IDENTIFYING 'HUMAN DIGNITY' WITHIN DEPLOYMENT OF LETHAL AUTONOMOUS WEAPON SYSTEMS

Shagufta Sen, Ph.D., Faculty of Law, University of Delhi

ABSTRACT

Advancements in means and methods of warfare have continually created military revolutions, leading to the changing nature of war. The advent of Artificial Intelligence has paved the path for the creation of lethal autonomous weapon systems(AWS), which can identify and engage in targets without human intervention. It is the very aspect of the removal of humans from warfare and the development of automated decision-making that has raised considerable concerns within the international community. The spectrum of autonomy, which defines machine independence has remained a defining factor in underlining the reaction of human control within the operation and use of AWS. The international community continues to debate upon the very aspect of conceptualising these weapon systems and negotiating an appropriate regulatory framework but has continually failed to understand the essential human rights implications, given that these weapon systems certainly pose catastrophic risks. The human rights argument remains centre stage even after establishing the fact that AWS would remain in compliance with humanitarian law and its principles. The paper considers the human rights argument for analysing and understanding how 'human dignity' can continue to be preserved if appropriate human control is exercised during the deployment of the weapon system.

Keywords: autonomous weapon systems, autonomy, meaningful human control, human rights, human dignity, dehumanisation

I. Introduction

The revolution in military affairs (RMA) has changed the outlook on warfare with the introduction of smart weapons such as , drones, cyber weapons and now autonomous weapons. The emergence of lethal autonomous weapon systems or AWS, has introduced the usage of Artificial Intelligence within the military domain, highlighting that weapon technology is headed towards a revolution of removing the 'human element' from warfare. The removal of 'humans' from warfare also stems from the casualty aversion States have begun to have due to the number of casualties usually suffered in an armed conflict.

The ability of AWS to identify and engage in targets without any human intervention has paved the way for removing 'humanness' at the deployment level. Autonomy or machine independence is the functional constraint responsible for making decisions with respect to identifying and engaging in targets. A 'human-out-of-the-loop' model of autonomy highlights the removal of 'humans' from warfare leading to dehumanisation, and within the context of automated decision-making(ADM), it has been described as the usage of cruel weapons and the withdrawal of human beings from the battlefield. [1] From an interdisciplinary point of view, the term implies undermining of 'human dignity', 'humanity' or emotional desensitisation of a human operator on the loop (HOTL). Dehumanisation when read in conjunction with anthropomorphisation, highlights the removal of humanness from warfare, equipping a machine with the sense of conducting hostilities as per its whim or the programmed 'decision-making process'. It is multilayered because the perception of human control has been highly linked with the threshold of autonomy. [2] Even with the complex and interconnected relationship between 'human control' and autonomy, the threshold of autonomy of the weapon systems is a determiner of how much human control may be necessary in intervening in the decision-making process, if a violation of IHL may occur from such targeting decision.

The dehumanisation of warfare in the context of autonomous weapon systems is not limited to the 'removal of humans' or 'combatants' from warfare but moves beyond this premise of exercising or integrating human control, raising ethico-legal and moral concerns. The widening responsibility and accountability gap is attributed to the reduced cognitive involvement of

[1] Jan C. Jordan, "Dehumanization: The Ethical Perspective" in Wolff Heinschel von Heinegg *et al.* (eds.), *Dehumanisation of Warfare: Legal Implications of New Weapon Technologies* (Springer 2018).
[2] Wolff Heinschel von Heinegg *et al.* (eds.), *Dehumanisation of Warfare: Legal Implications of New Weapon Technologies* (Springer 2018).

human agents, complicating the process of assigning responsibility to an 'individual' based on existing criminal law principles.

It should be inferred from this term that the human operator/agent would not become cruel or inhumane but would be pushed into the background of decision-making processes in warfare.

II. Understanding the relevance of meaningful human control

'Meaningful Human Control' or MHC doesn't have a fixed definition but has been understood to be the 'human agency' or appropriate human control that is required to be exercised during the deployment of an autonomous weapon system (AWS). MHC has emerged as the concept which draws the distinction between lawful and unlawful uses of autonomous weapons, while also underscoring the need for maintaining relevant 'human supervision' during the operation of the weapon system. Human control can also be understood in the context of the level of human judgement required in executing a particular targeting or use of force decisions. Having 'meaningful' human control also suggests that it should be an explicit condition that humans are supervising the operation of such weapon systems, irrespective of the level of autonomy they have been deployed. When understanding from a development and design perspective, integration of human control would not only govern the operation of the weapon system but also define the limitations of design, delineating the functionalities wherein full autonomy can be exercised or appropriate human control would be established.

From a legal point of view, human control or MHC has to be understood in the wider context of IHL rules and principles, wherein respecting 'human dignity' or the human right to life remain central. For instance, compliance with the principle of distinction is necessary to ensure that no civilians are harmed during an attack while the principle of proportionality highlights that the military attack should not be in excess of the military objective to be achieved. [3] Given the importance of humanitarian law, it should be pointed out that the Group of Governmental Experts (GGE) on Lethal Autonomous Weapon Systems established in 2019 within its Guiding Principles that these weapon systems would continue to comply with IHL rules and regulations.[4] Hence, from a humanitarian lens it is essential that these weapon systems would

[3] International Committee of the Red Cross (ICRC), *Proportionality.*

[4] Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Annex III, 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/MSP/2019/9.

go through rigorous review procedures as established within the Geneva Conventions of 1949 an only then be deployed in the battlefield. But the argument of human rights continues to be mulled over, as legal review process would not be adequate enough to assess the human rights implications of development and deployment of these weapon systems. Hence, it is essential to understand the status of AWS in context of human rights, especially the right to life and concept of 'human dignity', as it has been pointed out before that deployment of AWS runs the risk of dehumanising warfare.

III. AWS within a human rights context

There are serious concerns raised by the development and deployment of lethal autonomous weapons systems, especially within the context of human rights. The aspect of violations of humanitarian law principles has continued to be debated. Still, it has been established that these weapon systems would continue to comply with the current IHL rules. The debate regarding the violation of human rights has remained an underdeveloped portion. The fundamental rule of international human rights is that no one may be arbitrarily deprived of his or her life. Article 3 of the UDHR upholds this right to "life, liberty and security of person".[5] Article 6 of ICCPR, provides that: 'Every human being has the inherent right to life. This right shall be protected by law, no one shall be arbitrarily deprived of his life.[6] [7]

Given the extent of this right, it is a peremptory norm and any derogation from it would call for accountability on the part of the State violating the same. Hence, within the context of autonomous weapon systems, deprivation of life by mechanised and autonomous targeting and decision-making may pave for the deprivation of this right, dehumanising warfare eventually.

The violation of the right to life is the first instance of dehumanisation that the development and deployment of autonomous weapon systems may create. Further, imbuing weapon systems with human-like qualities is the second instance that would lead to the dehumanisation of warfare, as the involvement of human operators is 'supposedly' going to be reduced. The right to life and liberty is inclusive of human dignity, and violation of it highlights that the human-centric approach to warfare would eventually dwindle as the world moves towards a more post-

[5] Universal Declaration of Human Rights, G.A. Res. 217A (III), U.N. Doc. A/810 at 71 (1948).
[6] International Covenant on Civil and Political Rights, art. 4(2).
[7] Human Rights Committee, General Comment No. 29 on States of Emergency (31 August 2001), para. 7; Human Rights Committee, General Comment No. 6 on the Right to Life (30 April 1982), paras. 1–3.

heroic age.

IV. Identifying the right to human dignity

From a legal point of view, warfare has continued to be human-centric even with advancements in technology and conducting drone strikes, and the aspect of dehumanisation has only been understood in the context of the elimination of humans during warfare and not within the purview of advanced weapon technology. Humans continue to be 'in-the-loop' and 'on-the-loop' of conducting operations involving dangerous means and methods of warfare. It is only with the emergence of autonomous webcam systems that the contentious issue of 'dehumanisation of warfare' cropped up.

From an ethical perspective, the power of the decision-making process within machines would raise responsibility and accountability gaps, and giving machines the ability to make life-and-death decisions certainly undermines human dignity. [8] The argument with respect to the violation of human dignity through the deployment of autonomous weapon systems lies in the fact that even if there is a 'human-on-the-loop' there remains the possibility of permeation of automation complacency. The unbridled trust within the decision-making and functionality of the weapon system would certainly lead to the emotional disengagement of the human operator in supervisory control. The argument for integrating MHC has continually been reiterated, but irrespective of that the threat to the human right to life is apparent.

Human dignity is a set of basic human rights including the right to minimum freedom of action and decisions, the right to the minimum level of quality of life and relief of offering and the right not to be treated as a means to other people's ends, clearly highlights that 'decision-making process' is central to human dignity, and delegating this power to a machine would certainly reduce humans to just algorithms.[9]

Placing human dignity within the deployment of AI-enabled weapon systems remains a contentious issue and requires in-depth analysis and deliberation. Hence, to mitigate the effects of dehumanisation that results in the undermining of human dignity, integrating 'Meaningful Human Control' within the operation and deployment of autonomous weapon systems remains

[8] Bonnie Docherty, "Heed the Call: A Moral and Legal Imperative to Ban Killer Robots", *Human Rights Watch* (August 21 2018).
[9] D. Birnbacher, "Are Autonomous Weapon Systems a Threat to Human Dignity?" N. Bhuta *et al.* (eds) in *Autonomous Weapons Systems: Law, Ethics, Policy* 105, 105–121 (Cambridge University Press 2016).

essential. MHC has continued to be characterised by the factum of increasing predictability, reliability and understandability of weapon systems during deployment. MHC is more of a judgement-based criteria that makes a subjective assessment of the fact 'when and where' human control may have to be integrated irrespective of the level of autonomy of the weapon system or at what stage the AWS is in. From a socio-technical perspective, it should be understood that MHC would create a human-machine team, balancing human-machine interaction and striking a balance between the levels of automation and relevant human control for making use of force decisions.

The spectrum of autonomy or the level of machine independence is the factor that anthropomorphises the weapon systems, providing them with the ability to make use of force decisions without any human control.

V. How does meaningful human control preserve human dignity?

The purpose of understanding meaningful human control is to create a morally and ethically compliant weapon system , which does not violate the very right to life and human dignity. Dehumanisation of warfare has supposedly been in existence for centuries due to the killing of soldiers during wars, but the power of killing delegated to a weapon system clearly poses a higher existential threat and risk. MHC has continued to be a 'guiding frame of reference' for the regulation and mitigation of the impact of the development of AWS. The aspect of meaningful within human control has been given varied interpretations such as 'effective', 'substantive' 'sufficient' and other synonymous words which provide for a wider emanating to the aspect of 'human judgement' required for maintaining a low threshold of autonomy within the deployment of the weapon system.

Irrespective of ambiguity within the definitional conceptualisation of MHC, the maximalist school of thought highlights that MHC is a standalone principle that exists outside the purview of IHL principles, while the minimalist school of thought emphasises that it is more of a 'guiding principle' that is complementary to the design and weapon usage requirements. [10][11]

[10] Michael Horowitz, Paul Scharre, "Meaningful Human Control In Weapon Systems: A Primer" *Centre for a New American Security* 7–8 (March 6 2015),

[11] Shin-Shin Hua, "Machine Learning Weapons and International Humanitarian Law: Rethinking Meaningful Human Control" 51 *Georgetown Journal of International Law* 51 (2019).

Human control is understood to increase the predictability and understandability of the weapon system; these characteristics are essentially 'metrics' that determine the fact whether the machine has been designed for the purpose of weaponeering and is also functioning as per its purpose.[12] The reason for the characteristics being termed as 'metrics is that they anticipate the outcome of the deployment of an AWS, and since they cannot be quantified, a more 'subjective assessment' of human control would have to be made.

The drawback of making a more 'subjective assessment' is that it creates a more 'inherent operational unpredictability', as weapon systems may be capable of navigating situations that a human operator cannot.[13] This 'unpredictability' may seem like a drawback to the integration of 'effective' or 'meaningful' human control, but it would objectively define the mandatory 'threshold of autonomy' during deployment of AWS in such non-navigable areas. This 'inherent' unpredictability may also create 'emotional disengagement' of the 'human-on-the-loop'(HOTL).

This emotional disengagement may further result in the permeation of both automation bias and complacency. Wherein automation bias is simply the human tendency to overlook information outputs by machines through an automated process because of the perceived notion that they are infallible and far superior to humans. The over-reliance on the functional capability of the machine highlights automation complacency, wherein a HOTL may overlook the limitations of the algorithmic outputs, posing an existential risk or threat and making the targeted decision a grave violation of human rights. Hence, over-reliance on the operational and functional capabilities of the machine would provide legitimacy to automated decision-making would make it more of a 'control problem' as there would be minimum to zero human control. The legitimisation of decisions making process of AWS would need become a 'control problem'. Such over-reliance would consequently legitimise the decision-making process of the weapon systems, which may eventually become a 'control problem'.

In addition to the subjective and objective assessments of the threshold of autonomy, integrating MHC would add a cognitive value to the understanding of more abstract

[12] Heather M. Roff & Richard Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons", Briefing Paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016.
[13]Michel Arthur Holland, The Black Box, Unlocked: Predictability and Understandability in Military AI (United Nations Institute for Disarmament Research 2020)

circumstances emerging from such deployment, for instance, the knowledge of armed conflict or a certain activity being part of an overall plan.[14] Since autonomy is a functional and technical constraint and confined to predetermined boundaries, it would also be a determiner of the level of cognitive involvement required during the deployment of the weapon system. [15] MHC has been envisioned as the adequate response to maintain control at a functional and operational level, but how can it be integrated within weapon systems remains a question.

An example of an automated decision-making system, engaging in targets autonomously even with a human supervisor, can be seen in the 'Lavender' and 'Gospel' , two ADM technologies which have recently been on the news for their autonomous capabilities. The former identifies 'human 'targets, highlighting the beginning of 'dehumanised 'warfare while the latter identifies infrastructural targets and the latter identifies. Israel has developed a reputation for utilising advanced military technologies, in 2021 it had developed a a semi-autonomous military robot,[16] which was promoted as 'one of the first many robots that can substitute soldiers on the border';[17] clearly highlighting that the replacement of combatants with machines is not far.

As per +972 magazine, 'Lavender' is a problematic automated decision-making system as the authorised personnel operating it is only provided a 'rubber stamp' of approval to execute the decisions made by the machine.[18] This not only highlights the pervasion of automation complacency but also the responsibility gap. Testing, evaluation, verification and validation (TEVV) are important processes within the assessment of an autonomous weapon system. From a design perspective, these processes highly influence the predictability, understandability and reliability of the weapon system. This approval of kill lists without examining the data, also highlights that automated decision-making systems are not capable of making distinctions in accordance with IHL obligations, as the system killed thousands of

[14] Thilo Marauhan, Meaningful Human Control – and the Politics of International Law in n Wolff Heinschel von Heinegg et al. (eds.), *Dehumanisation of Warfare: Legal Implications of New Weapon Technologies* (Springer 2018) 209.
[15] William Boothby, 'Dehumanisation: Is There a Legal Problem Under Article 36', in Wolff Heinschel von Heinegg et al. (eds.), *Dehumanisation of Warfare: Legal Implications of New Weapon Technologies* (Springer 2018).
[16] Israel Defense Forces, 'Jaguar: The IDF's Newest, Most Advanced Robot '*Israel Defense* forceshttps://www.idf.il/en/mini-sites/technology-and-innovation/jaguar-the-idf-s-newest-most-advanced-robot/ (Accessed on 3 January 2025).
[17] Access Now, 'Artificial Genocidal Intelligence: Israel's AI in Gaza '*Access Now* (27 October 2023) https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/ (Accessed on 2 January 2025).
[18] Yuval Abraham, ''Lavender': The AI Machine directing Israel's bombing spree in Gaza '+972 Magazine, April 3 2024 https://www.972mag.com/lavender-ai-israeli-army-gaza/

civilians.[19] The ability of Lavender software to learn from the training data and analyse information collected on 2.3 million Palestinians, through a system of mass surveillance and assessing their ranks or the likelihood of them being active as a Hamas operative.[20] The data used for training the ADM system and the information used for allowing the system to recognise individuals in the Gaza Strip, is a clear example of automation complacency of the human beings, as the bias that continually permeates during the data training.[21] The accuracy of the system being only 90% highlights that, in the 10% margin of error rate, the automation bias and complacency have killed thousands of people in the Gaza Strip underscoring that 'meaningful human control' in automated decision-making systems become highly important. [22] The example of these ADM technologies highlights that having HOTL is not the end alternative for maintaining then the required amount of human control, there has to be an established thorough mechanism for ensuring that MHC is maintained throughout the 'life-cycle 'of the weapon system. MHC, being a moral and ethical consideration, essentially differs from ordinary human control i.e. HITL. In the case of HITL, the machine or weapon system is fully controlled by the human operator and has zero autonomy of its own. HOTL is essential to the model of autonomy that maintains the balanced role of both humans and weapon systems at an operational level. MHC is subjective in nature and cannot be integrated like training data, it would have to be established within the features/functions of the weapon systems. A design and development approach would have to be adopted to integrate human control 'effectively'.

VI. Conclusion

Given this advancement of automated decision-making within weapon systems, it becomes clear that delegating the task of making life-and-death decisions is not possible and humans would continue to have control over weapon systems. Furthermore, even within a higher level of autonomy, the human control retained would be 'supervisory' in nature, terminating the decision when there is doubt that a violation of international human rights and humanitarian law would occur.

It should be highlighted that even though dehumanisation continues to be the existential risk, it may culminate into reality at some point if autonomous weapon are not defined in a concrete

[19] Ibid.
[20] Ibid.
[21] Ibid.
[22] Ibid.

manner. Even though autonomous weapon systems may offer an irrational advantage during warfare or conducting use-of-force decisions, there certainly remain risks which can only be mitigated if there is a concrete conceptualisation of what exactly would constitute an autonomous weapon and how its definitional elements would define the limitations regarding its usage in the theatres of war.

Furthermore, establishing adequate human control within weapon systems would be an ethical safeguard wherein the accountability and responsibility gaps can be fulfilled. Since it is an ethical and moral safeguard, it would also act as a mechanism for preventing catastrophic consequences such as targeted killings or unlawful uses of force during deployment. The anthropomorphisation of weapon systems has to be understood beyond the scope of use-of-force decisions and target selection, as varied military tasks may also highlight human-like qualities. The spatiotemporal nature of the weapon technology clearly highlights a higher level of dehumanisation and emotional desensitisation in the context of target selection and the decision-making process.

It is essential that the essence of 'human dignity' is maintained, as AI-enabled weapons having independent power to make use-of-force decisions would only further increase the gap between humans and the machines. Hence, it is essential to maintain an appropriate balance between human judgement and weapon autonomy to achieve human-machine teaming.