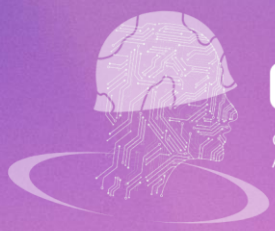

GC REAIM

GLOBAL COMMISSION ON RESPONSIBLE AI IN THE MILITARY DOMAIN

Foreign, Commonwealth & Development Office

GC REAIM Expert Policy Note Series

Restrictions on AI Weapons in Specific Situations

Mun-eon Park

April 2025

POWERED BY The Hague Centre for Strategic Studies

GC REAIM Expert Policy Note Series

Restrictions on AI Weapons in Specific Situations

Author: Mun-eon Park

April 2025

Cover photo: Unsplash

The Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM) is an initiative of the Government of the Netherlands that was launched during the 2023 REAIM Summit on Responsible Artificial Intelligence in the Military Domain in The Hague. Upon request of the Dutch Ministry of Foreign Affairs, the Hague Centre for Strategic Studies acts as the Secretariat of the Commission.

The GC REAIM Expert Policy Note Series was funded by the Foreign, Commonwealth and Development Office (FCDO) of the United Kingdom. GC REAIM Experts maintained full discretion over the topics covered by the Policy Notes. The contents of the GC REAIM Expert Policy Note series do not represent the views of the Global Commission as a whole. The Policy Notes are intended to highlight key issues related to the governance of AI in the military domain and provide policy recommendations.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl

1. Introduction

Assessing Recently, artificial intelligence (AI) in the military domain has been discussed in terms of ethics, law, policy, and technology. However, discussions have remained at the level of abstract principles and policies, not having moved beyond specific uses or prohibitions. In order to discuss more specific uses and limitations of military AI, it is necessary to identify the military applications of AI in advance.

The creation, collection and management of military data, the use of generative AI, and the use of AI for non-weapon systems are open for debate on the same basis as civilian areas. As long as uses of AI do not pose a risk to civilians or violate their rights, they can be employed. However, in the case of weapon systems, more specific regulations are needed because the killing of civilians is a potential risk.

There are two types of restrictions on the use of weapons: First, the weapon itself is illegal and the use of the weapon is not permitted.[1] Second, the weapon itself is not illegal, but becomes illegal depending on how it is used. It is prohibited to employ weapons, projectiles, materials, and means of warfare that, by their nature, cause superfluous injury or unnecessary suffering (Article 35. 2 of the First Additional Protocol of Geneva Convention). Although AI-enabled weapons themselves are not illegal weapons under international law, their use should be prohibited if they cannot comply with basic principles of international humanitarian law (IHL) as a 'red line'.[2]

In an armed conflict, the use of force is permitted only when there is a military necessity (principle of military necessity), and armed forces should attack enemy combatants and military objects (principle of distinction). Wounded and incapacitated persons, as well as religious and medical personnel, are prohibited from being attacked even though they are military personnel. Conversely, armed forces may use force against civilians who are directly participating in hostilities, with several exceptions. The use of force is also prohibited if it causes excessive injury or unnecessary suffering to civilians that exceeds the military advantage (principle of proportionality). It is forbidden to kill or harm enemy combatants in a way that causes undue injury or unnecessary suffering (principle of humanity). Finally, the principle of precaution, which refers to all preventive measures to avoid harm to civilians and civilian objects, must be observed.

Some argue that AI-enabled weapons should be prohibited from killing humans because this would violate human dignity.[3] However, if weapons prohibited under international humanitarian law are not used and there is compliance with the principles of IHL, use of

[1] Claude Pilloud et al., 'Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949', Library of Congress, Washington, D.C. 20540 USA, 1987, https://www.loc.gov/item/2011525357/, para 1402.

[2] Yasmin Afina and Giacomo Persi Paoli, 'Governance of Artificial Intelligence in the Military Domain: A Multi-Stakeholder Perspective on Priority Areas', 9 May 2024, https://unidir.org/publication/governance-of-artificial-intelligence-in-the-military-domain-a-multi-stakeholder-perspective-on-priority-areas/, p 14.

[3] Robert Sparrow, 'Robots and Respect: Assessing the Case Against Autonomous Weapon Systems', *Ethics & International Affairs* 30, no. 1 (April 2016): 93–116, https://doi.org/10.1017/S0892679415000647, p. 107.

AI-enabled weapons may be permitted.[4] Whether the use of an AI weapon system is illegal is not definitive, but must be judged comprehensively based on factors such as the degree of human control and intervention, the scope of the attack object, the effectiveness of the attack, and the possibility of stopping the attack by humans.

Therefore, this paper will examine the specific situations when AI-enabled weapons could be used or should be prohibited on a case-by-case basis. Taking human control as a practical criterion for determining compliance with international law, the following discussion will focus on autonomous weapon systems (AWS). AWS are weapon systems that, once activated, can select and engage targets without further human intervention[5] by reviewing information and changes of external circumstances. In this case, humans can interrupt or neutralize the weapon system's attack if problems are foreseen after the attack.[6] Meanwhile, fully autonomous weapon systems not maintaining human control should be prohibited in principle.

[4] Martin Lark, 'The Future of Killing: Ethical and Legal Implications of Fully Autonomous Weapon Systems', *Salus Journal* 5, no. 1 (2017): 62–73; Marco Sassoli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified', *International Law Studies* 90, no. 1 (31 December 2014), https://digital-commons.usnwc.edu/ils/vol90/iss1/1, p. 314-315.

[5] United Nations. *Chair's Summary – Second 2024 Session of the Group of Governmental Experts on Lethal Autonomous Weapons Systems.* CCW/GGE.1/2024/WP.11, 2 October 2024. https://documents.un.org/doc/undoc/gen/g24/181/91/pdf/g2418191.pdf.

[6] US Department of Defense, 'DoD Announces Update to DoD Directive 3000.09, "Autonomy In Weapon Systems"', U.S. Department of Defense, 2023, https://www.defense.gov/News/Releases/Release/article/3278076/dod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems/https%3A%2F%2Fwww.defense.gov%2FNews%2FReleases%2FRelease%2FArticle%2F3278076%2Fdod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems%2F, p. 21-22.

2. Military AI and Human Control

2.1 Using Military AI and Applicable Principles

As technology advances, the scope of military AI application will expand further. The same general principles of AI that are being discussed in the civil sector, such as human dignity, human control, human responsibility, safety, diversity, transparency, non-discrimination, non-bias, privacy, explainability, reliability etc, should be respected. In that case, AI can reduce repetitive and analytical tasks, increase the speed and effectiveness of military operations as well as save defence budget.

These days, the defence sector is broadly using AI in various fields, including information collection and management, human resource management, military medical, logistics, war games, modelling & simulation (M&S), education, training, cyber, transportation, communication, and integrated analysis. The use of AI in data-related fields and non-weapon systems is actively promoted as long as it does not infringe on civilians' privacy or personal information and does not cause damage to civilians.

In addition, there may be some room for using AI in the military for specific purposes. AI could be allowed to verify identity using biometrics or to analyse the structure of genes or biochemicals for therapeutic or protective purposes. However, the use of AI for the sole purpose for surveillance of civilians or the development of weapons of mass destruction (WMD) should be restricted. Provided AI decisions do not generally pose a risk to human life, but are linked to a weapon system or could cause harm to civilians in special circumstances, military AI should be subject to the same review standards as weapon systems. This paper discusses cases when AI is connected to a weapon system or directly used by weapon systems.

Under international law, a weapon that can be used for accurate attacks while also being abused and causing civilian harm is not a prohibited weapon *per se*, but may be restricted by the means or method of warfare.[7] When reviewing a new weapon, it is necessary to determine not only whether the weapon itself is prohibited but also whether its use may cause indiscriminate harm or unnecessary suffering, in order to comply with the principles of distinction, proportionality, and precaution under IHL (Article 36 of the First Additional Protocol). In the case of AI-enabled weapons, they operate on the basis of a combination of hardware and software. Consequently, answering how to evaluate the software or algorithms used in the AI weapon system becomes even more important.

The requirements for the lawful use of weapons, as recognized by the International Court of Justice (ICJ), can be applied to AI weapons systems.[8] The first requirement is to determine whether a system falls under a general or specific prohibition in certain

[7] IRRC, 'A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977', International Review of the Red Cross, 31 December 2006, http://international-review.icrc.org/articles/guide-legal-review-new-weapons-means-and-methods-warfare-measures-implement-article-36.
[8] International Court of Justice, Legality of the Threat or Use of Nuclear Weapons, accessed 16 April 2025, p. 226.

circumstances. The second requirement, in the absence of a specific prohibition; is to establish whether it may ⅰ) cause superfluous injury or unnecessary suffering, ⅱ) be used indiscriminately, ⅲ) not be used proportionately, ⅳ) cause serious damage to the natural environment, and ⅴ) affect current or future trends in international humanitarian law.[9] In addition, the principle of precaution, which seeks to minimize civilian harm, must be observed throughout the entire process of the military operations. This is summarized in the table below, where some elements may overlap across multiple fields.

<table>
<tr><th colspan="2" rowspan="2">Application Field</th><th colspan="4">Discussion Issues</th></tr>
<tr><th>Technology</th><th>Ethics</th><th>Legal Issues</th><th>Policy</th></tr>
<tr><td rowspan="2">Data</td><td rowspan="5">CYBER</td><td>Safety,
Explainability,
Trustworthy,
Transparency,
Human control</td><td>Fairness,
Diversity,
Inclusion,
Non-discrimination,
Non-bias</td><td>Human dignity,
Privacy,
Protection of Information</td><td>Security,
Hacking</td></tr>
<tr><td colspan="4">Collecting, Creating, Managing Data, Generative AI</td></tr>
<tr><td>Non-weapon Systems</td><td colspan="4">Basically, the same discussions are applied as the civil sector,
if non-weapon systems are connected to weapon systems, further discussions are required as follow</td></tr>
<tr><td></td><td colspan="4">Human Resource, Medical care, Logistics, Education, Training, Installation Security, Modeling and Simulations, War games, Transportation, Information, Communication, Integrated Analysis</td></tr>
<tr><td rowspan="2">Weapon Systems</td><td>Technical Standard,
Test and Evaluation,
Verification, Validation, Accreditation</td><td>Killing human by machine (human dignity)
Dictates of public conscience</td><td>Responsibility (civil, criminal, command, state),
Principle of distinction, proportionality, precaution,
Legal review,
Legally Binding</td><td>Risk of Outbreak War,
Terrorism,
Hacking,
Arms control
Non-proliferation</td></tr>
<tr><td colspan="4">Surveillance, Reconnaissance, Command-Control, Automation, Unmanned, Autonomy, Fully autonomy</td></tr>
</table>

[9] United States Department of Defense. *Department of Defense Law of War Manual.* Originally published June 2015; updated July 2023. https://media.defense.gov/2023/Jul/31/2003271432/-1/-1/0/DOD-LAW-OF-WAR-MANUAL-JUNE-2015-UPDATED-JULY%202023.PDF, p. 347.

Common	Definition, Terminology, Taxonomy, Meaningful Human Control, Human Responsibility, Human-Machine Interaction, Explainability, Risk Mitigation, Confidence Building Measure

Table 1: Scope of military AI and applicable principles

2.2 Human Control

For Human control, which is linked to human responsibility, is not a legal concept, but it provides a key resource for examining human intentional or negligent liability. If AI-enabled weapons kill a human without human control, it is difficult to establish human responsibility because the illegal result could not have been foreseen. In such cases, the causal link between the unlawful result and human behavior is likely to be denied.[10] This would violate the dictates of public conscience and human dignity. Even consequentialists may argue for using AI-enabled weapons that can minimize damage compared to the illegal consequences caused by humans, but this also presupposes human control to some degree.

If humans can continue to oversee and control AI weapon systems, ethical issues may be reduced because casualties are caused by humans, not AI-enabled weapons themselves. However, the ethical and legal issues remain continuously, depending on how the scope of meaningful human control is defined. Even if AI-enabled weapons have the capability to make their own decisions, human control should be maintained by identifying the collected data, AI decisions, selection and engagement of the targets in real time and the ability to stop an attack if necessary.

A more problematic situation is when nominal human control gives the appearance of human control (human in the loop), when in reality, AI weapon systems could not be controlled by humans.[11] In other words, this relates to instances where humans cannot understand and review the information or decisions of the weapon systems, and only approve it as a rubber stamp. Even if it was a human who approved the attack, the human actually followed the AI's decision without being able to provide oversight. It would then appear that the human operator has killed the enemy combatants in a formal sense but the actual killer would be the AI weapon systems. When it comes to this case, killing by the AI weapon system, and not a human, is considered a violation of human dignity.

---

[10] Mun-eon Park, 'Autonomous Weapon Systems and Command Responsibility', *Seoul International Law Journal* 31, no. 2 (2024), https://m.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&control_no=9cdc66c46a6f7cf947de9c1710b0298d, p. 24-28.

[11] State of Palestine. *Proposal for the Normative and Operational Framework on Autonomous Weapons Systems.* Working paper CCW/GGE.1/2023/WP.2/Rev.1, submitted to the Group of Governmental Experts on Lethal Autonomous Weapons Systems, Convention on Certain Conventional Weapons, Geneva, 3 March 2023. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.2_Rev.1.pdf, p. 3; Mun-eon Park, *"The Permissibility and Regulation of Autonomous Weapon Systems in International Law"* (PhD diss., Seoul National University, 2019).

The rest of the discussion assumes that AI-enabled weapons are basically autonomous in their selecting and attacking, but human control remains continuously. Considering the situation where human control seems to exist but human supervision is impossible due to nominal human control, this paper examines whether AWS or AI-enabled weapons could be allowed or should be prohibited in specific situations.

3. Principles of Distinction

3.1 Enemy Combatants and Civilians

On the battlefield, AI weapon systems that can utilize video, audio, and biometric information may be superior to humans in their ability to distinguish between regular armed forces, wearing uniform and carrying firearms, and civilians. However, there are many practical difficulties in distinguishing militia, *levée en mass*[12] who are not wearing military uniforms, military religious or medical personnel who are protected under IHL, and those who have lost the will or ability to fight (*hors de combat*) from civilians. In particular, civilians who are directly participating in hostilities are legitimate targets of attacks under international law, and it will be even more difficult for AWS to distinguish them from civilians who are protected under international law. The scope of directly participating in hostilities (DPH) can include the use of weapons as well as logistical support and intelligence gathering activities.[13] Furthermore, it is more difficult for AI to decide whether someone's use of force is in self-defence or whether he is forcibly used as a human shield protected by IHL. Allowing AI to make decisions about what constitutes a DPH is likely to be problematic. Namely, if AI-enabled weapons are used to make decisions on legal and normative matters that have not been resolved among states or in international courts, different results may be produced depending on situations.

Therefore, AWS could be allowed in situations both where only enemy combatants are present on the battlefield but also where combatants and civilians are clearly distinguishable by appearance. Moreover, human authorization for AI-enabled weapons should always be required when the distinction between combatant and civilian may be problematic depending on the combat situation or legal interpretation. In particular, the use of AWS should be limited to situations where there is a mix of protected civilians, civilians directly participating in hostilities, and enemy combatants (regular and irregular), such as in urban warfare. In other words, this is a case where it is difficult to distinguish between legitimate targets and protected targets under IHL, as well as to exercise meaningful human control or supervision.

3.2 Military and Civilian Objects

The principle of distinction requires the distinction between military and civilian facilities, or between military and civilian objects. Military objects, distinguished from civilian objects, are conceptually defined as those that contribute to the enemy's military action

[12] Inhabitants of a non-occupied territory, who on the approach of the enemy spontaneously take up arms to resist the invading forces, without time to form themselves into regular armed units, provided they carry arms openly and respect the laws and customs of war: United Nations, 'Geneva Convention Relative to the Treatment of Prisoners of War', OHCHR, 1949, https://www.ohchr.org/en/instruments-mechanisms/instruments/geneva-convention-relative-treatment-prisoners-war, Article 4. A. 6.
[13] Nils Melzer, 'Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law | International Committee of the Red Cross', 2020, https://www.icrc.org/en/publication/0990-interpretive-guidance-notion-direct-participation-hostilities-under-international, p. 50.

and whose total or partial destruction, capture or neutralization offers a definite military advantage (Article 52. 2 of the First Additional Protocol).

Whether the object is a military target or not is determined by its nature, location, current use, and future purpose. Military targets can be categorized into planned targets and targets of opportunity, according to the target processing method, and sensitive targets and time-limited targets according to their specific purpose.[14] The use of AWS to attack on sensitive targets leading to significant damage, loss of life, or economic impact (Article 56 of the First Additional Protocol), should not be permitted without approval of authorization officer. The use of AWS could be allowed for targets of opportunity and time-sensitive targets as long as they are under continuous human control.

Determining whether a dual-use facility, such as a broadcasting station or power plant that is normally a civilian facility but may be used for military purposes in wartime, is a complicated problem for humans. The purpose of the object or its installation is sometimes not easy to determine externally, so it is difficult to make a judgment without evaluating the specific use of the facility and its impact on the war as a whole. Though the decision can be made through comprehensive data and information by AWS itself, use of AWS against dual-use objects or facilities should be authorized with human approval.

Therefore, AWS can be used against military installations and munitions. Even if attacks against planned targets could be allowed, attacks against targets of opportunity or time-limited targets should be permitted on a restricted basis. The use of AWS should be prohibited against civilian targets and sensitive targets as well as dual-use facilities or objects.

[14] US Joint Chiefs of Staff, 'Joint Chiefs of Staff Instruction: No-Strike and the Collateral Damage Estimation Methodology | Public Intelligence', 15 November 2013, https://publicintelligence.net/cjcs-collateral-damage/.

4. The Principle of Proportionality

4.1 Military Advantage and Collateral Damage

How In AI-enabled weapons, a key question is how to objectively determine the specific and direct military advantage of an attack against the expected civilian casualties or damage to civilian objects. When ratifying the First Additional Protocol, many states made reservations to Article 52. The interpretation of military advantage means benefit from an attack considered as a whole, not an isolated or specific part of an attack.[15] However, in the case of AWS, military advantage must be considered tactically at the time of the specific attack, not strategically. This is because, if strategic interest is included in military advantage, there is a risk of unlimited expansion of collateral damage to civilians. Only military advantage resulting directly from the destruction or neutralization of the enemy target should be considered, rather than the impact on the overall battlefield or war.

In the case of military installations or munitions, an objective economic value can be calculated from the number, area, or type of facilities destroyed. However, the casualties of enemy combatants or civilians cannot be calculated as an economic value, which creates a problem for comparative evaluation. Although the life of the commander-in-chief and a single civilian are equally precious, it may differ in terms of military advantages. However, it is not appropriate for AI weapon systems to make such a decision.

While the type of weapon, its capabilities, the presence and type of nearby civilian facilities, and the presence of civilians are important factors in determining when an attack causes disproportionate, incidental loss of civilian life and destruction of civilian property, it is not easy to predict the indirect damage that may occur to civilians due to destruction of electricity, water, or sanitation facilities. To make objectively predictable comparisons, it is appropriate to exclude damage that can occur after a long period of time or that can cause civilian harm indirectly from destruction of infrastructures. Such indirect or long-term damage, however, must be included in collateral damage as AI technology develops in the future. Of course, if indirect or long-term damage is clear, this should also be included in the calculation of incidental damages.

4.2 Military Advantage vs. Collateral Damage

AWS might be permitted if an attack does not cause collateral damage to civilians and if the military advantage can be mathematically compared to the civilian collateral damage so the former outweighs the latter, like destroying a large number of military facilities or munitions while only a few civilian facilities are destroyed. However, AWS should not be

[15] Julie Gaudreau. "The Reservations to the Protocols Additional to the Geneva Conventions for the Protection of War Victims." *International Review of the Red Cross* 85, no. 849 (March 2003): 143–184. https://www.icrc.org/sites/default/files/external/doc/en/assets/files/other/irrc_849_gaudreau-eng.pdf, p. 16.

permitted even if there is damage to civilian facilities alone, causing significant and prolonged environmental damage.[16]

When it comes to the mathematical impossibility of weighing military advantage against collateral damage, AWS must be limited and subject to human approval. Namely, there are civilian and enemy combatant's casualties at the same time. In cases where the killing of enemy combatants and the destruction of civilian facilities occur at the same time, even though it is not possible to make an arithmetical assessment, AWS could be allowed by principle of military necessity unless the damage to civilians is significant.

Therefore, AWS is permitted if there is only a military advantage and no risk of civilian collateral damage as well as the military advantage clearly and objectively outweighs the potential civilian collateral damage. An exception may be allowed if there are casualties of enemy combatants and the destruction of some civilian facilities, even though the comparison is not mathematically possible. The table below summarizes the discussion.

<table>
<tr><th colspan="2" rowspan="2">Classification</th><th colspan="4">Anticipated Collateral damage</th></tr>
<tr><th>No civil damage</th><th>Destroying civil facility</th><th>Civilian casualty</th><th>Both</th></tr>
<tr><td rowspan="3">Military advantage</td><td>Destroying military facility</td><td rowspan="3">Using AWS possible</td><td>Using AWS possible</td><td rowspan="3">Using AWS impossible</td><td rowspan="3">Using AWS impossible</td></tr>
<tr><td>Killing of enemy combatant</td><td>Using AWS possible depending on situation</td></tr>
<tr><td>Both</td></tr>
</table>

Table 2: Applicable cases of using AWS under proportionality principle[17]

[16] Isabel Robinson and Ellen Nohle, 'Proportionality and Precautions in Attack: The Reverberating Effects of Using Explosive Weapons in Populated Areas', International Review of the Red Cross, 15 April 2016, http://international-review.icrc.org/articles/proportionality-and-precautions-attack-reverberating-effects-using-explosive-weapons, p. 126-127.

[17] Mun-eon Park, 'The Concept of Autonomous Weapon System and Principle of Proportionality', *The Korean Journal of International Law* 64, no. 2 (2019): 85–110, https://doi.org/10.46406/kjil.2019.06.64.2.85, p. 105.

5. Principle of Precaution

5.1 Characteristics of AWS and Prevention

The First Additional Protocol requires that all possible precautions should be taken to avoid harm to civilians and civilian objects in planning or deciding the attack, or in the conduct of military operations (Article 57). The principle of precaution must be implemented throughout the entire process of conducting military operations, not just when an attack is imminent. These principles should also be applied at all life cycle stages of design, production, deployment, use, and post-accident measure as well as training of commanders and operators.

In the case of AWS, the testing & evaluation, verification & validation of software and algorithms are critical because they are the elements that essentially allow the weapon system to operate autonomously. In order for human oversight and human accountability to be maintained for AWS, the decisions or outputs made by AI-enabled weapons must be understandable to humans. Since AI that draws conclusions through deep or reinforcement learning is essentially a black box, humans must be able to examine the output of AI.

The legal review of weapons under Article 36 of the First Additional Protocol is one of ways to ensure that precautionary measures are thoroughly implemented. Practicable precautionary measures rely on the technology and information available at the time of an attack so the performance of the weapon system, testing & evaluation and verification & validation procedures for the software, and the capabilities of the operators are important factors to evaluate. Consequently, AWS should be prohibited in situations where the outputs are unpredictable in a variety of battlefield situations, such as when the software or algorithms have not been evaluated and verified properly.

5.1 Contextual Restriction

In order to maintain human control, the information provided by an AI system must be technically visible, audible, and comprehensible, so that its operators can recognize, understand its content and make a judgment. Moreover, humans must be given the time and opportunity to review the information or decisions provided by the AI and take actions before a violation of international law occurs. The operator must also be knowledgeable about the performance of the AI-enabled weapons and how to correct a problem if it arises.

Above all, to ensure that human control is maintained, AWS should be limited in the situation of communication breakdown by electronic attack (jamming). Even with pre-programmed safeguards in place, it is inappropriate to acknowledge human oversight in the absence of communication.

As an additional safeguard (fail-safe), the use of lethal weapons should be prohibited if communications are lost during operations and only the exercise of the right of self-defence through non-lethal weapons is allowed. The same issue arises underwater, an unmanned submarine that is incommunicado should be prohibited to use torpedoes. However, a mission such as minesweeping could be permitted while incommunicado.

From the *jus ad bellum* perspective, WMD use, including nuclear or biological weapons, should not be permitted to be decided through AI, even in the exercise of the right of self-defence under Article 51 of the UN Charter or customary international law. That is because the risk of damage to mankind caused by AI misjudgement or error is too enormous. This would also violate the principle of proportionality. On the other hand, active use of military AI should be allowed for operational or training-related applications that are difficult to test or evaluate in real-world battlefield situations.

Thus, AWS should be prohibited if humans cannot understand the information or decisions of the AI or cannot review them in time. The use of lethal weapons by AWS should be prohibited in situations where communications are lost and only non-lethal weapons should be used for self-defence. Notwithstanding, AWS could be allowed to perform tasks such as mine clearance even in instances where communication is difficult.

6. Conclusion

The use of AI in the military domain will continue to improve the efficiency of military operations and can minimize civilian casualties in some cases.[18] Even in peacetime, the use of AI should be prohibited to monitor civilians or to develop WMD. In addition, AI that does not normally pose a risk of lethality, but is expected to be used in conjunction with a weapon system to cause harm to civilians, should be subject to the same review as a weapon system.

While AI-enabled weapons are, by themselves, not considered to be illegal weapons *per se* under international law, their use should be prohibited unless they can comply with the fundamental principles of IHL: military necessity, distinction, proportionality, and precaution. Furthermore, human control should be used as a practical criterion for determining whether AI-enabled weapons can comply with international law. The use of fully AWS, where it is impossible to maintain human control, should be prohibited in principle. In addition, the use of AI-enabled weapons should be reviewed in the same way as the use of AWS, when there is nominal human control.

This paper concludes with the following policy recommendations related to the basic principles of distinction, proportionality, and precaution are as follows.

1. AWS may be permitted when only enemy combatants are present or civilians and enemy combatants can be clearly distinguished on the battlefield. Its use should be prohibited in situations where it is difficult to distinguish between civilians and combatants (regular armed forces) and between civilians and irregular forces, or where there is a mix of protected civilians and civilians directly participating in hostilities.
2. AWS could be used for attacks against military installations and munitions. Though attacks against planned targets would be possible, attacks against targets of opportunity or time-limited targets should be limited. AWS should be prohibited against sensitive targets, civilian targets as well as dual-use facilities or objects.
3. AWS could be allowed if there is only a military advantage and no collateral damage to civilians and if the military advantage mathematically outweighs the damage to civilians. As an exception, AWS may be permitted in instances of the use of force against enemy combatants and the destruction of small parts of civilian facilities.
4. AWS should be prohibited in situations where the result of using AI weapons is unpredictable in a fog of wars, such as when software or algorithms have not been evaluated and validated and when human operators cannot objectively verify the operational capabilities of the AI-enabled weapons.
5. AWS should be prohibited if humans cannot understand the information or decisions of the AI or review them in time. The use of lethal weapons by AWS should be prohibited in situations where communications are lost or underwater where communications are not possible.

[18] Ronald Arkin, *Governing Lethal Behavior in Autonomous Robots* (New York: Chapman and Hall/CRC, 2009), https://doi.org/10.1201/9781420085952.

About the Author

Dr. Mun-eon Park

Park, Mun-eon became a Judge Advocate in March 2001. He graduated from Kyungpook University in February 1999 (Bachelor), Indiana School of Law (Bloomington) in June 2011 (LL.M.) and became a member of the New York Bar Association in July 2013. He received his Ph.D. in International Law from Seoul National University in February 2019 (The Permissibility and Regulation of Autonomous Weapon Systems in International Law). He retired in March 2020 as Colonel. He works for the Korean Institution for Defence Analyses now.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies