Aspects of Realizing (Meaningful) Human Control

a legal perspective

Boutin, B.; Woodcock, T.

DOI
10.4337/9781800377400.00016
Publication date
2024
Document Version
Author accepted manuscript
Published in
Research Handbook on Warfare and Artificial Intelligence

Research paper series

ASSER INSTITUTE
Centre for International & European Law

May 2022

Aspects of Realizing (Meaningful) Human Control: A Legal Perspective

Berenice Boutin and Taylor Woodcock

07

Centre for International & European Law
www.asser.nl

Forthcoming in: Geiß, R. and Lahmann, H. (eds.), *Research Handbook on Warfare and Artificial Intelligence*, Edward Elgar Publishing.

Link to SSRN Asser page
www.asser.nl

Cite as: ASSER research paper 2022-07

Author Contact Details: b.boutin@asser.nl t.woodcock@asser.nl

Abstract

The concept of ‘meaningful human control’ (MHC) has progressively emerged as a key frame of reference to conceptualize the difficulties posed by military applications of artificial intelligence (AI), and to identify solutions to mitigate these challenges. At the same time, this notion remains relatively indeterminate and difficult to operationalize. If MHC is to support the existing framework of international law applicable to military AI, it needs to be clarified in order to deal with the challenges of AI broadly construed, not limited to ‘autonomous weapons systems’ (AWS). This chapter seeks to refine the notion of MHC by exploring its nature and purpose, and reflecting on how MHC relates to core concepts of human agency and responsibility. Building on this analysis, we propose ways to operationalize MHC, in particular by putting greater emphasis on pre-deployment stages. A legal ‘compliance by design’ approach is advanced by the authors as a means to address the complex realities when military decision-making processes are mediated by AI-enabled technologies.

Keywords

autonomous weapons systems, meaningful human control, artificial intelligence, human agency, human-machine interaction, compliance by design

Aspects of Realizing (Meaningful) Human Control:

A Legal Perspective

Bérénice Boutin and Taylor Woodcock*

Forthcoming in:

Robin Geiß and Henning Lahmann (eds.), *Research Handbook on Warfare and Artificial Intelligence* (Edward Elgar)

* Bérénice Boutin is Senior Researcher in International Law at the Asser Institute, and Project Leader of the DILEMA Project on Designing International Law and Ethics into Military Artificial Intelligence (b.boutin@asser.nl). Taylor Woodcock is PhD Researcher in International Law at the Asser Institute, within the DILEMA Project (t.woodcock@asser.nl). The DILEMA Project received funding from the Dutch Research Council (NWO) Platform for Responsible Innovation (NWO-MVI).

1. Introduction

The emergence of artificial intelligence (AI) in the military domain elicits a number of questions about the lawfulness and ethics of new digital modes of warfare. Applications of AI developed and used by the military have the potential to bring about transformations in how armed conflict is conducted, resulting in new human-machine relationships and impacting how human actors exercise control. The concept of 'meaningful human control' (MHC) has progressively emerged as a key frame of reference to conceptualize the difficulties posed by these technologies and to identify solutions to mitigate these challenges. At the same time, this notion remains relatively indeterminate and difficult to operationalize. If MHC is to support the existing framework of international law applicable to military AI, it needs to be refined in order to deal with the challenges of AI broadly construed, not limited to 'autonomous weapons systems' (AWS). Clarifying how this notion can be operationalized in technical and practical terms is also necessary to ensure respect for international law. A legal 'compliance by design' approach is advanced by the authors as a means to address the complex realities when military decision-making processes are mediated by AI-enabled technologies.

In this Chapter, we first introduce the origins, meaning, and limitations of the concept of MHC (Section 2). Further, we seek to refine the notion of MHC by exploring its nature and purpose and reflecting on how MHC relates to core concepts of human agency and responsibility (Section 3). Building on this analysis, we propose ways to operationalize MHC, in particular by putting greater emphasis on pre-deployment stages (Section 4). Section 5 then offers some concluding remarks.

2. Origins, meaning, and limitations of the concept of MHC

2.1. Origins and development of the concept of MHC

Since its inception, the debate on the legality and morality of AWS has revolved around the necessity of retaining human control over these systems. The International Committee for Robot Arms Control (ICRAC), an NGO founded in 2009, was amongst the first to advocate that 'machines should not be allowed to make the decision to kill people'.[1] At a time when the United States issued a Directive defining AWS as '[a] weapon system that, once activated, can select and engage targets without further intervention by a human operator',[2] a number of scholars and NGOs began extensively debating what forms and degrees of human intervention or human control ought to be maintained over AWS.[3] It is in this context that views progressively converged towards the

[1] ICRAC Mission Statement, September 2009, https://www.icrac.net/statements.
[2] US Department of Defense, 'Autonomy in Weapons Systems' (2012) Directive 3000.09, 13.
[3] Peter Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 International Review of the Red Cross 687; Human Rights Watch (HRW) and Harvard Law School's International Human Rights Clinic (IHRC), 'Losing Humanity: The Case against Killer Robots' (2012); Mark

qualification that human control over autonomous systems needed to be 'meaningful'. Launched in 2013, the Campaign to Stop Killer Robots is a large coalition of NGOs that relentlessly advocated for a 'preemptive ban on weapons systems that would be able to select and attack targets without meaningful human intervention',[4] and significantly contributed to framing the debate on this topic. Along the same lines, the United Kingdom-based NGO Article 36 issued early calls for a 'commitment not to develop [...] systems that could undertake attacks without meaningful human control.'[5]

The notion of MHC crystallized in the context of inter-governmental debates on lethal autonomous weapons systems (LAWS) conducted under the auspices of the Convention on Certain Conventional Weapons (CCW).[6] These debates, taking place amongst a Group of Governmental Experts (GGE) with the involvement of civil society and academia, have constituted the prime forum to address concerns regarding the challenges of autonomy in weapon systems and human control over the lethal use of force. Since the first informal meeting within the CCW in 2014, MHC has served as a central pillar of deliberations on the legality and ethics of AWS.[7] Background papers submitted by NGOs in this context were particularly influential in affirming the need for MHC and attempting to flesh out this concept.[8] MHC has since remained a guiding frame of reference for thinking about AWS,[9] and has become a prolific topic of scholarly reflection.[10] To this day, a majority of States considers human control to be the relevant conceptual framework against which to consider AWS.[11]

Gubrud, 'The Principle of Humanity in Conflict' (2012), ICRAC, https://www.icrac.net/the-principle-of-humanity-in-conflict/.

[4] Campaign to Stop Killer Robots, 'Statement to the UN General Assembly First Committee on Disarmament and International Security', 29 October 2013, https://www.stopkillerrobots.org/wp-content/uploads/2013/10/KRC_StatementUNGA1_29Oct2013_delivered.pdf.

[5] Article 26, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (2013), https://article36.org/wp-content/uploads/2020/12/Policy_Paper1.pdf, 1.

[6] 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 (10 October 1980).

[7] 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, 'Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)' (2014), UN Doc CCW/MSP/2014/3, para. 20.

[8] Campaign to Stop Killer Robots, 'The Convention on Conventional Weapons and Fully Autonomous Weapons', Background Paper (2013) https://www.stopkillerrobots.org/wp-content/uploads/2013/09/KRC_BackgrounderCCW_26Sep2013.pdf; Article 36, 'Structuring Debate on Autonomous Weapons Systems: Memorandum for delegates to the Convention on Certain Conventional Weapons' (2013) https://article36.org/wp-content/uploads/2020/12/Autonomous-weapons-memo-for-CCW.pdf; Heather M Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf.

[9] Chair of the GGE LAWS, 'Commonalities in national commentaries on guiding principles' (2020) https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2020/gge/documents/chair-paper-commonalities.pdf, para 8.

[10] See e.g., Thompson Chengeta, 'Defining the Emerging Notion of "Meaningful Human Control" in Weapon Systems' (2017) 49 New York University Journal of International Law and Politics 833; Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal 53; Merel AC Ekelhof, 'Complications of a Common Language: Why It Is so Hard to Talk about Autonomous Weapons' (2017) 22 Journal of Conflict and Security Law 311; Michael C Horowitz, 'Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems' (2016) 30 Temple International and Comparative Law Journal 85.

[11] See e.g., Chair of the GGE LAWS, 'Commonalities in national commentaries on guiding principles' (2020), para 8; Statement to the GGE LAWS submitted by Republic of Costa Rica, the Republic of Panama, the Republic of Peru, the Republic of the Philippines, the Republic of Sierra Leone and the Eastern Republic of Uruguay (2021)

Participants from civil society have similarly considered that 'the desire to retain human control over the use of force provides a sound basis for collective action'.[12]

2.2. Defining what MHC entails

While the concept of MHC has developed and been debated for over a decade, it has remained relatively elusive, its concrete meaning being difficult to capture. As noted by the United Nations Institute for Disarmament Research (UNIDIR), 'the idea of Meaningful Human Control is intuitively appealing even if the concept is not precisely defined'.[13] Calls to elaborate on what constitutes MHC have been ongoing in the context of the GGE debates,[14] with recognition that the concept requires further clarification.[15] The 'nature', 'type', 'degree', 'quality', and 'extent' of human control, and combinations thereof, have all been raised in these debates as requiring elaboration.[16] Most recently, in a working paper submitted in 2021, the Non-Aligned Movement, representing 120 States, pointed to the importance of reaching 'a common understanding of what "meaningful" or "effective" human control entails in practice'.[17] Whilst the qualifying 'meaningful' has been dropped by some participants in debates in favour of other formulations such as 'substantive', 'appropriate', 'sufficient', 'effective' or 'adequate' human control,[18] or 'human control' alone,[19] others focus on the related but different concepts of 'human judgement' or 'human involvement'.[20]

---

https://documents.unoda.org/wp-content/uploads/2021/06/Costa-Rica-Panama-Peru-the-Philippines-Sierra-Leone-and-Uruguay.pdf, 2.

[12] Reaching critical will, 'Civil society perspectives on the Group of Governmental Experts on lethal autonomous weapons systems of the Convention on Certain Conventional Weapons' (21 September 2020), 4.

[13] UNIDIR, 'The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward' (2014) UNIDIR Resources No 2, 2.

[14] See e.g. 'Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)' (2014), para 20; Chair of the GGE LAWS, 'Chair's summary of discussion' (2018) 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, 6.

[15] GGE LAWS, 'Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', (2019) UN Doc CCW/GGE.1/2019/3/Add.1, para 19; 'Possible outcome of 2019 Group of Governmental Experts and future actions of international community on Lethal Autonomous Weapons Systems UN Doc', Working paper submitted by Japan (22 March 2019) CCW/GGE.1/2019/WP.3, para 26.

[16] See e.g., Chair of the GGE LAWS, 'Commonalities in national commentaries on guiding principles' (2020), paras 11, 21(c) ('nature'); 'Working Paper on Lethal Autonomous Weapons Systems submitted to the GGE LAWS by Poland' (2018) UN Doc CCW/GGE.1/2018/WP.3, para 3; 'Statement to the GGE LAWS submitted by the European Union' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/European-Union.pdf, 2, ('type and degree'); 'Statement to the GGE LAWS submitted by Brazil, Chile and Mexico' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/Brazil-Chile-Mexico.pdf, 7 ('quality and extent'); 'Joint Working Paper submitted by Republic of Costa Rica, the Republic of Panama, the Republic of Peru, the Republic of the Philippines, the Republic of Sierra Leone and the Eastern Republic of Uruguay' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/Costa-Rica-Panama-Peru-the-Philippines-Sierra-Leone-and-Uruguay.pdf, 2 ('type and extent').

[17] 'Working paper submitted to the GGE LAWS by the Bolivarian Republic of Venezuela on behalf of the Non-Aligned Movement (NAM) and Other States Parties to the CCW' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/NAM.pdf, para 10.

[18] See, Chair of the GGE LAWS, 'Chair's summary of discussion' (2018), 7, Agenda item 6(b); ICRC, 'Ethics and autonomous weapon systems: An ethical basis for human control?' (29 March 2018) UN Doc CCW/GGE.1/2018/WP.5, 2.

[19] See e.g., 'Statement to the GGE LAWS submitted by Austria, Brazil, Chile, Ireland, Luxembourg, Mexico and New Zealand' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/Austria-Brazil-Chile-Ireland-Luxembourg-Mexico-and-New-Zealand.pdf, 2.

[20] See e.g., 'Canadian response to the Chair's request for input on potential consensus recommendations' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/Canada_Commentary-on-potential-consensus-

Essentially, MHC is usually seen as entailing a sufficient degree of cognitive awareness of the operator using an AWS, and allowing for the exercise of human judgement, thus leading to informed decision-making.[21] It captures the idea that human involvement, without more, may not be sufficient to overcome the legal and ethical challenges raised by autonomy in military technologies. Yet, there is no consensus on who should exercise control (i.e. the operator, the commander, other relevant actors), over what control is to be exercised (i.e. individual attacks, certain targeting decisions, all stages in the lifecycle of an autonomous system),[22] and how control can be achieved in practice (e.g. real time approval of selected targets, intervention with override capabilities, overall supervision of an operation, control at the pre-deployment stages of design, manufacturing and testing).[23]

Amoroso and Tamburrini argue for a threefold role for humans in their relations with autonomous systems in the military domain: a 'fail safe actor' to prevent unlawful attacks due to system malfunctions; an 'accountability attractor' to ensure the ascription of responsibility in case of breaches; and a 'moral agency enactor', to ensure human dignity is respected by leaving decisions to resort to the lethal use of force to human beings.[24] Human involvement in this sense aims to address the main concerns of preventing unlawful attacks, attributing responsibility and respecting human dignity. They further suggest that, in any case, autonomy in selecting and engaging targets on the basis of pre-programmed goals may only be acceptable in terms of MHC if subject to strict restrictions in spatial and temporal parameters of operation so that 'targeting decisions [are] entirely and reliably traceable to human operators'.[25]

Another interesting perspective is provided by Roff and Moyes, who propose a three-layered approach to MHC across the stages of 'ante bellum' (design, development, acquisition, and testing), 'in bello' (deployment) and 'post bellum' (command structures and accountability frameworks).[26] At the ante bellum stage, they call for embedding MHC in autonomous systems, although no

recommendations.pdf, 1–2, suggesting that there is a need to consider how 'appropriate human involvement' is connected with compliance with IHL, where involvement comprises both judgment and control.

[21] Heather M Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) Briefing paper, 1; Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal 53, 56–57; Merel Ekelhof, 'Autonomous Weapons: Operationalizing Meaningful Human Control' (2018) ICRC Blog.

[22] UNIDIR, 'The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward' (2014), 2; Heather M Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) Briefing paper, 2–4; 'Categorizing lethal autonomous weapons systems – A technical and legal perspective to understanding LAWS', Working paper submitted by Estonia and Finland (24 August 2018) UN Doc CCW/GGE.2/2018/WP.2, para 19.

[23] GGE LAWS, 'Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', (2019) UN Doc CCW/GGE.1/2019/3/Add.1, para 19; 'Categorizing lethal autonomous weapons systems – A technical and legal perspective to understanding LAWS', Working paper submitted by Estonia and Finland (24 August 2018) UN Doc CCW/GGE.2/2018/WP.2, para 19; 'Statement of the GGE Laws submitted by Spain' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/Spain1.pdf, 1; 'Statement to the GGE LAWS submitted by Brazil' (2021) https://documents.unoda.org/wp-content/uploads/2021/06/Brazil.pdf, 1–2.

[24] Daniele Amoroso and Guglielmo Tamburrini, 'In Search of the "Human Element": International Debates on Regulating Autonomous Weapons Systems' (2021) 56 The International Spectator 20, 30.

[25] Ibid., 31.

[26] Heather M Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) Briefing paper, 3.

solutions are provided on how to concretely achieve this. During deployment, they emphasize control exercised by commanders over individual attacks. They submit that 'human evaluations and judgments are necessary for adherence to the law' and that such exercise of 'human legal judgment' must in particular occur at the tactical level.[27] Building on this, they suggest that, post bellum, operators or commanders can be held accountable for failure to exercise MHC.[28]

In its proposal for a draft treaty on AWS, the Campaign to Stop Killer Robots also elaborated on the meaning of MHC.[29] MHC is at the heart of the proposed instrument, forming the basis of a general obligation to ensure MHC over the use of force, prohibitions on weapons that function without MHC and positive obligations to ensure MHC is retained in all autonomous systems. Here, MHC is conceived of having 'decision-making', 'technological', and 'operational' components which work in parallel to enhance human control.[30] The decision-making component relates to the 'information and ability [of humans] to make decisions about whether the use of force complies with legal rules and ethical principles', requiring understanding of the operational environment and functionality of the system, as well as sufficient time to decide.[31] The technological aspect comprises 'embedded features' of a system relating to predictability, reliability, and transparency, as well as the possibility for human override. Finally, the operational component functions as limitations on the parameters of use of the system in terms of time, duration, geography, and types of targets (e.g. personnel or materiel).[32]

While these different approaches are useful to frame the discussion on MHC, they often still fall short of providing sufficient grounds for understanding and implementing MHC in practice. It can however be observed that, as international debates on the topic continue, more and more refinement is brought into what MHC may entail. For instance, in a draft normative framework submitted to the CCW GGE debates in 2021, France and Germany indicated that 'appropriate/sufficient' human control should allow humans to:

- understand - depending on their role and level of responsibilities - the systems' way of operating, effect and likely interaction with its environment;
- evaluate and monitor the reliability of the systems;
- validate the usability/serviceability of the systems;
- define and validate rules of use and rules of engagement;

[27] Ibid., 4–5.
[28] Ibid., 6.
[29] Campaign to Stop Killer Robots, 'Key Elements of a Treaty on Fully Autonomous Weapons (2019) https://www.stopkillerrobots.org/wp-content/uploads/2020/04/Key-Elements-of-a-Treaty-on-Fully-Autonomous-WeaponsvAccessible.pdf. See also, Human Rights Watch, 'New Weapons, Proven Precedent: Elements of and Models for a Treaty on Killer Robots' (2020) www.hrw.org/sites/default/files/media_2020/10/arms1020_web.pdf.
[30] Campaign to Stop Killer Robots, 'Key Elements of a Treaty on Fully Autonomous Weapons (2019), 3-6.
[31] Ibid., 4.
[32] Ibid., 4.

- define and validate a precise framework for the mission assigned to the system (objective, type of targets, restrictions in time and space, etc.);
- exercise their judgement with regard to compliance with IHL in the framework and context of an attack, and thus take critical decisions over the use of force.[33]

If debates on military AI are to continue to advance calls for MHC in order to support respect for international law obligations, the limitations and shortcomings of this notion should be mitigated. Only then can steps towards operationalization of MHC and compliance with international law be promoted.

2.3. Limitations and shortcomings of the concept of MHC

Although it remains at the heart of policy and scholarly debates on AWS – and military applications of AI more generally – the notion of MHC as it is currently conceptualized presents significant shortcomings and limitations. These must be identified and addressed if the international community is to rely on this concept as a framework for dealing with autonomy in the military domain.

As MHC has primarily been discussed in the context of the CCW GGE whose mandate concerns 'emerging technologies in the area of lethal autonomous weapons systems (LAWS)',[34] discussions have mainly focused on autonomy in weapons, leaving aside the broader array of military applications of AI that also pose significant legal and ethical challenges. Moreover, there has been a tendency to center discussions on autonomy in 'critical functions', namely the identification, selection, and engagement of targets.[35] As such, there has been a lack of conceptualization of MHC in contexts other than AWS, where AI technologies also transform the ways in which warfare is conducted and how control is typically exercised. These other applications of AI, most notably within military decision-making processes, arguably also require MHC. These applications span algorithms for target recognition,[36] decision-making aids,[37] and 'tasking' more broadly,[38] as well as for mobility and navigation capabilities,[39] including the use of AI for 'collaborative autonomy', allowing

[33] 'Franco-German contribution: Outline for a normative and operational framework on emerging technologies in the area of LAWS' (2021) https://documents.unoda.org/wp-content/uploads/2021/08/France-and-Germany.pdf, 2.
[34] Meeting of the High Contracting Parties to the CCW, 'Final report' (13 December 2019) UN Doc CCW/MSP/2019/9, para 1.
[35] 'Statement of the International Committee of the Red Cross (ICRC)', Meeting of Experts on Lethal Autonomous Weapons Systems, 13-16 May 2014, Geneva.
[36] See e.g., Amanda Miller, 'AI Algorithms Deployed in Kill Chain Target Recognition' (21 September 2021) Air Force Magazine www.airforcemag.com/ai-algorithms-deployed-in-kill-chain-target-recognition/.
[37] See e.g., Karel van den Bosch and Adelbert Bronkhorst, 'Human-AI Cooperation to Benefit Military Decision Making' (2018) Proceedings of the NATO Specialist meeting on Big Data and Artificial Intelligence for Military Decision Making, STO-MP-IST-160.
[38] Defense Advanced Research Projects Agency (DARPA), 'Creating Cross-Domain Kill Webs in Real Time' (18 September 2020) https://www.darpa.mil/news-events/2020-09-18a.
[39] Vincent Boulanin and Maaike Verbruggen, 'Mapping the Development of Autonomy in Weapon Systems' (2017) SIPRI Report, 21-23.

synchronized 'swarms' of drones to communicate and operate together;[40] or the collection, prioritization and analysis of data in the context of intelligence, surveillance, and reconnaissance operations.[41] Emphasis on control over one specific military application of AI may also present limitations, given the way that numerous algorithms may be embedded into military systems and are not necessarily single discrete applications.[42] Compounding decision-making that is already complex and highly distributed in the military context, all these uses of AI form part of broader sociotechnical networks that are composed of both humans and technologies.[43]

Discussions surrounding MHC have often not taken these broad webs comprising humans and AI technologies duly into account, focusing primarily on the need for control to be exercised by individual operators over individual attacks. The underlying idea is that the ability of autonomous systems to undertake multiple attacks requiring 'timely consideration of the target, context and anticipated effects' threatens the ability of military personnel to exercise control meaningfully.[44] This rightly highlights the difficulties of ensuring meaningful human oversight yet fails to recognize the need for control to be exercised beyond just in the course of executing attacks. Limiting the need for control to the operation of military engagements centers both the operator and the individual attack in a targeting 'loop', when in reality targeting processes are much more complex and multidimensional, involving a number of different actors across time and space. Decision-making on targeting in the military is distributed, with the actual control exercised by an operator being at times limited, particularly in circumstances where situational awareness is reduced, such as in poor weather conditions.[45] It is for this reason that Ekelhof suggests that MHC does not fully capture the operational realities of military decision-making in which 'various human beings exercise different forms of control at various junctures' in the targeting process.[46] It is questionable whether, due to time pressure, stress, fear, and biases experienced by humans executing attacks as part of dynamic targeting practices, as well as the fact that automated and fire and forget

[40] James Johnson, 'Artificial Intelligence, Drone Swarming and Escalation Risks in Future Warfare' (2020) 165 The RUSI Journal 26; David Hambling, 'Israel's Combat-Proven Drone Swarm May Be Start Of A New Kind Of Warfare' Forbes (21 July 2021) www.forbes.com/sites/davidhambling/2021/07/21/israels-combat-proven-drone-swarm-is-more-than-just-a-drone-swarm/.
[41] See e.g. US DARPA Project Maven, which uses computer vision to process and analyse huge amounts of drone footage, a task ordinarily undertaken by hundreds of analysists: US Deputy Secretary of Defense, 'Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven),' Memorandum (26 April 2017), https://www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf; Cheryl Pellerin, 'Project Maven to Deploy Computer Algorithms to War Zone by Year's End' (21 July 2017) US Department of Defense News https://dod.defense.gov/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/; Merel Ekelhof, 'Lifting the Fog of Targeting: "Autonomous Weapons" and Human Control through the Lens of Military Targeting' 71 Naval War College Review 62, 77.
[42] Elke Schwarz, 'Autonomous Weapons Systems: Artificial Intelligence, and the Problem of Meaningful Human Control' (2021) V The Philosophical Journal of Conflict and Violence 53, 57.
[43] See e.g. Lucy Suchman and Jutta Weber, 'Human–Machine Autonomies' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems* (Cambridge University Press 2016) 75–102, 98–102.
[44] Article 36, 'Killer Robots: UK Government Policy on Fully Autonomous Weapons' (2013), https://article36.org/wp-content/uploads/2020/12/Policy_Paper1.pdf, 4.
[45] Merel Ekelhof, 'Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation' (2019) 10 Global Policy 343, 345.
[46] Ibid., 347.

systems are already being deployed, MHC is really exercised during attacks today.[47] For static targeting in particular, human control exercised when the target is approved may be more relevant than when the mission is executed.[48] In light of the specificities of AI technologies, it may be argued that critical decision-making takes place at an earlier stage than execution of attacks, when these systems are deployed.[49] Current approaches to MHC also fail to recognize the need for the exercise of human judgment at other stages and levels, including design and development, as well as strategic decision-making. Accounting for the legal and ethical challenges of military AI at the pre-deployment stages, discussed in more detail below,[50] has important implications for how human control can be meaningfully exercised over a range of military applications of AI, not limited to AWS.

The inherent characteristics of modern AI technologies also call into question the extent to which humans can actually exercise control meaningfully. Human oversight in the form of deactivation or override functions is no panacea, as such forms of human control may be merely perfunctory due to the speed and complexity at which AI functions and the very limited cognitively actionable information that may be available. The 'black box' and self-learning character of machine learning algorithms in particular may present obstacles for humans to understand and predict how and why a system generated a specific output. This is due to the fact that these algorithms are driven by huge amounts of data, complete tasks through non-deterministic methods that are qualitatively different to that of human beings in terms of complexity and scale and are largely opaque, even to experts.[51] Furthermore, concerns that 'automation bias' may lead operators of AI technologies to over-trust autonomous systems and fail to subject output to sufficient scrutiny have been well documented.[52] The cognitive barriers to reasoned decision-making that come about as a result of reliance on AI systems may thus impact the exercise of MHC. Accounting for the characteristics of AI – including its opacity, complexity, speed, scale, unpredictability, and potential biases – therefore requires accounting for human control in the design of these technologies and how humans interact with them. Current approaches to MHC often frame a binary dichotomy between the autonomy of a system and the control of the human operator,[53] failing to recognize that human-machine relationships and military processes are complex, distributed, intermediated, and

[47] Missy L Cummings, 'Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification?' (2019) 38 IEEE Technology and Society Magazine 20, 24.
[48] Ibid. 24.
[49] Mark Roorda, 'NATO's Targeting Process: Ensuring Human Control Over (and Lawful Use of) "Autonomous" Weapons' in Williams Andrew P and Scharre Paul D (eds), *Autonomous systems: issues for defence policymakers* (NATO HQ SACT 2015) 152-168, 162-163.
[50] See below Section 4.2.
[51] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1, 4-5.
[52] Missy L Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems' (AIAA 1st Intelligent Systems Technical Conference, Chicago, Illinois, 20-22 September 2004) https://arc.aiaa.org/doi/10.2514/6.2004-6313.
[53] See, e.g., Daniele Amoroso, '*Jus In Bello* and *Jus Ad Bellum* Arguments against Autonomy in Weapons Systems: A Re-Appraisal' (2017) QIL 5, 11; Linell A Letendre, 'Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers' (2020) 96 International Law Studies 274, 281.

multidimensional.[54] These relations must be accounted for when assessing the relevance of the concept of MHC and overcoming the difficulties posed by military applications of AI.

3. Refining (meaningful) human control in relation to the international legal framework

3.1 Nature and purpose of the notion of MHC

Whilst MHC plays a central role in existing debates on AWS and has the potential to inform discussions around AI in the military more broadly, there is a lack of consensus regarding the nature and purpose of this notion. Some consider it a means to achieve legal compliance, without necessarily being an existing or emerging principle of international law itself.[55] As such, many States hold the view that 'control over critical targeting functions is necessary for compliance with international law'.[56] Nevertheless, it is unsettled whether MHC is an existing or emerging international legal standard. There is also disagreement in scholarship as to the status of MHC. Horowitz and Scharre point to the divergent positions on whether MHC is 'a principle for the design and use of weapon systems in order to ensure that their use can comply with [IHL]', or alternatively, whether it should constitute a new legal obligation as the existing legal framework is not sufficient to deal with emerging autonomous technologies.[57] Reflecting a third position, Marauhn suggests that a contextual application of the existing framework of IHL is sufficient to address autonomy in weapon systems, with MHC failing to add any additional normative requirements.[58] McFarland and Galliott go further to say that recourse to MHC is based on false premises relating to both technology and the law, rather arguing that autonomy itself is a function of human control and that the existing legal framework of IHL already operates to preserve human control.[59] Crootof, on the other hand, views MHC as a useful tool to supplement the requirements of existing law, with IHL

[54] UNIDIR's 'Iceberg' infographic illustrates that military conduct and decision-making with respect to targeting goes far beyond mission execution, including planning across the strategic, operational and tactical levels: UNIDIR, 'The human element in decisions about the use of force' (2020) https://unidir.org/sites/default/files/2020-03/UNIDIR_Iceberg_SinglePages_web.pdf. See further below, Section 4.1.

[55] See e.g. 'Categorizing lethal autonomous weapons systems – A technical and legal perspective to understanding LAWS', Working paper submitted by Estonia and Finland (24 August 2018) UN Doc CCW/GGE.2/2018/WP.2, para 13; 'Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', Working paper submitted by the United States (28 August 2018) UN Doc CCW/GGE.2/2018/WP.4 para 48

[56] GGE LAWS, 'Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', (2019), para 23; Chair of the GGE LAWS, 'Commonalities in national commentaries on guiding principles' (2020), para 8. For instance, Finland has suggested that assessing compliance with IHL needs to be an ongoing process and will determine 'how the human element materialises in relation to a machine' in specific operational circumstances: 'Elements for possible consensus recommendations Contribution by Finland' (June 2021) https://documents.unoda.org/wp-content/uploads/2021/06/Finland.pdf, 1.

[57] Michael C Horowitz and Paul Scharre, 'Meaningful Human Control in Weapon Systems: A Primer' (2015) CNAS Working Paper, 7.

[58] Thilo Marauhn, 'Meaningful Human Control – and the Politics of International Law' in Wolff Heintschel von Heinegg and others (eds) *Dehumanization of Warfare: Legal Implications of New Weapon Technologies* (Springer, 2018), 216-217.

[59] Tim McFarland and Jai Galliott, 'Understanding AI and Autonomy: Problematizing the Meaningful Human Control Argument against Killer Robots' in Jai Galliot and others (eds) *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (OUP 2021) 41-56, 51.

standards serving as an 'interpretative floor' whereby any conception of MHC that risks the safety of soldiers and civilians in conflict with existing IHL requirements cannot be accepted.[60]

MHC is further regarded by some as a mechanism through which to ensure there is responsibility for violations of international law caused with the use of AWS. Here, technology is conceived of as a tool in the hands of actors bound by the accountability framework in which they are operating.[61] The purpose of MHC is thus to ensure responsibility by requiring that targeting decisions be made by humans in real time, and subject to continuous monitoring by human operators.[62] From this view, 'pre-programmed instructions' are not considered sufficient for MHC; humans should remain 'decisively influential' in targeting practices and the decision to use lethal force.[63]

Questions on the nature and purpose of MHC that arise in debates on AWS are also relevant when reflecting on military applications of AI more broadly. As data-driven algorithms are embedded into military decision-making processes – including but not limited to targeting processes – and digital modes of warfare become more prevalent, there is a need to critically engage with the concept of MHC and assess its relevance to military AI. Indeed, the concept of MHC appears to more broadly reflect concerns about the increasingly remote link between human decision-making relying on AI and the resulting conduct. From a philosophical perspective, it has been used to conceptualize design requirements that would ensure that conduct can be tracked back to human decision-making and reasoning.[64] Yet, questions remain about the connection between the concept of MHC and the existing framework of international law. Until and unless MHC becomes enshrined in an international legal instrument as a stand-alone principle, the utility of this concept lies in its function of supporting existing legal frameworks regulating armed conflict. Clarifying the connection between MHC and international law is a therefore a necessary step towards promoting legal compliance when it comes to military AI.

3.2. Human control, human agency, and responsibility in the international legal framework regulating armed conflict

As the nature and purpose of MHC in relation to the existing international legal framework remains to a certain extent unsettled, it is worth diving deeper into the theoretical underpinnings of this concept and its connection with the international norms applicable to the conduct of warfare, including when using military AI. In the authors' view, human agency and responsibility are central

[60] Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal 53, 62.
[61] Heather M Roff and Richard Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (2016) Briefing paper, 6. See also, Thompson Chengeta, 'Defining the Emerging Notion of "Meaningful Human Control" in Weapon Systems' (2017) 49 New York University Journal of International Law and Politics 833, 869.
[62] Thompson Chengeta, 'Defining the Emerging Notion of "Meaningful Human Control" in Weapon Systems' (2017) 49 New York University Journal of International Law and Politics 833, 838-839.
[63] Ibid., 878.
[64] Filippo Santoni de Sio and Jeroen van den Hoven, 'Meaningful Human Control over Autonomous Systems: A Philosophical Account' (2018) 5 Frontiers in Robotics and AI 15.

notions that underlie and enable the fulfilment of international law norms. However, the emergence of data-driven algorithms in the military domain fundamentally alters the way in which armed conflicts are conducted and problematizes the exercise of human agency, such that the role of the human in the fulfilment of international obligations is no longer assured. From this perspective, (meaningful) human control can be considered a means through which to preserve human agency and responsibility in warfare and thus promote compliance with the relevant international norms.

Legal and moral responsibility in traditional Western thinking are grounded in the notion of human agency, which, although is a contested concept,[65] generally captures the capacity of individuals to act for reasons.[66] The notion of a responsible agent is also reflected in the international norms that bind states during armed conflict. In some instances, international obligations are explicitly directed towards individuals who launch attacks, such as the duty to take feasible precautions to verify the military character of objectives, avoid or minimize harm to civilians and ensure attacks are proportionate.[67] In others, the presence of a responsible human agent is reflected in the connection between the fulfilment of duties owed under the laws of armed conflict and ethical considerations, whereby the application of legal norms on the ground has a 'situated subjectivity'[68] and is inherently linked to value-laden judgments,[69] which can arguably only be made by human beings.

The principle of distinction is a prime example, requiring an assessment that weighs and balances two qualitatively different factors, namely whether the expected collateral damage would be excessive in light of the anticipated military advantage.[70] It has been highlighted that '[m]uch has been written on the indeterminacy of the principle of proportionality and on the unworkable test of comparing the incommensurable values of military advantage and civilian lives.'[71] Concerns have been raised about the quantification of such considerations for the purposes of proportionality assessments when it comes to autonomous systems. For instance, the US National Security Commission on Artificial Intelligence (NSCAI) stated in a 2021 report that '[t]he moral reasoning involved in this calculus [...] remains the responsibility of a human commander'.[72] The ICRC also asserts the connection between agency, morality, and the laws of armed conflict, suggesting that

[65] Mireille Hildebrandt, 'Introduction: A multifocal view of human agency in the era of autonomic computing' in Mireille Hildebrandt and Antoinette Rouvroy (eds) *Law, Human Agency and Autonomic Computing* (Routledge, 2011) 1-11, 5.
[66] Stephen J Morse, 'Determinism and the Death of Folk Psychology: Two Challenges to Responsibility from Neuroscience' (2008) 9 Minnesota Journal of Law, Science and Technology 1, 20.
[67] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977 ('AP1'), Art. 57(2)(a)(i)-(iii). See also, Nikolas Stürchler and Michael Siegrist, 'A "Compliance-Based" Approach to Autonomous Weapon Systems" (EJIL: Talk!, 1 December 2017) www.ejiltalk.org/a-compliance-based-approach-to-autonomous-weapon-systems.
[68] Ioannis Kalpouzos, 'Double Elevation: Autonomous Weapons and the Search for an Irreducible Law of War' (2020) 33 Leiden Journal of International Law 289, 311.
[69] Frédéric Mégret, 'Theorizing the Laws of War' in Anne Orford and Florian Hoffmann (eds) *The Oxford Handbook of the Theory of International Law* (OUP, 2016) 762-778, 767, 772.
[70] Art 51(5)(b) AP1; Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law, Volume I, Rules* (Cambridge University Press 2005), 46-50 (Rule 14).
[71] Gabriella Blum, 'On a Different Law of War' (2011) 52 Harvard Journal of International Law 163, 189.
[72] NSCAI, 'Final Report' (1 March 2021), 92. See also, Jeroen van den Boogaard, 'Proportionality and Autonomous Weapons Systems' (2015) 6 Journal of International Humanitarian Legal Studies 247, 266.

the need to preserve human agency in warfare 'derives from broader ethical considerations of humanity, moral responsibility, human dignity and the dictates of public conscience'.[73] The intrinsic and necessary agency of human actors with the capacity for reasoned conduct is further reflected in the standard to which those who assess proportionality when conducting attacks are held, namely that of a reasonable commander.[74] As such, the framework of international law regulating armed conflict is not solely an effects based regime; how warfare is conducted matters. Accounting for the human addressees of international law, as well as the ethical considerations that must inform battlefield decision-making and the applicable standards of reasonableness to which military personnel are held is therefore necessary in the development and use of AI systems that transform how warfare is conducted and alter the role of the human in this context.

The adoption of AI in the military affects and problematizes how human agency is traditionally understood and exercised. Data-driven algorithms compound the spatial and temporal distance between those launching attacks and the effects of these attacks on the ground by further distributing decision-making processes. Additionally, these algorithmic techniques create a further cognitive barrier experienced by humans when decision-making becomes abstracted from temporally earlier decisions in geographically distant places, whether of those involved in the acquisition of these technologies, the programmers encoding it, or the military commanders planning attacks. As discussed above, the opaque, self-learning character of algorithms that constitute data-driven techniques of AI such as machine learning alter the role of humans in the military domain. More specifically, AI alters the competencies of human beings through the delegation and automation of existing and new tasks, as well as the capacity individuals ordinarily possess to undertake certain tasks. This may create additional barriers to *reasoned* decision-making that is genuine and informed on the part of military personnel, limiting human agency at both the micro level of specific attacks and at the macro level of the wider military organization through the reduction of conduct in military affairs to datapoints. As put by Schwarz, 'where the moral stakes are high, such as in warfare, being able to predict and understand the decision system is important'.[75] These challenges to the exercise of human agency when AI technologies are relied upon may further be exacerbated by biases that may manifest in the data, design, or outcomes of a system, as well as the human interacting with it.[76]

---

[73] ICRC, 'Artificial intelligence and machine learning in armed conflict: A human-centred approach' (2020) 102 International Review of the Red Cross 463, 474.
[74] Frits Kalshoven and Liesbeth Zegveld, *Constraints on the Waging of War* (ICRC 2001), 109; Geoffrey S Corn, 'Humanitarian Regulation of Hostilities: The Decisive Element of Context' (2018) 51 Vanderbilt Journal of Transnational Law 763, 765.
[75] Elke Schwarz, 'Autonomous Weapons Systems: Artificial Intelligence, and the Problem of Meaningful Human Control' (2021) V The Philosophical Journal of Conflict and Violence 53, 61. Schwarz argues that the 'technological features and the underlying logic of the AI system progressively close the space as limit the capacities required for human moral agency', 54.
[76] See above Section 2.3.

As the exercise of human agency is challenged and eroded in the context of AI-powered warfare, this has bearings on how responsibility can be established. When violations of international law are committed as a result of the use of AI-enabled autonomous systems, the transformed role of humans impacts in particular the assignment of individual responsibility.[77] As argued by Bo, ascribing individual responsibility under international criminal law becomes difficult or impossible, due to an insufficiently genuine subjective intent (*mens rea*) that could be traced to the individual operator.[78]

Understanding how the exercise of human agency relates to the fulfilment and implementation of international law by the military thus provides theoretical grounds for seeking to ensure (meaningful) human control over AI-enabled tools adopted in the military. The notion of MHC thus emerges as a conceptual lens that could contribute to preserving human agency, and thereby responsibility, in a setting where it can no longer be taken for granted. The next section specifically addresses ways in which human control can be operationalized to these ends.

4. Operationalizing (meaningful) human control

4.1. Recognizing the complex and distributed nature of human-machine interaction in the military domain

As we have seen, the introduction of AI into the military changes how warfare is conducted and reshapes the roles of humans in this context. Military processes are already complex and distributed, which is compounded by reliance on AI-enabled technologies. As a result, when data-driven algorithms become embedded into military decision-making processes, human agency and responsibility may be compromised. As put by Brehm, '[i]n addition to constraints based on ethical and other imperatives, compliance with the law presupposes a measure of human agency in the use of force that places limitations on permissible "human-machine configurations"'.[79] As these technologies become integrated into military architectures that comprise both humans and machines, human conduct becomes mediated in a way that is no longer self-evident, with the role of the human no longer assured. The complex reality of AI technologies in the military domain suggests that there is no clear proportional relationship between human control and autonomy. Rather, recognizing the dynamic interaction between AI-enabled systems and human agents provides a more appropriate frame of reference for the challenges posed by these technologies.

[77] See, generally: Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175.
[78] Marta Bo, 'Autonomous Weapons and the Responsibility Gap in light of the *Mens Rea* of the War Crime of Attacking Civilians in the ICC Statute' (2021) Journal of International Criminal Justice 1.
[79] Maya Brehm, 'Defending the Boundary: Constraints and Requirements on the Use of Autonomous Weapon Systems Under International Humanitarian and Human Rights Law' (2017) Geneva Academy Briefing No 9, 20.

Capturing the connection between humans and AI technologies is possible through the lens of 'human-machine interaction', which originates from the fields of philosophy of science and computer design.[80] This notion has recently emerged in the debates on AWS as a potential way forward to ensure and safeguard human control. The 2019 GGE Guiding Principles affirmed that a certain 'quality and extent of human-machine interaction' is necessary to ensure that autonomous systems can be used in compliance with international law.[81] The progressive shift in the debate from focusing on MHC to human-machine interaction could be explained by an increased recognition that technological mediation affects human decision-making in many ways, and therefore binary notions of control and delegation do not sufficiently grasp the complexity and nuances of the relationship between human decision-making and machine autonomy. The theory of technological mediation essentially suggests that technological tools mediate our relationship to the world, and our use of technologies shapes and impacts human processes of decision-making.[82] However, human-machine interaction need not be divorced from our understandings of MHC. Rather, it may serve as a lens to view the complex interactions between human actors and various applications of AI mediating their conduct in the military domain. In this way, human-machine interaction may provide a useful analytical framework to assess whether MHC – and therefore legal compliance – can be ensured.

Importantly, reflecting on human-machine interaction in the context of military AI allows us to broaden the scope of analysis and overcome the limitations presented by MHC alone. First, it acknowledges that the complexity, scale, and speed of AI impacts human control and goes beyond human cognitive capabilities. If a machine-learning algorithm pre-selects a target on the basis of millions of data points analyzed in a split-second, the role of human control or supervision becomes superficial. When soldiers are faced with the decision of whether or not to follow an algorithmic recommendation to engage a particular target, they will have limited cognitively actionable information to act upon in a time-critical environment. Second, it sheds more light on the relevance of the pre-deployment stages with respect to the development and acquisition of military AI. When we design military applications of AI, we are likewise constructing humans' competencies and capacities through the ways in which humans interact with the system.[83] As the performance of international obligations owed during armed conflict is inherently connected to the exercise of

[80] Batya Friedman and Peter H Kahn, 'Human Agency and Responsible Computing: Implications for Computer System Design' (1992) 17 Journal of Systems and Software 7, 10.
[81] Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (25 September 2019) UN Doc CCW/GGE.1/2019/3, Annex IV, Guiding Principle (c). See also, United Kingdom, 'Expert paper: The human role in autonomous warfare' (18 November 2020) UN Doc CCW/GGE.1/2020/WP.6.
[82] Peter-Paul Verbeek, 'Toward a Theory of Technological Mediation: A Program for Postphenomenological Research', in Jan Kyrre Berg O. Friis and Robert P. Crease (eds), *Technoscience and Postphenomenology: The Manhattan Papers* (2016), 189-204.
[83] Schwarz suggests that with respect to AI, human 'modes of reasoning are shaped along the techno-logics of algorithmic data processing': Elke Schwarz, 'Autonomous Weapons Systems: Artificial Intelligence, and the Problem of Meaningful Human Control' (2021) V The Philosophical Journal of Conflict and Violence 53, 55ff.

human agency and responsibility,[84] the design of AI technologies thus has an impact on how States can implement military AI in compliance with international law. Human control measures must therefore be incorporated into the design and testing of military AI, in order to engineer human agency and responsibility into an architecture where these can no longer be taken for granted. To safeguard the possibility of MHC, military AI systems should be designed and developed in a way that, for instance, reduces automation bias and supports human cognitive functions. Indeed, as will be discussed in the next section, more emphasis on the design, development and testing of these technologies is crucial in operationalizing MHC.

4.2. Ensuring MHC at the pre-deployment stages

The interaction between humans and AI technologies is not limited to the moment the technology is used to engage targets on the battlefield, but indeed begins and is designed at the pre-deployment stages. This takes place at various moments, between various actors, and at various levels. Whilst, as explained above, discussions around MHC have centered on the need for control by an individual operator over an attack, the distributed nature of military decision-making and complexities of AI-enabled technologies requires us to pay closer attention to the pre-deployment stages. Recent interventions in the debate on AWS have acknowledged the need to take into account various stages, spanning '(0) political direction in the pre-development phase; (1) research and development; (2) testing, evaluation and certification; (3) deployment, training, command and control; (4) use and abort; (5) post-use assessment.'[85] While the GGE LAWS has moved towards recognizing the importance of human-machine interaction across the 'life cycle' of a system,[86] it has not yet further elaborated on how to operationalize MHC in each of these stages. When it comes to military applications of AI, some elements of critical human decision-making may be relocated at the earlier stages of conception, testing, acquisition, and deployment. Indeed, various human touchpoints can cumulatively assist in reaching a sufficient level of control. Identifying how, in practical terms, MHC can be ensured at these pre-deployment stages thus constitutes a crucial step forward in determining how military AI can be compliant with the framework of international law by design.

This idea of compliance with international law and MHC by design builds upon approaches in ethics of technology, such as value-sensitive design and responsible innovation, which seek to identify and integrate ethical values within the design and development of technologies.[87] By analogy, it can be argued that military technologies should be designed and developed in ways that reflect

[84] See above, Section 3.2.
[85] 'Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (23 October 2018) UN Doc CCW/GGE.1/2018/3, para 23.
[86] Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (25 September 2019) UN Doc CCW/GGE.1/2019/3, Annex IV, Guiding Principle (c).
[87] Jeroen van den Hoven, 'Value Sensitive Design and Responsible Innovation' (2013), in R. Owen, J. Bessant, M. Heintz (eds.) *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, 75–83

international obligations and enable compliance. As expressed by Letendre, 'legal requirements for the use of force cannot be an after-thought in the development of autonomous weapon systems; it must be built-in from the beginning.'[88] In this sense, compliance with international law can be promoted by integrating legal standards as system requirements. This forward-looking approach would further help in minimizing the risk of occurrence of violations of international law once such technologies have been deployed. Encoding the law into AI systems certainly has limits, as context-based principles such as proportionality are difficult to quantify and reduce to code.[89] Nevertheless, it is worth taking the effort to seek to integrate certain ground principles, values, and requirements for control into the design of AI systems.

Operationalization of MHC at pre-deployment stages must cut across different stages and levels. Touchpoints for implementing human control measures begin with the state policy decisions to acquire, through either development or procurement, military AI technologies. At the stage of development of these technologies, processes in which such measures should be implemented include ideation and design, research and development, and testing and certification. Design choices at each stage should take into account human-machine interaction and seek to safeguard humans' ability to exercise critical judgment. In the case of procurement of technologies from third parties, testing procedures, including evaluation and certification of MHC in the system's design, must also take place.[90] Such evaluation procedures could notably take place in the context of Article 36 AP1, although it is not clear whether military applications of AI that are outside or only indirectly related to targeting processes qualify as 'weapons, means and methods' of warfare for the purpose of Article 36 review.[91]

Within the military sphere, there are a number of processes outside of the deployment of weapons and execution of attacks that require human control measures. This includes in particular the training of military personnel to raise awareness as to the possible limits of control at the stage of deployment and attacks, and the complexities of human-machine interaction in practice. Training should also cover a certain amount of technical understanding of AI systems, their capabilities, parameters, and limits. This would contribute to reinforcing the capacities of operators and commanders to make informed decisions once AI systems are deployed. The availability and understanding of information regarding the technical features and reliability of a given system in terms of MHC is of particular importance for military commanders. Indeed, in the context of AI

[88] Linell A Letendre, 'Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers' (2020) 96 International Law Studies 274, 275.
[89] Ashley Deeks, 'Coding the Law of Armed Conflict: First Steps' (2020) Virginia Public Law and Legal Theory Research Paper No. 2020-49 (SSRN); Alan L. Schuller, Artificial Intelligence Effecting Human Decisions to Kill: The Challenge of Linking Numerically Quantifiable Goals to IHL Compliance (2019) 15 I/S: A Journal of Law and Policy for the Information Society 105-122.
[90] See also, Cummings, who argued for requirements of 'meaningful human certification'. Missy L Cummings, 'Lethal Autonomous Weapons: Meaningful Human Control or Meaningful Human Certification?' (2019) 38 IEEE Technology and Society Magazine 20, 25.
[91] Klaudia Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare' (2022) 23 Yearbook of International Humanitarian Law 123.

systems, the duty of commanders to act reasonably translates into a duty to seek information on the design of systems that are being deployed.

In order to further operationalize MHC certification and verification procedures, it will be necessary to move outside the legal realm and into technical standards. A degree of collaboration between policy and industry is required here to develop processes and standards that can assess the technical capabilities, reliability, and parameters of a system in terms of MHC and human-machine interaction.

5. Concluding remarks

The integration of AI into the military domain represents a shift in the way that warfare is conducted. Armed conflicts will increasingly become digitalized through the use of AI, producing novel relationships between humans and technologies, and impacting the ways in which control is typically understood to be exercised. The concept of MHC has been conceptualized so as to support or supplement the existing framework of international law in addressing these new challenges. Yet, if this concept is to move forward as a cardinal principle in the regulation of military applications of AI, it needs to be refined in legal terms and operationalized in technical and policy terms. Failure to do so will run the risk of MHC comprising an empty shell, propped up by states seeking to develop and deploy these technologies in an unimpeded manner. Compliance with international law and MHC by design with respect to military AI technologies is an approach that can assist in this pursuit, acknowledging that the inherent characteristics of AI and the new types of interactions with human beings that result need to be accounted for when these technologies are designed and developed. Precisely how to integrate compliance with international law through design choices is an area requiring further research. Inquiry into the feasibility of embedding legal norms into algorithmic formats is needed to assess from both a legal and technical perspective whether international standards governing armed conflict are amenable to a 'legal compliance by design' approach. Only then can we advance in determining whether particular applications of military AI are compatible with the international legal framework.