hypothetical situations. Nevertheless, it also strives to bring the debate into the context of current warfare and longstanding rules of IHL. It echoes the view expressed by many that the development of LAWS should bear the objectives of international humanitarian law in mind, to limit the suffering caused by warfare and alleviate its effects. The debate over LAWS mirrors the delicate balance between the strive for efficiency in action on the one hand and the laws of humanity on the other. And hopefully, if humans retain meaningful control over lethal autonomous weapon systems, we could do this balance justice.

Bibliography

Treaties and Conventions

American Convention on Human Rights, adopted on 22 November 1969, entered into force on 18 July 1978

Geneva Conventions of 12 August 1949, adopted on 12 August 1949, entered into force on 21 October 1950

Hague Convention II with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land, adopted on 29 July 1899, entered into force on 4 September 1900

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflict, adopted on 8 June 1977, entered into force on 7 December 1978

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), adopted on 8 June 1977, entered into force on 7 December 1978

Second Protocol to The Hague Convention of 1954 for the Protection of Cultural Property in the Event of Armed Conflict, adopted on 26 March 1999, entered into force 9 March 2004

UN General Assembly, Rome Statute of the International Criminal Court, adopted on 17 July 1998, entered into force on 1 July 2002

Judicial Decisions

ICC, Decision on the Confirmation of Charges, Lubanga (ICC-01/04-01/06), Pre-Trial Chamber I (29 January 2007)

ICJ, Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, ICJ Reports 1996

ICJ, North Sea Continental Shelf, Judgment, ICJ Reports 1969

ICJ, Right of Passage over Indian Territory (Portugal v. India), Merits, Judgment, ICJ

Reports 1960

ICJ, The Case of the S.S. Lotus, 1927 PCIJ Series A, No. 10

ICTY, Prosecutor v. Kupreškić et al, IT-95-16-T, Judgment, Trial Chamber (14 January 2000)

ICTY, Prosecutor v. Stanislav Galić, Case No. IT-98-29-T, Judgment, Trial Chamber (5 December 2003)

Books and publications

Akerson, D. The Illegality of Offensive Lethal Autonomy. In: Saxon, D. *International Humanitarian Law and the Changing Technology of* War. The Netherlands: Martinus Nijhoff Publishers, 2013.

Asaro, P. Jus nascendi, robotic weapons and the Martens Clause. In: *Robot Law*. Cheltenham, UK: Edward Elgar Publishing, 2016, 367–386.

Bode, I. and Huelss, H. The Future of Remote Warfare? Artificial Intelligence, Weapons Systems and Human Control. In: McKay, A., Watson, A. and Karlshøj-Pedersen, M. *Remote Warfare: Interdisciplinary Perspectives.* E-International Relations, 2021, 218-233.

Boothby, W. How Far Will the Law Allow Unmanned Targeting to Go? In: Saxon, D. *International Humanitarian Law and the Changing Technology of* War. The Netherlands: Martinus Nijhoff Publishers, 2013.

Boothby, W. *Weapons and the Law of Armed Conflict.* Oxford: Oxford University Press, 2009.

Boutin, B. and Woodcock, T., Aspects of Realizing (Meaningful) Human Control: A Legal Perspective. Forthcoming in: Geiß, R. and Lahmann, H., *Research Handbook on Warfare and Artificial Intelligence.* Cheltenham: Edward Elgar Publishing, 2022.

Cheng, B. United Nations Resolutions on Outer Space: 'Instant' International Customary Law? In: Cheng, B. *International Law: Teaching and Practice*, London: Stevens, 1982.

Dinstein, Y. *The Conduct of Hostilities under the Law of International Armed Conflict*. Cambridge: Cambridge University Press, 2004.

Gaeta, P. and Jain, A. G. Individualisation of IHL rules through criminal responsibility for war crimes and some (un)intended consequences. In: Akande, D. and Welsh, J. *The Individualisation of War*. Oxford: Oxford University Press, 2021.

Geiß, R. and Lahmann, H. Autonomous weapons systems: A paradigm shift for the law of armed conflict? In: Ohlin, J. D. *Research Handbook on Remote Warfare*. Cheltenham: Edward Elgar Publishing, 2017, 371–404.

Guzman, A. T. *How International Law Works: A Rational Choice Theory*. Oxford: Oxford University Press, 2008.

Hawley, J.K. and Mares, A.L. Human Performance Challenges for the Future Force: Lessons from Patriot after the Second Gulf War. In: Savage-Knepshield, P. *Designing Soldier Systems: Current Issues in Human* Factors. Burlington, VT: Ashgate, 2012, 3-34.

ICRC. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*. The Netherlands: Martinus Nijhoff Publishers, 1987.

ICRC. *Customary International Humanitarian Law.* Volume I: Rules. Cambridge: Cambridge University Press, 2005.

Jain, A. G. Autonomous Cyber Capabilities and Individual Criminal Responsibility for War Crimes. In: Liivoja, R. and Väljataga, A. *Autonomous Cyber Capabilities Under International Law*. NATO Cooperative Cyber Defence Centre, 2021.

Kahneman, D. *Thinking, Fast and Slow*. London: Penguin, 2011.

Kantowitz, B. H. and Sorkin, R. D. Allocation of Functions. In: Salvendy, G. *Handbook of Human Factors*. New York: Wiley, 1987, 355–69.

Lipovský, M. Mental Element (Mens Rea) of the Crime of Aggression and Related Issues. In: Šturma, P. *The Rome Statute of the ICC at Its Twentieth Anniversary*. Leiden, The Netherlands: Brill | Nijhoff, 2018.

Mosier, K. L. and Skitka, L. J. Human decision makers and automated decision aids: Made for each other? In: Parasuraman, R. and Mouloua, M. *Automation and human performance: Theory and applications*, CRC Press: 1996, 201–220.

Osinga, F. P. B. *Science, Strategy and War: The Strategic Theory of John* Boyd. London and New York: Routledge, 2006.

Pospisil, L. The Attributes of Law. In: Bohannon, P. *Law and Warfare: Studies in the Anthropology of Conflict.* New York: American Museum of Natural History, 1967, 25-41.

Sassòli, M. Humanitarian Law and International Criminal Law. In Cassese, A. *The Oxford Companion to International Criminal Justice.* Oxford: Oxford University Press, 2009.

Sassòli, M. *International Humanitarian Law Rules, Controversies, and Solutions to Problems Arising in Warfare*. Cheltenham: Edward Elgar Publishing, 2019.

Scharre, P. Army of None: *Autonomous Weapons and the Future of War*. New York and London: W. W. Norton, 2018.

Sharkey N. Staying in the Loop: Human Supervisory Control of Weapons. In: Bhuta, N., Beck, S., Geiss R., Kress, C. and Liu, Hin Yan. *Autonomous Weapons Systems: Law, Ethics, Policy.* Cambridge: Cambridge University Press, 2016, 23-38.

Thurnher, J. Examining Autonomous Weapon Systems form a Law of Armed Conflict Perspective. In: Nasu, H. and McLaughlin, R. *New Technologies and the Law of Armed Conflict*. The Netherlands: T.M.C Asser Press, 2014.

van den Boogaard, J. C., and Roorda, M. P. Autonomous Weapons and Human Control. In: Bartels, R., van den Boogaard, J. C., Ducheine, P. A. L., Pouw, E. and Voetelink, J. *Military Operations and the Notion of Control Under International Law*. Berlin: Springer, 2021, 421-39.

Wagner, M. Autonomy in the Battlespace. In: Saxon, D. *International Humanitarian Law and the Changing Technology of* War. The Netherlands: Martinus Nijhoff Publishers, 2013.

Werle, G. and Jeßberger, F. *Principles of International Criminal Law*. Oxford: Oxford University Press, 2020.

Winfield, A. F. T. *Robotics: A Very Short Introduction.* Very Short Introductions 330. Oxford: Oxford University Press, 2012.

Academic articles, reports

Amoroso, D. and Tamburrini, G. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. *Current Robotics Reports*. 2020. 1(4). 187–94. At: https://doi.org/10.1007/s43154-020-00024-3 (last accessed 13 April 2022).

Anderson, K. and Waxman, M. C. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can.* Stanford University, The Hoover Institution (Jean Perkins Task Force on National Security and Law Essay Series). 10 April 2013. At: https://ssrn.com/abstract=2250126 (last accessed 13 April 2022).

Asaro, P. On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanisation of Lethal Decision-Making. *International Review of the Red Cross*. 2013. 94(886). At: https://international-review.icrc.org/articles/banning-autonomous-weapon-systems-human-rights-automation-and-dehumanization-lethal (last accessed 14 April 2022).

Bo, M. Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute. *Journal of International Criminal Justice*. 2021. 19(2). 275–299. At: https://doi.org/10.1093/jicj/mqab005 (last accessed 2 May 2022).

Bo, M. Meaningful Human Control over Autonomous Weapon Systems: An (International) Criminal Law Account. *Opinion Juris*. 18 December 2020. At: http://opiniojuris.org/2020/12/18/meaningful-human-control-over-autonomous-weapon-systems-an-international-criminal-law-account/ (last accessed 27 April 2022).

Bode, I. and Watts, T. *Meaning-less Human Control.* Centre for War Studies, University of Southern Denmark with Drone Wars UK. 2021. At: https://dronewars.net/wp-content/uploads/2021/02/DW-Control-WEB.pdf (last accessed 13 April 2022).

Boulanin, V. and Verbruggen, M. *Mapping the Development of Autonomy in Weapons Systems*. Stockholm: Stockholm International Peace Research Institute. 2017. At: https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_ systems_11171.pdf (last accessed 15 April 2022).

Boulanin, V., Davison, N., Goussac N. and Peldán Carlsson M. *Limits of Autonomy in*

Weapon Systems: Identifying Practical Elements of Human Control. Stockholm International Peace Research Institute and International Committee of the Red Cross. 2020. At: https://www.icrc.org/en/document/limits-autonomous-weapons (last accessed 15 April 2022).

Chambers, A.B. and Nagel, D.C. Pilots of the future: Humans or computer? *Communications of the ACM*. 1985(28). 1187 - 1199.

Crootof, R. A Meaningful Floor for Meaningful Human Control. *Temple International & Comparative Law Journal*. 2016 (30). At: https://ssrn.com/abstract=2705560 last accessed 15 April 2022).

Cummings, M. L. *Automation bias in intelligent time critical decisions support systems*. American Institute of Aeronautics and Astronautics Third Intelligent Systems Conference, Chicago. 2004.

Cummings, M.L. *Automation Bias in Intelligent Time Critical Decisions Support Systems*. American Institute of Aeronautics and Astronautics. AIAA 3rd Intelligent Systems Conference Chicago. 2004. At: https://arc.aiaa.org/doi/10.2514/6.2004-6313 (last accessed 15 April 2022).

Desimone, R., and Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 1995. 18. 193–222. doi: 10.1146/annurev.ne.18.030195.001205.

Ekelhof, M. A. C. Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting. *Naval War College Review.* 2018 71(3). At: https://digital-commons.usnwc.edu/nwc-review/vol71/iss3/6/ (last accessed 13 April 2022).

Epstein, R. *The empty brain*. Aeon. 2016. At: https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer (last accessed 13 April 2022).

Evans, S. B. T. and Stanovich, K. E. Dual- process theories of higher cognition: advancing the debate. *Perspectives on Psychological Science*. 2013. 8(3). 223–41.

Franzoni V. and Poggioni V. Emotional book classification from book blurbs. *Proceedings – 2017 IEEE/WIC/ACM International Conference on Web Intelligence.* WI 2017. At: https://dl.acm.org/doi/10.1145/3106426.3109422 (last accessed 14 April 2022).

Franzoni, V., Milani, A., Nardi, D. and Vallverdú, J. Emotional machines: The next revolution. *Web Intelligence*. 2019. 17. 1–7. At: https://content.iospress.com/articles/web-intelligence/web190395 (last accessed 13 April 2022).

Franzoni, V., Milani, A., Pallottelli, S., Leung C.H.C. and Li Y. Context-based image semantic similarity. *12th International Conference on Fuzzy Systems and Knowledge Discovery.* FSKD 2015. 2016. At: https://ieeexplore.ieee.org/document/7382127 (last accessed 14 April 2022).

Goodfellow, I. J., Shlens, J. and Szegedy, C. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572.* 2014. At: https://arxiv.org/abs/1412.6572 (last accessed 13 April 2022).

Graziano M. S. A. and Webb T. W.,The attention schema theory: a mechanistic account of subjective awareness. *Frontiers in Psychology*. 2015(6). At: https://www.frontiersin.org/article/10.3389/fpsyg.2015.00500 (last accessed 15 April 2022).

Haladjian, H. H. and Montemayor, C. Artificial consciousness and the consciousness-attention dissociation. *Consciousness and cognition.* 2016(45). 210–225, At: https://pubmed.ncbi.nlm.nih.gov/27656787/ (last accessed 13 April 2022).

Hawley, J. K. Looking Back at 20 Years on MANPRINT on Patriot. *Army Research Laboratory*. 2007. At: https://apps.dtic.mil/sti/pdfs/ADA472740.pdf (last accessed 15 April 2022).

Hawley, J. K. Patriot Wars: Automation and the Patriot Air and Missile Defense System. Center for a New American Security Project on Ethical Autonomy Working Paper. 2017. At: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf?mtime=20170106135013&focal=none (last accessed 2 May 2022).

Heyns, C. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions.* UN General Assembly, A/HRC/23/47. 9 April 2013. At: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf (last accessed 13 April 2022).

Horowitz, M. and Scharre, P. Meaningful Human Control in Weapon Systems: A Primer. Center for a New American Security Project on Ethical Autonomy Working Paper. 2015.

At: www.cnas.org/sites/default/files/publications-pdf/Ethical_Autonomy_Working_Paper_031315.pdf (last accessed 15 April 2022).

Human Rights Watch. *Losing Humanity: The Case Against Killer Robots*. 2012. At: http://www.hrw.org/reports/2012/11/19/losing-humanity (last accessed 13 April 2022).

Jensen, E. T. The (Erroneous) Requirement for Human Judgment (and Error) in the Law of Armed Conflict. SSRN Electronic Journal. 2020. 96. 26–57. At: https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=2916&context=ils (last accessed 15 April 2022).

Kwik, J. A Practicable Operationalisation of Meaningful Human Control. *Laws.* 2022. 11(43). At: https://doi.org/10.3390/laws11030043 (last accessed 15 April 2022).

Lindsay, G. W. Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*. 2020 (14). At: https://www.frontiersin.org/article/10.3389/fncom.2020.00029 (last accessed 15 April 2022).

Liu, Hin-Yan. Categorisation and Legality of Autonomous and Remote Weapon Systems. *International Review of the Red Cross*. 2012. 627(94). At: https://international-review.icrc.org/sites/default/files/irrc-886-liu.pdf (last accessed 14 April 2022).

Marchant, G., Allenby, B., Arkin, R., Barrett, E., Borenstein, J., Gaudet, L., Kittrie, O., Lin, P., Lucas, G., O'Meara, R. and Silbermann, J. International Governance of Autonomous Military Robots. *Columbia Science and Technology Law Review.* 2011. XII. 272-315. At: https://academiccommons.columbia.edu/doi/10.7916/D8TB1HDW (last accessed 13 April 2022).

Mero, T. The Martens Clause, Principles of Humanity, and Dictates of Public Conscience. *American Journal of International Law*. 2000. 94(1), 78-89. At: doi:10.2307/2555232.

Oken, B. S., Salinsky, M. C. and Elsas, S. Vigilance, alertness, or sustained attention: physiological basis and measurement. Clin. Neurophysiol. 2006(117). 1885–1901. doi: 10.1016/j.clinph.2006.01.017.

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z. B. and Swami A. The limitations of deep learning in adversarial settings. *Security and Privacy (EuroS&P), 2016 IEEE European Symposium*. 2016. 372–387. At: https://arxiv.org/abs/1511.07528 (last accessed 13 April 2022).

Parasuraman R. and Manzey, D. H. Complacency and Bias in Human Use of Automation: An Attentional Integration. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2010. 52(3). 381–410. At: https://doi.org/10.1177/0018720810376055 (last accessed 15 April 2022).

Parasuraman, R. Human-computer monitoring. *Human Factors*. 1987. 29(6), 695-706.

Queguiner, J. Precautions under the Law Governing the Conduct of Hostilities. *International Review of the Red Cross.* 2006. 88(864). At: https://international-review.icrc.org/articles/precautions-under-law-governing-conduct-hostilities (last accessed 14 April 2022).

Roff, H. M. The Strategic Robot Problem: Lethal Autonomous Weapons in War. *Journal of Military Ethics*. 2014. 13(3). 211-227. At: https://doi.org/10.1080/15027570.2014.975010 (last accessed 15 April 2022).

Rose, N. and Osborne, T. Do the Social Sciences Create Phenomena: The Case of Public Opinion Research. 1999. *BRIT. J. SOC*. 367(50). At: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-4446.1999.00367.x (last accessed 15 April 2022).

Roth, B.R. The Enduring Significance of State Sovereignty. *Florida Law Review*. 2004. 1017(56). At: https://digitalcommons.wayne.edu/lawfrp/188/ (last accessed 15 April 2022).

Russell, S., Dewey, D. and Tegmark, M. Research Priorities for Robust and Beneficial Artificial Intelligence. *AI Magazine.* 2015. 36(4), 105-114. At: https://doi.org/10.1609/aimag.v36i4.2577 (last accessed 14 April 2022).

Sassòli, M. Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified. *International Law Studies / Naval War College*. 2014(90). 308-340. At: https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1017&context=ils (last accessed 14 April 2022).

Scharre, P. and Horowitz, M.C. *An Introduction to Autonomy in Weapon Systems*. Center for a New American Security. February 2015. At: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/Ethical-Autonomy-Working-Paper_021015_v02.pdf?mtime=20160906082257&focal=none (last accessed 15 April 2022).

Schmitt, M. Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics. *Harvard National Security Journal: Features Online*. 2013. At: https://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/ (last accessed 14 April 2022).

Schneider, W. and Chen, J. M. Controlled and automatic processing: behavior, theory and biological mechanisms. *Cognitive Science.* 2003(27). 525–59.

Searle, J. R. Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association.* 1990. 64(3). 21-37. At: https://philosophy.as.uky.edu/sites/default/files/Is%20the%20Brain%20a%20Digital%20Computer%20-%20John%20R.%20Searle.pdf (last accessed 13 April 2022).

Signorelli, C. M. Can Computers Become Conscious and Overcome Humans? *Frontiers in robotics and AI*. 2018/5. At: https://doi.org/10.3389/frobt.2018.00121 (last accessed 13 April 2022).

Signorelli, C. M. Types of cognition and its implications for future high-level cognitive machines. *AAAI Spring Symposium Series* (Berkeley, CA). 2017. At: http://aaai.org/ocs/index.php/SSS/SSS17/paper/view/\penalty-\@M15310 (last accessed 13 April 2022).

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. Mastering the game of Go without human knowledge. *Nature*. 2017. 550. 354–359. doi: 10.1038/nature24270.

Slijper, F. *Where to Draw the Line: Increasing Autonomy in Weapon Systems - Technology and Trends*. Utrecht, Neth.: PAX, 2017. At: www.paxvoorvrede.nl/ (last accessed 14 April 2022).

Smeulders, A. et al. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 22(12). 2000. 1349–1380. At: https://ieeexplore.ieee.org/document/895972 (last accessed 2022).

Sparrow, R. *Ethics as a source of law: The Martens clause and autonomous weapons*. 2017. At: https://blogs.icrc.org/law-and-policy/2017/11/14/ethics-source-law-martens-clause-autonomous-weapons/ (last accessed 15 April 2022).

Sparrow, R. Killer Robots. *Journal of Applied Philosophy*. 2007. 24 (1). 62–77.

Su, J., Vasconcellos Vargas, D. and Sakurai, K. One Pixel Attack for Fooling Deep Neural Networks. *IEEE Transactions on Evolutionary Computation.* 2019. 23(5). 828-841.

Taddeo, M. and Blanchard, A. *A Comparative Analysis of the Definitions of Autonomous Weapons.* 10 May 2021. At: http://dx.doi.org/10.2139/ssrn.3941214 (last accessed 13 April 2022).

Taddeo, M. and Taylor, I. *Ethical Principles for Artificial Intelligence in the Defence and Security Domain - Part 1 of 2*. The Alan Turing Institute. 2021.

Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. Deepface: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014. 1701-1708. At: https://ieeexplore.ieee.org/document/6909616 (last accessed 13 April 2022).

Waxman, M. and Anderson, K. Law and Ethics for Robot Soldiers. *Policy Review*. 2012 (176). At: https://scholarship.law.columbia.edu/faculty_scholarship/1742/ (last accessed 14 April 2022).

Weil, P. The Court Cannot Conclude Definitively... Non Liquet Revisited. *Columbia Journal of Transnational Law*. 1998. 109(36). At: https://heinonline.org/HOL/LandingPage?handle=hein.journals/cjtl36&div=14&id=&page= (last accessed 15 April 2022).

Wiener, E. L. *Complacency: Is the Term Useful for Air Safety*. Proceedings of the 26th Corporate Aviation Safety Seminar (Denver, CO: Flight Safety Foundation, Inc.). 1981.

International organisations

Ansell, D. Research and Development of Autonomous ‘Decision Making’ Systems. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

Biontino, M. (Chairperson of the Informal Meeting of Experts). Report of the 2015 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS). Geneva: 2015. At: http://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2015/Draft Report.pdf (last accessed 15 April 2022).

CCW. Canadian response to the Chair’s request for input on potential consensus recommendations. 2021. At: https://documents.unoda.org/wp-content/uploads/2021/06/Canada_Commentary-on-potential-consensus-recommendations.pdf (last accessed 28 May 2022).

CCW. Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, Annex III. CCW/GGE.1/2020/WP.7. 19 April 2021. At: https://documents.unoda.org/wp-content/uploads/2020/07/CCW_GGE1_2020_WP_7-ADVANCE.pdf (last accessed 13 April 2022).

CCW. Meeting of the High Contracting Parties to the CCW, Final report, Annex III. CCW/MSP/2019/9. 13 December 2019. At: https://undocs.org/Home/Mobile?FinalSymbol=CCW%2FMSP%2F2019%2F9&Language=E&DeviceType=Desktop&LangRequested=False (last accessed 13 April 2022).

CCW. Meeting of the High Contracting Parties to the CCW, Final report, Annex III, CCW/MSP/2019/9, 13 December 2019. At: https://undocs.org/Home/Mobile?FinalSymbol=CCW%2FMSP%2F2019%2F9&Language=E&DeviceType=Desktop&LangRequested=False (last accessed 13 April 2022).

CCW. Report of the 2014 informal Meeting of Experts on Lethal Autonomous Weapons Systems. CCW/MSP/2014/3. 11 June 2014. At: https://meetings.unoda.org/section/ccw-gge-2014-documents/ (last accessed 13 April 2022).

CCW. Report of the 2016 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS). Submitted by the Chairperson of the Informal Meeting of Experts. 2016. At: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Informal_Meeting_of_Experts_(2016)/ReportLAWS_2016_AdvancedVersion.pdf (last accessed 13 April 2022).

CCW. Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. CCW/GGE.1/2018/3. At: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/323/29/PDF/G1832329.pdf?OpenElement (last accessed 15 April 2022).

CCW. 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/Add.1. 8 November 2019. At: https://documents.unoda.org/wp-content/uploads/2020/09/1919338E.pdf (last accessed 13 April 2022).

Chair of the GGE LAWS. Chair's summary of discussion, Agenda item 6(b). 2018. At: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2018)/Summary%2Bof%2Bthe%2Bdiscussions%2Bduring%2BGGE%2Bon%2BLAWS%2BApril%2B2018.pdf (last accessed 15 April 2022).

Chairperson's Summary of the CCW GGE Meeting of 19 April 2021. CCW/GGE.1/2020/WP.7, Annex III, Commentaries on the 11 guiding principles. At: https://documents.unoda.org/wp-content/uploads/2020/07/CCW_GGE1_2020_WP_7-ADVANCE.pdf (last accessed 15 April 2022).

Davison, N., Weizmann, N. and Robinson, I. Background Paper by the International Committee of the Red Cross. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

Heyns, C. Increasingly Autonomous Weapon Systems: Accountability and Responsibility. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

ICRC. Addressing Internal Displacement in Times of Armed Conflict and Other Violence, 2018. At: https://shop.icrc.org/addressing-internal-displacement-in-times-of-armed-conflict-and-other-violence-pdf-en.html (last accessed 13 April 2022).

ICRC. Artificial intelligence and machine learning in armed conflict: A human-centred approach. Geneva: 6 June 2019. At: https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach (last accessed 13 April 2022).

ICRC. Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control? At: https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control (last accessed 15 April 2022).

ICRC. Ethics and autonomous weapon systems: An ethical basis for human control? UN Doc CCW/GGE.1/2018/WP.5. 29 March 2018. At: https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control (last accessed 15 April 2022).

ICRC. Guide to the Legal Review of New Weapons, Means and Methods of Warfare. Geneva: 2006. At: http://www.icrc.org/eng/assets/files/other/icrc_002_0902.pdf (last accessed 13 April 2022).

ICRC. Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law. Geneva: 2009. At: http://www.icrc.org/eng/assets/files/other/irrc-872-reports-documents.pdf (last accessed 13 April 2022).

ICRC. Paper prepared by the International Committee of the Red Cross relating to the crimes listed in article 8, paragraph 2 (e) (i), (ii), (iii), (iv), (ix) and (x), of the Rome Statute of the International Criminal Court. Doc. PCNICC/1999/WGEC/INF.2/Add.4. 15 December 1999. At: https://www.legal-tools.org/doc/dc889c/pdf (last accessed 15 April 2022).

ICRC. Position on Autonomous Weapon Systems. Geneva. 12 May 2021. At: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems (last accessed 13 April 2022).

ICRC. *Report of the ICRC Expert Meeting on 'Autonomous weapon systems: Technical, military, legal and humanitarian aspects'*. Geneva: March 2014. At: https://shop.icrc.org/expert-meeting-autonomous-weapon-systems-technical-military-legal-and-humanitarian-aspects.html?___store=en (last accessed 13 April 2022).

ICRC. Statement to the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts on Lethal Autonomous Weapons Systems under agenda item 6(b). Geneva: 27-31 August 2018. At: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2018)/2018_GGE%2BLAWS%2B2_6b_ICRC.pdf (last accessed 13 April 2022).

ICRC. Statements to the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts on Lethal Autonomous Weapons Systems. Geneva: 25–29 March 2019. At: https://www.unog.ch/__80256ee600585943.nsf/(httpPages)/5c00ff8e35b6466dc125839b003b62a1?OpenDocument&ExpandSection=7#_Section7 (last accessed 15 April 2022).

ICRC. The Element of Human Control, Working Paper, Convention on Certain Conventional Weapons (CCW) Meeting of High Contracting Parties. CCW/MSP/2018/WP.3. 20 November 2018. At: https://undocs.org/Home/Mobile?FinalSymbol=CCW%2FMSP%2F2018%2FWP.3&Language=E&DeviceType=Desktop&LangRequested=False (last accessed 15 April 2022).

ICRC. Views of the ICRC on Autonomous Weapon Systems. November 2016. At: https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system (last accessed 13 April 2022).

NATO. Allied Joint Doctrine for Joint Targeting. Brussels: NATO Standardization Office, 2016.

Righetti, L. Civilian robotics and developments in autonomous systems. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

Roff, H. M. Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons. Paper presented at Technical Report Briefing Paper for the Delegates at the Review Conference on the Convention on Certain Conventional Weapons. Geneva, Switzerland: December 12–16, 2016. At: https://article36.org/wp-content/uploads/2016/12/Control-or-Judgment_-Understanding-the-Scope.pdf (last accessed 13 April 2022).

Sharkey, N. Autonomous weapons and human supervisory control. In: ICRC. *Autonomous weapon systems: Technical, military, legal and humanitarian aspects*. Expert meeting, Geneva: March 2014. At: https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems (last accessed 13 April 2022).

UN Security Council. Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011). S/2021/229. 8 March 2021. At: https://undocs.org/Home/Mobile?FinalSymbol=S%2F2021%2F229&Language=E&DeviceType=Desktop (last accessed 13 April 2022).

UNIDIR. Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies. A Primer. August 2018. At: https://unidir.org/publication/algorithmic-bias-and-weaponization-increasingly-autonomous-technologies (last accessed 13 April 2022).

UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches. A Primer. Geneva: UNIDIR, 2017. At: https://www.unidir.org/publication/weaponization-increasingly-autonomous-technologies-concerns-characteristics-and (last accessed 13 April 2022).

UNIDIR. The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward. 2014. At: https://unidir.org/publication/weaponization-increasingly-autonomous-technologies-considering-how-meaningful-human (last accessed 15 April 2022).

United Kingdom. Expert Paper: The Human Role in Autonomous Warfare, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System. Agenda Item 5. Technical Report CCW/GGE.1/2020/WP.6. Geneva: 21–25 September 2020 and 2–6 November 2020. At: https://documents.unoda.org/wp-content/uploads/2020/07/CCW_GGE1_2020_WP_7-ADVANCE.pdf (last accessed 28 May 2022).

United States of America. Intervention on Appropriate Levels of Human Judgment over the Use of Force delivered by John Cherry. Paper presented at Technical Report Convention on Certain Conventional Weapons (CCW), Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS). Geneva, Switzerland, 15 November 2017. At: https://geneva.usmission.gov/2017/11/16/u-s-statement-at-ccw-gge-meeting-intervention-on-appropriate-levels-of-human-judgment-over-the-use-of-force/ (last accessed 28 May 2022).

Yusuf, A. A. Statement of the President of the International Court of Justice before the Sixth Committee of the General Assembly. New York: 1 November 2019. At: https://www.icj-cij.org/public/files/press-releases/0/000-20191101-STA-01-00-EN.pdf (last accessed 15 April 2022).

Miscellaneous

Anderson, W. R., Husain, A. and Rosner, M. The OODA Loop: Why Timing Is Everything. Cognitive Times. December 2017. At: https://www.europarl.europa.eu/cmsdata/155280/WendyRAnderson_CognitiveTimes_OODA%20LoopArticle.pdf (last accessed 15 April 2022).

Article 36. Key Areas for Debate on Autonomous Weapons Systems. May 2014. At: http://www.article36.org/wp- content/uploads/2014/05/A36-CCW-May-2014.pdf (last accessed 15 April 2022).

Article 36. Key elements of meaningful human control. Paper presented at Technical Report Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva, Switzerland, April 11–15 2016.

Article 36. Killer Robots: UK Government Policy on Fully Autonomous Weapons. April 2013. At: http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf (last accessed 15 April 2022).

Article 36. Structuring Debate on Autonomous Weapons Systems. November 2013. At: http://www.article36.org/wp- content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf (last accessed 15 April 2022).

Bennett, G. The Precision-Recall Trade-Off. 21 June 2020. At: https://datascience-george.medium.com/the-precision-recall-trade-off-aa295faba140 (last accessed 13 April 2022).

Czech Republic. Statement at the CCW Meeting of Experts on Lethal Autonomous Weapons Systems. 2015. At: https://www.mzv.cz/public/29/e/7d/1448252_1299062_CZ_statement_general_debate_LAWS_ver2_1.pdf (last accessed 15 April 2022).

Government of the Netherlands. Government Response to AIV/CAVV Advisory Report No. 97, 'Autonomous Weapon Systems: The Need for Meaningful Human Control'. 2 March 2016. At: http://aiv-advies.nl/ (last accessed 15 April 2022).

Hennessy, M. Clearpath Robotics Takes Stance Against 'Killer Robots. Clearpath Robotics press release. 13 August 2014. At: https://www.clearpathrobotics.com/2014/08/clearpath-takes-stance-against-killer-robots/ (last accessed 13 April 2022).

ICRAC. Berlin Statement. At: http://icrac.net/statements/ (last accessed 15 April 2022).

Kellenberger, J. International Humanitarian Law and New Weapon Technologies. ICRC, Keynote address at 34th Round Table on Current Issues of International Humanitarian Law, San Remo. 8-10 September 2011. At: https://international-review.icrc.org/articles/international-humanitarian-law-and-new-weapon-technologies-34th-round-table-current-issues (last accessed 15 April 2022).

Lendave, V. Python Guide to Precision-Recall Tradeoff. Developers Corner. June 10, 2021. At: https://analyticsindiamag.com/python-guide-to-precision-recall-tradeoff/ (last accessed 13 April 2022).

Leung, R. The Patriot Flawed? CBS News. 19 February 2004. At: https://www.cbsnews.com/news/the-patriot-flawed-19-02-2004/ (last accessed 15 April 2022).

Manyika, J., Silberg, J. and Presten, B. What Do We Do About the Biases in AI? Harvard Business Review. 25 October 2019. At: https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai (last accessed 13 April 2022).

Missile Defense Project. "Patriot," Missile Threat. 2018. At:

https://missilethreat.csis.org/system/patriot/ (last accessed 15 April 2022).

Oxford English Dictionary. “Artificial intelligence”. At: https://www.lexico.com/definition/artificial_intelligence (last accessed 13 April 2022).

Oxford English Dictionary. “Feasible”. At: https://www.lexico.com/definition/feasible (last accessed 13 April 2022).

Oxford English Dictionary. “Machine learning”. At: https://www.lexico.com/definition/machine_learning (last accessed 13 April 2022).

Pearson, T. The Ultimate Guide to the OODA Loop. 2017. At: https://taylorpearson.me/ooda-loop/ (last accessed 15 April 2022).

Sauer, F. ICRAC Statement on Technical Issues to the 2014 UN CCW Expert Meeting, ICRAC INT‘L COMM. FOR ROBOT ARMS CONTROL. 14 May 2014. At: https://www.icrac.net/icrac-statement-on-technical-issues-to-the-2014-un-ccw-expert-meeting/ (last accessed 15 April 2022).

Sharp, C. Cognitive Lethal Autonomous Weapons Systems (CLAWS). Articles of War. 5 November 2021. At: https://lieber.westpoint.edu/cognitive-lethal-autonomous-weapons-systems/ (last accessed 13 April 2022).

UK Ministry of Defence. Aircraft Accident to Royal Air Force Tornado GR MK4A ZG710. 2004. At: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/82817/maas03_02_tornado_zg710_22mar03.pdf (last accessed 15 April 2022).

UK Ministry of Defence. Joint Doctrine Note 2/11, The UK Approach to Unmanned Aircraft Systems. 2011. At: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505JDN_211_UAS_v2U.pdf (last accessed 13 April 2022).

UK Ministry of Defence. Joint Doctrine Publication 030.2 Unmanned Aircraft Systems. August 2017. At: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf (last accessed 13 April 2022).

US Department of Defense. Directive 3000.09 on Autonomy in Weapon Systems. 2012. At: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf (last accessed 13 April 2022).

US DoD Defense Science Board. The role of autonomy in DoD systems, Task Force Report. July 2012. At: www.fas.org/irp/agency/dod/dsb/autonomy.pdf (last accessed 13 April 2022).

US DoD, Defense Science Board. Task Force Report: The Role of Autonomy in DoD Systems. 19 July 2012. At: https://irp.fas.org/agency/dod/dsb/autonomy.pdf (last accessed 13 April 2022).

US DoD. Report of the Defense Science Board Task Force on Patriot System

Performance. Report Summary. Washington, DC: Office of the Under-Secretary of Defense for Acquisition, Technology, and Logistics, January 2005. At: https://dsb.cto.mil/reports/2000s/ADA435837.pdf (last accessed 15 April 2022).

US DoD. Unmanned Systems Integrated Roadmap 2013, FY2013-2038. At: https://www.hsdl.org/?abstract&did=747559 (last accessed 13 April 2022).

Vincent, J. Have autonomous robots started killing in war? The Verge. 3 June 2021, At: https://www.theverge.com/2021/6/3/22462840/killer-robot-autonomous-drone-attack-libya-un-report-context (last accessed 13 April 2022).

Meaningful Human Control in Autonomous Weapons

ABSTRACT

The research on autonomy in robotic systems is flourishing in many areas, but none is deemed as troubling as the development of lethal autonomous weapon systems (LAWS). It raises various compelling questions, legal and ethical ones. Discussions on the potential challenges posed by these emerging technologies highlighted the desirability of a certain level of human control. The notion of meaningful human control (MHC) over LAWS has gained widespread support. However, the principle itself and its requirements are yet to be defined.

To this end, this paper analyses the emerging principle of MHC and explores its elements. It aims to clarify questions such as where the principle stems from and how it should be perceived and integrated into State practice. First, the definition and categorisation of LAWS are shortly addressed to provide an introduction to the topic. Second, it is argued that it is necessary to insist on the requirement of MHC, particularly because of technological limitations of current and future technology, such as object recognition and classification, bias, or unpredictability. The arguments stemming from the rules of international humanitarian law (IHL) on the conduct of hostilities are explored, mainly the rules of distinction, proportionality, and precautionary measures. Briefly, the possibility of attributing individual criminal responsibility for acts carried out by LAWS is debated. Third, a case study of an air defence system is analysed with the conclusion that systems with automated functions may already be setting a precedent for what is considered meaningful in terms of human control. Fourth, it is argued that while the requirement of MHC does not (yet) constitute a rule of customary international law, IHL rules implicitly require human control to be maintained over LAWS. Fifth, the requirement of MHC is analysed in detail, particularly what should be control exercised over and at which level. The central part focuses on the technological, conditional, and decision-making elements which influence how meaningful the control is. Finally, it is argued that the approach to defining the appropriate level of human control should be flexible.

KEYWORDS: lethal autonomous weapon systems, meaningful human control, international humanitarian law

Smysluplná lidská kontrola v kontextu autonomních zbraní

ABSTRAKT

Výzkum autonomie v robotických systémech vzkvétá v mnoha oblastech, ale žádná z nich není považována za tak znepokojující jako vývoj smrtících autonomních zbraňových systémů, který vyvolává naléhavé otázky, jak právní, tak i etické. Diskuse zdůrazňují nezbytnost určité úrovně lidské kontroly. Pojem smysluplné lidské kontroly nad autonomními zbraněmi získal na popularitě, avšak samotný princip a jeho požadavky doposud nebyly definovány.

Za tímto účelem tato práce analyzuje pojem smysluplné lidské kontroly a zkoumá jeho prvky. Cílem je objasnit určité otázky, jako například z jakých pramenů tato zásada vyvěrá a jak by měla být vnímána a začleněna do praxe Států. Práce se nejprve stručně věnuje definici a kategorizaci autonomních zbraní, a to za účelem poskytnutí úvodu do tématu. Zadruhé práce předkládá důvody, proč je nutné trvat na požadavku smysluplné lidské kontroly, a to zejména kvůli omezením současných a budoucích technologií, jako je rozpoznávání a klasifikace objektů, zkreslení nebo nepředvídatelnost. Následně se analýza zaměřuje na argumenty vyplývající z pravidel mezinárodního humanitárního práva, především co se týče principů rozlišování, proporcionality a preventivních opatření. Stručně je diskutována možnost přičítání individuální trestní odpovědnosti za následky plynoucí z použití autonomních zbraní. Zatřetí práce analyzuje případovou studii systému protivzdušné obrany a to závěrem, že systémy s automatizovanými funkcemi již mohou vytvářet precedens smysluplné lidské kontroly. Začtvrté práce argumentuje, že ačkoli požadavek smysluplné lidské kontroly (prozatím) nepředstavuje pravidlo mezinárodního obyčejového práva, principy a normy mezinárodního humanitárního práva implicitně vyžadují, aby nad autonomními zbraněmi byla vykonávána lidská kontrola. Následně je podrobně analyzován požadavek smysluplné lidské kontroly, a to zejména s ohledem na to, co je předmětem kontroly a na jaké úrovni by měla být vykonávána. Jádrem práce je potom část zaměřující se na prvky, které ovlivňují smysluplnost kontroly, z hlediska technologie, podmínek užití autonomních zbraní a rozhodovacího procesu. Závěrem je argumentováno, že přístup k definování vhodné úrovně lidské kontroly by měl být flexibilní a brát v potaz realitu ozbrojených konfliktů.

KLÍČOVÁ SLOVA: smrtící autonomní zbraňové systémy, smysluplná lidská kontrola, mezinárodní humanitární právo