

Scientific Article

DOI: 10.51940/2024.1.167-188

UDC: 341.3:342.7:004.8

623.09:004.8

*Yuval Shany**

To Use AI or Not to Use AI? Autonomous Weapon Systems and Their Complicated Relationship with the Right to Life

Abstract

The increased prevalence of AI technology developed or adapted for military use raises difficult questions about the compatibility of this new technology with international law in general, and international human rights law (IHRL) in particular. The Human Rights Committee, the expert body entrusted with monitoring the application of the International Covenant on Civil and Political Rights, expressed its view in 2018 on the relationship between the emergence of new military AI and respect for the right to life. The article reviews the terms of the IHRL debate surrounding the introduction of AI technology into military contexts and its relationship to the right to life. Section one briefly reviews some actual and potential applications of AI in military contexts. Section two deals with three principal objections to introducing military AI to battlefield environments: the capacity of autonomous or semi-autonomous AI systems to properly apply international humanitarian law (IHL), concerns about *de facto* lowering of standards of humanitarian protection, and the ethical and legal implications of transferring certain life-and-death decisions from humans to machines. Section three reviews, in light of these three principled objections, specific proposals by the ICRC to limit the use of AI in military contexts (limiting the scope and manner of use of autonomous weapon systems, and excluding unpredictable and lethal systems). Section four reviews the main issues discussed in this article from the vantage point of the right to life under IHRL, as elaborated in General Comment No. 36.

Key words

autonomous weapon systems, right to life, international humanitarian law, human dignity, accountability, transparency, meaningful human control, ICRC, military AI.

* *Hersch Lauterpacht Chair in Public International Law, The Hebrew University of Jerusalem. Prof. Shany served in 2013–2020 as a member of the Human Rights Committee.*

The increased prevalence of AI technology developed or adapted for military use raises difficult questions about the compatibility of this new technology with international law in general, and international human rights law (IHRL) in particular.²² This is because moving away from human discretion and agency towards decision-making by machines in contexts involving the use of lethal force implicates some of the most basic human rights, including the right to life. Indeed, the Human Rights Committee, the expert body entrusted with monitoring the application of the International Covenant on Civil and Political Rights,¹ expressed its view in 2018 on the relationship between the emergence of new military AI and respect for the right to life. General Comment No. 36 on the Right to Life alludes, in the following language, to legal concerns relating to the development and use of autonomous weapon systems:

“65. States parties engaged in the deployment, use, sale or purchase of existing weapons and in the study, development, acquisition or adoption of weapons, and means or methods of warfare, must always consider their impact on the right to life. For example, the development of autonomous weapon systems lacking in human compassion and judgement raises difficult legal and ethical questions concerning the right to life, including questions relating to legal responsibility for their use. The Committee is therefore of the view that such weapon systems should not be developed and put into operation, either in times of war or in times of peace, unless it has been established that their use conforms with article 6 and other relevant norms of international law.”²

The present article will review the terms of the IHRL debate surrounding the introduction of AI technology into military contexts and its relationship with the right to life. Due to time and space limitations, it will not deal with other human rights implicated by the use of AI in military contexts, including equality, privacy, and the emerging right not to be subject to automated decision-making.³ Furthermore, it will address, only to a limited degree, the parallel debate on the normative implications of military AI under international humanitarian law (IHL).⁴

²² *Thanks are due to Dr. Shereshevsky for his comments on an earlier draft of this article. The research for this article was supported by ERC Grant No. 101054745 (DigitalHRGeneration3).*

¹ International Covenant on Civil and Political Rights (ICCPR), 16 December 1966, 999 UNTS 171.

² Human Rights Committee, General Comment No. 36: The Right to Life, UN Doc. CCPR/C/GC/36 (2018), § 65.

³ See, e.g., Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Article 22, OJ L 119, 4 May 2016, p. 1.

⁴ For a comprehensive discussion of the legality of military AI under IHL, see Hua, 2019; Brenneke, 2018; Jensen & Alcalá, 2019.

In section one, I will briefly review some actual and potential applications of AI in military contexts. These applications will serve as a factual background against which normative questions will be discussed in subsequent parts of the article. Section two deals with three principal objections to introducing military AI to battlefield environments: the capacity of autonomous or semi-autonomous AI systems to properly apply IHL, concerns about the *de facto* lowering of standards of humanitarian protection, and other ethical and legal implications of transferring certain life-and-death decisions from humans to machines. I will maintain in that part that, whereas some objections to military AI are compelling, others are contingent on the actual technological state of the art, on a distinction between lethal and non-lethal AI that is hard to maintain, and on an idealised—and, ultimately, unrealistic—portrayal of the qualities of human decision making. Section three reviews, in light of these three principled objections, specific proposals by the ICRC to limit the use of AI in military contexts (namely, limiting the scope and manner of use of autonomous weapon systems, and excluding unpredictable and lethal systems). Finally, section four reviews the main issues discussed in this article from the vantage point of the right to life under IHRL, as elaborated in General Comment No. 36.

1. The Growing Use of AI in Military Contexts

The ‘AI revolution’—involving the transfer of decision-making power from human beings to computerised systems run by AI⁵—has not bypassed military organisations. In fact, these organisations are proving to be a particular hotbed for the development of new AI technologies in light of the complex, multi-factor environments in which they operate; the vital need for speedy, precise and reliable decisions in military contexts; the possibility of increasing troop safety by placing machines rather than humans in the line of fire; and the considerable resources that security bodies can command—especially in an “arms race” context. Indeed, the world’s leading militaries have already introduced a number of sophisticated AI systems into their ranks, and increasingly rely on them in their operations.

Among the AI systems long in use by the US military, for example, one might mention Joint Assistant for Development and Execution (JADE)—a set of software tools, employing AI technology, capable of quickly developing time-sensitive troop deployment plans on the basis of past and existing operational plans adapted to changing mission environments.⁶ In the field of air defence, the US Navy already makes use of the Aegis Ballistic Missiles Defense (BDM) system, which automatically intercepts incoming missiles; its

⁵ See, e.g., Makridaki, 2017.

⁶ Morgan *et al.*, 2020.

capacity is currently being upgraded by the introduction of AI technology to enable better identification of incoming threats and a faster selection of outgoing responses.⁷

In the field of offensive capabilities, a relatively straightforward weapon system used by the US Air Force is the High-Speed Anti-Radiation Missile (HARM) system, which is programmed to identify and target enemy air-defence systems.⁸ The increased reliance of such weapon systems on AI technology significantly enhances their loitering capacity.⁹ Other AI-based systems currently under development by the US military are Collaborative Operations in Denied Environments (CODE)—a weapon system consisting of autonomous aircrafts that can fly in swarms, engage in long-term loitering over targets and carry out a variety of intelligence and targeting missions,¹⁰ and the Combined Joint All-Domain Command and Control (CJADC2)—an integrative system comprising data collection (sense), threat identification and response selection capacity (make sense), and reaction through AI-supported or controlled weapon systems (act).¹¹ A final example is Project Maven—an AI-based imagery analysis software (which also utilises facial recognition technology), developed by the US Department of Defense from 2017 onwards, with the aim of designating targets for military attacks.¹² Of course, while the US is a global leader in developing military AI, it is by no means the only developer and user of such technology. Other countries, such as China,¹³ Russia,¹⁴ France¹⁵ and Israel,¹⁶ also possess significant capacities in this field, and they, like the US, are expected to share these with their allies as well.

This brief survey thus suggests that military AI does not represent a “weapon of the future”, but rather forms part of the current state of the art. Furthermore, the more sophisticated these weapon systems become—due to the evolution of their data collection, data storage, data analysis and overall functional capacities—the greater the tendency of military organisations might be to rely on them and to vest them with autonomous

⁷ Center for Strategic and International Studies, Maritime Security Dialogue: The Aegis Approach with Rear Admiral Tom Druggan, 22 November 2021, <<https://www.csis.org/analysis/maritime-security-dialogue-aegis-approach-rear-admiral-tom-druggan>>.

⁸ Hollings, 2021.

⁹ Ibrahim, 2022.

¹⁰ UAS Vision, DARPA Reveals Details of CODE Program, 2019, <<https://www.uasvision.com/2019/03/25/darpa-reveals-details-of-code-program>>.

¹¹ Department of Defense, Summary of the Joint All-Domain Command and Control (CJADC2) Strategy, March 2022, <<https://media.defense.gov/2022/Mar/17/2002958406/-1/-1/1/SUMMARY-OF-THE-JOINT-ALL-DOMAIN-COMMAND-AND-CONTROL-STRATEGY.PDF>>.

¹² Brewster, 2021.

¹³ Morgan, 2020, pp. 60–82.

¹⁴ *Ibid.*, pp. 83–99.

¹⁵ See, e.g., Manuel, 2022.

¹⁶ See, e.g., Min, 2022; Mimran, Pacholska, Dahan & Trabucco, 2024; Swoskin, 2024.

or semi-autonomous decision-making power. This process of substituting human decision-makers with machines, including in matters of life and death, nonetheless raises difficult ethical and legal concerns.

2. The Case Against LAWS

Most ethical discussions of military AI focus on the development, deployment and use of lethal autonomous weapon systems (LAWS), and most legal discussions concerning LAWS revolve around their compatibility with IHL. Although IHL is the specific branch of international law governing the conduct of hostilities, its norms are highly relevant to IHRL as well, given the considerable substantive overlap between IHL and IHRL, and their concurrent application in situations of armed conflict.¹⁷ The ethical and legal debates around LAWS have accompanied the lengthy—and, so far, inconclusive—process of negotiating an agreement on their development, deployment and use by a Group of Governmental Experts (GGE) convened by the contracting parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons (CCW).¹⁸

Although no uniform definition of LAWS exists, the literature tends to regard them as weapon systems with the autonomous capacity to identify and select targets and to apply lethal force to them.¹⁹ While most military AI systems currently in use do not involve fully autonomous weapon systems—since they still feature a human “in the loop” or “on the loop”—there is little doubt that the combined effect of technologies for target identification (such as those developed by Project Maven) and autonomous targeting capacity (such as that developed in CODE) could be harnessed to develop weapon systems capable of identifying and killing human beings with no human involvement (i.e., with humans “off the loop”). Furthermore, even activating existing AI weapon systems programmed to target military objects—such as radar stations—may lead to loss of human life. Indeed, there is some anecdotal evidence that an attack carried out in 2020 by a Turkish-manufactured AI-powered drone on a militant convoy in Libya resulted in casualties.²⁰ Finally, as explained below, the difference between autonomous AI and semi-autonomous AI (involving humans “in” or “on the loop”) might not be as sharp as it seems, since human control over sophisticated military AI systems is eroding across the

¹⁷ See, e.g., *Legality of the Threat or Use of Nuclear Weapons*, 1996 ICJ 226, 240; Human Rights Committee, General Comment No. 36, § 64. See also Shany, 2023.

¹⁸ Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or Have Indiscriminate Effects, 10 October 1980, 1342 UNTS 137.

¹⁹ See, e.g., Taddeo & Blanchard, 2022. See also ICRC, 2022.

²⁰ See Nasu, 2021.

board, and the ability of human operators to exercise effective oversight is increasingly called into question.²¹

At the heart of the ethical and legal debate on LAWS lie three main issues:

1. Concerns regarding their ability to apply IHL properly;
2. The humanitarian implications of losing the moderating influence of human involvement in battlefield decisions; and
3. Other ethical and legal questions associated with letting machines decide to kill humans.

Taken together, these factors arguably cast doubt on the compatibility of LAWS with international law, generally, and, as section four will show, with the right to life under IHRL in particular.

2.1. Law Application

One common criticism levelled against the development, deployment, and use of LAWS concerns worries about mistakes in target identification, risk assessment and cost-benefit analysis, which could lead to the misapplication of IHL rules—especially the principles of distinction and proportionality. Given the interrelationship between IHL and IHRL, such misapplication of IHL is also likely to also entail a violation of IHRL.²² Arguably, the risk of misapplying IHL and IHRL might justify outlawing LAWS under existing law, regardless of the outcome of the negotiations under the auspices of the CCW GGE.

Human Rights Watch—at the forefront of the “Stop Killer Robots” campaign—published a report in 2021 in which its researchers, together with researchers from a Harvard Law School law clinic, stated:

“It would be difficult for fully autonomous weapons systems, which would select and engage targets without meaningful human control, to distinguish between combatants and noncombatants as required under international humanitarian law [...] [C]omplying with the principle of distinction frequently demands the ability to assess an individual’s conduct and intentions, not just appearance. Such assessments may require interpreting subtle cues in a person’s tone of voice, facial expressions, or body language or being aware of local culture [...] Humans possess the unique capacity to identify with other human beings and are thus equipped to understand the nuances of unforeseen behavior in ways that machines, which must be programmed in advance, simply cannot.”²³

In the same vein, the report contends that the principle of proportionality cannot be properly applied by a machine:

²¹ See, e.g., Renic & Schwartz, 2023.

²² See Human Rights Committee, General Comment No. 36, § 64.

²³ Human Rights Watch and International Human Rights Clinic – Harvard Law School, 2021, p. 7.

“First, because a machine would have trouble distinguishing military from civilian targets, it will face obstacles to assessing the military advantage and civilian harm that would come from a possible attack. Second, the proportionality principle involves a qualitative balancing test that requires the application of human judgment and moral and ethical reasoning [...] human characteristics that machines seem unlikely to possess through their programming [...] Third, proportionality requires contextual decisions at the moment of attack. The lawful response to a situation could change considerably by slightly altering the facts, and it would be impossible to pre-program a robot to be prepared for the infinite number of scenarios that it could face.”²⁴

Comparable objections, focusing on the capacity of LAWS to apply IHL properly, have also been raised or discussed by other NGOs,²⁵ UN officials²⁶ and academic researchers in this field.²⁷

It is hard to disagree that relying on LAWS to apply IHL in complex battlefield conditions may yield false negatives and false positives, leading to legal mistakes—if not outright violations—given the problems AI systems face when attempting to develop situational awareness and respond to unforeseen circumstances.²⁸ Serious doubts also remain as to whether the difficult, multi-factored and value-laden act of balancing between military necessity against humanitarian considerations that underlies IHL proportionality can be properly undertaken by an algorithm. Still, it is also difficult to deny that human beings applying IHL are prone to error, especially in the ‘fog of war’ and when responding to surprising developments on the ground; and some also commit intentional violations. Moreover, doubts regarding the feasibility of implementing the principle of proportionality in a fixed or predictable manner have been raised with regard to human decision makers as well.²⁹

At least from a rule-consequentialist point of view, a key question may be: who is the less accident-prone decision-maker—the human soldier or the AI-based weapon system? The answer to this question appears largely contingent on developments in the rele-

²⁴ *Ibid.*, p. 8.

²⁵ See, e.g., Article 36, 2019.

²⁶ See, e.g., Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, UN Doc. A/HRC/23/47 (2013), §§ 63–74 (presenting the main features of the debate around the ability of LAWS to properly apply IHL).

²⁷ See, e.g., Sharkey, 2012, pp. 787 and 788–790 (arguing that LAWS lack situational awareness and common-sense reasoning needed to apply the principles of distinction and proportionality). See also McFarland, 2015, pp. 1313 and 1335 (claiming that battlefield decisions by LAWS will be different than those reached by humans because they will be based on more general and pre-determined rules and on anticipated circumstances).

²⁸ For a discussion of the difference between legal mistakes and legal violations, see Pacholska, 2023.

²⁹ See, e.g., Statman *et al.*, 2020.

vant AI technology, including its ability to predict and emulate human decision-making. Importantly, any such comparison ought to be made not between a machine and an idealised version of a perfect human being, but rather between a machine and a realistic version of a human being, whose decisions are likely to suffer from human imperfections, biases, and frailties.³⁰

In the long run, as in other highly complex fields of decision-making requiring speedy reactions and multifaceted analysis (such as driving), it is unlikely that human decision-makers could keep pace with developments in machine sensory and analytical capacity, given the constant improvements in computer-based data collection and storage capacity, processing speed and power, and system resilience.³¹ As a result, machines are expected, sooner or later, to make better-informed, more accurate, and faster decisions than human soldiers regarding the choice of means and methods of warfare necessary to attain military objectives with greater efficiency, while inflicting the least possible extent of collateral damage.³² Furthermore, while there remain serious methodological difficulties in quantifying the many variables comprising IHL proportionality analysis, such difficulties are not likely to be insurmountable (and, as noted above, they also pose a serious challenge for human decision-making).³³

It is also noteworthy that—unlike human soldiers—machines do not grow tired, frustrated, or confused; nor do they rely on inaccurate heuristics (or hunches) as decision-making short-cuts, as humans do.³⁴ Rather, they are expected to strictly follow pre-determined rules of conduct—including IHL rules—even in the most stressful of circumstances (including when their own continued existence is on the line), and to apply them in the exact manner in which they were trained (for example by studying past patterns of human conduct or drawing statistical predictions about future human decisions). And the more sophisticated the algorithms, machine learning capabilities, and training data available to military AI become, the smaller the likelihood of their involvement in deadly mistakes or legal violations (still, as explained below, the more sophisticated military AI becomes, the harder it is for humans to exercise over them effective control and monitoring).

In fact, replacing humans with machines in the line of fire may enable decision-makers to adopt higher standards of IHL protection than would otherwise have been

³⁰ Cf. Zerilli *et al.*, 2019; Heller, 2023.

³¹ See, e.g., Korteling *et al.*, 2021; Schmitt, 2013.

³² See, e.g., Winter, 2022, pp. 18–19.

³³ See, e.g., Schuller, 2019; Winter, 2022, pp. 16–17.

³⁴ See, e.g., Walker, 2021, pp. 10 and 16. For more information on reliance on heuristics, see Tversky & Kahneman, 1974.

possible. Such standards may include a “shoot second”³⁵ or a “double check”³⁶ rule of engagement, a “no civilian casualties” proportionality formulation,³⁷ and adopting “limited tolerance for error” settings in battlefield operations.³⁸ Indeed, it has already been alleged that, once established that LAWS can offer higher levels of IHL protection than human soldiers, there may be a legal obligation to opt for the LAWS-based approach.³⁹

2.2. *The Moderating Impact of Human Involvement*

Another cluster of objections to the application of LAWS to battlefield situations revolves around the inability of machines to exercise human compassion and discretion, and to moderate the application of IHL in circumstances where following the letter of the law would result in harsh consequences from a moral standpoint. Examples of situations, in which human compassion and discretion might provide a higher level of humanitarian protection than strict application of IHL include: using non-lethal weapons against child soldiers;⁴⁰ choosing to capture rather than kill enemy combatants even in the absence of a legal obligation to do so;⁴¹ and refraining from targeting soldiers withdrawing from the battlefield under conditions in which they are unlikely to rejoin the armed conflict.⁴² Arguably, delegating decision-making in such cases from humans to machines that operate on the basis of “black letter” rules might result in the loss of the additional safeguards human soldiers sometimes afford as a matter of discretion, leading to an overall reduction in the level of humanitarian treatment in and around the battlefield.⁴³ This sentiment, regarding a potential increase in the lethality of battlefield conditions due to the introduction of LAWS, appears to underlie some of the concerns voiced in paragraph 65 of General Comment No. 36, which reads into the right to life under IHRL certain humanitarian considerations that go beyond those found in the language of IHL rules.⁴⁴

Here too, doubts have been expressed in the literature concerning the comparative advantage of human beings over machines in affording enemy soldiers, civilians, and *hors de combat* humane treatment—over and above applicable legal obligations. Some

³⁵ Geiss, 2016. But see Sassóli, 2014, pp. 308 and 336 (alleging that “conservative programming” is not likely to be sustainable, given the loss of military advantage).

³⁶ See, e.g., Geiss, 2016.

³⁷ Cf. Runkle, 2015.

³⁸ Cf. Bellotti, 2021.

³⁹ Cf. Jensen, 2020, pp. 26 and 55.

⁴⁰ See, e.g., Barrett, 2019.

⁴¹ See, e.g., Schmitt, 2013.

⁴² See, e.g., Cook & Hamann, 1994.

⁴³ See, e.g., Human Rights Watch, p. 9.

⁴⁴ Human Rights Committee, General Comment No. 36, § 65.

commentators point out that while certain human emotions, such as compassion and empathy, may lead to higher standards of humanitarian treatment, other human emotions, such as fear, anger, or revenge, can generate the opposite result.⁴⁵ Furthermore, although algorithms cannot experience emotions, they can be programmed to emulate emotion-driven human conduct or to follow a course of action deemed consistent with positive human emotions, such as compassion or empathy (e.g., they can be programmed to avoid targeting with lethal weapons children under any circumstances).⁴⁶

2.3. *Other Ethical and Legal Concerns*

Even if LAWS are capable of affording an equivalent level of humanitarian protection to that afforded by human soldiers, the very delegation of decision-making from humans to machines raises difficult ethical and legal concerns for which no satisfactory technological solution appears available. First, referring decisions over life and death to a computer algorithm engaged in risk assessment and cost-benefit analysis, without effective human supervision and control, is difficult to reconcile with moral norms requiring respect for human dignity and life.⁴⁷ Arguably, treating a human being as nothing more than a node generating data about risks arising from his/her predicted conduct, or about protections due by virtue of the low-risk category to which he/she belongs, is dehumanising in a profound sense.⁴⁸

Second, delegating decisions to machines creates an agency problem, potentially resulting in a lack of moral responsibility and legal accountability.⁴⁹ AI weapon systems do not have moral agency, and it is possible that none of the human actors involved in developing, introducing, and deploying them in specific theatres of hostilities will have a full grasp of the system's shortcomings and the precise battlefield conditions in which it is deployed. This hampers any attempt to assign ethical or legal responsibility for breaches of IHL or IHRL.⁵⁰

Third, as with other applications of AI, military AI raises difficult questions of transparency—in particular, explainability and traceability.⁵¹ Difficulties in understanding the reasons underlying machine decisions, exercising control over them, and monitoring

⁴⁵ See, e.g., Sassóli, 2014, p. 318; Price, 2016.

⁴⁶ Cf. Xiao *et al.*, 2016.

⁴⁷ See, e.g., Asaro, 2012; Wagner, 2014.

⁴⁸ See, e.g., Laitinen & Otto Sahlgren, 2021, pp. 10–11.

⁴⁹ See, e.g., Taddeo and Blanchard, 2022, p. 37; Human Rights Watch, *Mind the Gap: The Lack of Accountability for Killer Robots*, 2015.

⁵⁰ See, e.g., Amoroso & Giordano, 2019.

⁵¹ See, e.g., Atherton, 2022.

their operations further undermine the conditions for ethical and legal accountability, and ultimately weaken the rule of law.⁵²

Simultaneously, one might acknowledge that even before the introduction of LAWS, military organisations had already come to rely on weapon systems subject to limited human control (e.g., torpedoes),⁵³ on long-distance control (e.g., drones),⁵⁴ and on big data for targeting decisions (e.g., “signature strikes” following long-term data collection to establish “patterns of life” for suspected militants).⁵⁵ In other words, they have long employed weapons, means and methods of warfare featuring some of the same ethical and legal issues afflicting LAWS: delegating significant decision-making capacity to machines, and operating with a reduced sense of accountability and limited transparency.

Furthermore, as noted above, even military AI systems that leave humans the ultimate decision whether or not to use lethal force on a specific target (“humans in the loop” or “humans on the loop”), often constrain or shape human decision-making—through “black box”⁵⁶ and “automation bias”⁵⁷ features—in ways that render such human supervision and control merely nominal, from a practical viewpoint. In other words, the increased reliance on military AI is leading to an erosion of human decision-making capacity across the board, and excessive reliance on distinctions between humans “in the loop”, “on the loop”, and “off the loop” might perpetuate an illusion of effective human supervision and control, which little resembles reality. Put differently, it is questionable whether opposition to LAWS can be meaningfully distinguished, over time, from broader opposition to military AI—with all the operational implications such opposition might entail.

3. The Position of the ICRC

It is against the background of the extensive discussion about the conformity of LAWS with IHL and IHRL that the position of the ICRC on the legality of LAWS is particularly interesting. This is both because of the pride of place the ICRC occupies in the field as guardian and promoter of IHL,⁵⁸ and because its position directly engages

⁵² See, e.g., Rosengrün, 2022.

⁵³ See, e.g., Work, 2021.

⁵⁴ See, e.g., Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Philip Alston, Study on Targeted Killings, UN Doc. A/HRC/14/24/Add.6 (2010), § 84.

⁵⁵ See, e.g., Gibson, 2021. The legality of such practices under IHL has been, however, challenged. See, Heller, 2013.

⁵⁶ See, e.g., Schwartz, 2018.

⁵⁷ See, e.g., Cabitza, 2019, pp. 283 and 293.

⁵⁸ For a discussion, see, e.g., Geiss & Zimmermann, 2017, p. 215.

with the relationship between LAWS, IHL, and broader humanitarian considerations, including other ethical and legal concerns.

In 2021, an ICRC policy paper proposed the following recommendations relating to the use of autonomous weapon systems (AWS):

1. Unpredictable AWS should be expressly ruled out, notably because of their indiscriminate effects. This would best be achieved through a prohibition on AWS that are designed or used in such a way that their effects cannot be sufficiently understood, predicted and explained.
2. In light of ethical considerations to safeguard humanity and to uphold IHL rules for the protection of civilians and combatants hors de combat, the use of AWS to target human beings should be ruled out. This would best be achieved through a prohibition on AWS that are designed or used to apply force against persons.
3. In order to protect civilians and civilian objects, uphold the rules of IHL, and safeguard humanity, the design and use of AWS that would not be prohibited should be regulated, including through a combination of:
 - limits on the types of target, such as constraining them to objects that are military objectives by nature
 - limits on the duration, geographical scope, and scale of use, including to enable human judgement and control in relation to a specific attack
 - limits on situations of use, such as constraining them to contexts where civilians or civilian objects are not present
 - requirements for human–machine interaction, notably to ensure effective human supervision, timely intervention, and deactivation.⁵⁹

It is noteworthy that the overarching framework for the ICRC recommendations is a call on states to “adopt new binding rules” to give effect to the recommendations.⁶⁰ In other words, the policy paper presents itself as a proposal for new *lex ferenda*. It does not claim that LAWS are strictly banned by existing IHL (the ICRC does maintain, however, that “it is difficult to envisage realistic combat situations where LAWS use against persons would not pose a significant risk of IHL violations”).⁶¹ Furthermore, a central recommendation found in the policy paper—to ban the use of autonomous weapon systems to target human beings—is based first and foremost on ethical, and not strictly legal, considerations relating to human dignity.⁶²

Other aspects of the ICRC position are grounded, however, in traditional legal considerations: concerns relating to unpredictability, accurate target selection and collateral harm mirror the concerns about the capacity of LAWS to properly apply IHL discussed

⁵⁹ ICRC Position On Autonomous Weapon Systems, 2021, p. 11.

⁶⁰ *Ibid.*

⁶¹ *Ibid.*, p. 9.

⁶² *Ibid.*, p. 8.

in Part Two. The policy paper notes in this regard certain specific concerns relating to military operations in urban environments and the impact of unforeseen circumstances in military operations involving AWS engaged in long-term loitering. In addition, the policy paper raises concerns about legal accountability due to the limited capacity for understanding, predicting, and explaining the effects of autonomous weapon systems, the inadequate level of human control over them, and the broad scope of discretion afforded by such systems to algorithms.

It may be noted in this regard that the recommendation for a “human on the loop” in the ICRC position paper, including retaining the power to deactivate autonomous weapon systems, appears to go beyond the guiding principles adopted by the 2019 GGE on LAWS, which only alluded to placing human–machine interaction within an accountability framework, including a responsible chain of command and control.⁶³ Whereas the GGE was unable to reach consensus on a definition of the “meaningful human control” standard,⁶⁴ the ICRC proposed specific criteria for the exercise of such power of supervision and control.

The upshot of the position espoused in the policy paper is that the use of LAWS against human beings is considered by the ICRC to be unethical and legally problematic—though not clearly legally impermissible. In practical terms, barring a specific agreement relating to the development, deployment, and use of LAWS, the legal problems identified in the policy paper would need to be reviewed in the course of new weapon legality assessments, pursuant to Article 36 of the First Additional Protocol to the Geneva Conventions.⁶⁵

⁶³ Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, UN Doc. CCW/GGE.1/2019/3 (2019), Annex IV (“(c) Human-machine interaction, which may take various forms and be implemented at various stages of the life cycle of a weapon, should ensure that the potential use of weapons systems based on emerging technologies in the area of lethal autonomous weapons systems is in compliance with applicable international law, in particular IHL. In determining the quality and extent of human-machine interaction, a range of factors should be considered including the operational context, and the characteristics and capabilities of the weapons system as a whole; (d) Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control”).

⁶⁴ Kwik, 2022.

⁶⁵ Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 8 June 1977, Article 36, 125 UNTS 3 (“In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party”).

As for the ethical issues raised, a fundamental dilemma that the ICRC policy paper avoids discussing is whether technological developments that ultimately result in better application of IHL by machines than by humans could justify resorting to them, despite the troubling implications of authorising machines to kill humans. Furthermore, one might ask whether placing “human on the loop” limits on the use of algorithms to ensure legal compliance and accountability—would be tenable in the long run, given the growing gap between algorithmic and human capacity. The better machines become in fast and complex decision-making, the less accessible and understandable their decisions will inevitably be to humans, and the less effective the supervision and control that humans can exercise over their operations.⁶⁶ In the long run, difficult ethical and legal trade-offs between performance quality and the quality of supervision and control over performance may present themselves to policy-makers and their legal advisers.

4. General Comment 36

The doubts as to whether IHL clearly prohibits the use of LAWS, as discussed in previous sections, underscore the significance of broadening the scope so as to include IHRL norms as well. The advantage of IHRL over IHL in this regard is that it explicitly and implicitly recognises many of the normative notions underlying concerns about the development, deployment, and use of LAWS—concerns for which IHL does not offer a dedicated vocabulary—such as humanitarian protection that goes beyond the strict requirements of IHL and the ethical and legal implications of authorising machines to kill humans without effective supervision and control by human beings. In other words, IHRL offers protection both in situations governed by IHL (where IHRL provides overlapping protection) and in cases where IHL does not appear to constrain decision-making, thus inviting the application of humanitarian and other ethical and legal considerations.

An example of a broad ethical and legal consideration influencing the scope of protections afforded in and around the battlefield is the objection to permitting machines to kill humans, which is based on the notion of human dignity. This notion is found in Article 1 of the Universal Declaration of Human Rights⁶⁷ and in the Preambles to both Covenants from 1966.⁶⁸ On that basis, the Human Rights Committee explained

⁶⁶ Cf. Milmo, 2024 (citing Geoffrey Hinton: “how many examples do you know of a more intelligent thing being controlled by a less intelligent thing”).

⁶⁷ Universal Declaration of Human Rights, 10 December 1948, Article 1, GA Res. 217A III (1948) (“All human beings are born free and equal in dignity and rights”).

⁶⁸ ICCPR, preamble (“Considering that, in accordance with the principles proclaimed in the Charter of the United Nations, recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world;

in General Comment No. 36 that the right to life “concerns the entitlement of individuals to be free from acts and omissions that are intended or may be expected to cause their unnatural or premature death, as well as to enjoy a life with dignity”.⁶⁹ It could be claimed that the development, deployment and use of LAWS, involving the delegation of life-and-death decisions to machines lacking human agency, are *prima facie* incompatible with the right to life with dignity. Such an approach seems consistent with the development, outside the context of military AI, of a right not to be subject to automated decisions over significant matters.⁷⁰

In the same vein, notions of transparency and accountability mentioned above are strongly related to procedural dimensions of IHRL protection. Here too, reviewing General Comment No. 36 could be instructive. The Comment reads into Article 6 a normative expectation to report, review and investigate certain lethal incidents;⁷¹ a recommendation for the evaluation and monitoring of the impact of certain weapons on the right to life;⁷² an obligation to effectively monitor and control the involvement of private actors in the application of lethal force;⁷³ a duty to “take adequate measures of protection, including continuous supervision, in order to prevent, investigate, punish and remedy arbitrary deprivation of life by private entities”;⁷⁴ and a requirement “to investigate and, where appropriate, prosecute the perpetrators of such incidents, including incidents involving allegations of excessive use of force with lethal consequences”.⁷⁵ Although these specific obligations—some of which represent soft law and some hard law—were not formulated with a view to addressing the risks to the right to life posed by LAWS, they could apply thereto *mutatis mutandis*, and entail requirements of transparency and accountability for all cases involving the use of military AI.

Indeed, the specific paragraph that addresses the challenge of autonomous weapon systems—paragraph 65 (whose text is provided in the Introduction to this article)—explicitly sets out an obligation to consider the impact on the right to life of all new weapons, and calls for a moratorium on the development, deployment and use of autonomous weapon systems until their compatibility with Article 6 of the ICCPR and

Recognizing that these rights derive from the inherent dignity of the human person”). See also International Covenant on Economic, Social and Cultural Rights, 16 December 1966, preamble, 999 UNTS 3.

⁶⁹ Human Rights Committee, General Comment No. 36, § 3.

⁷⁰ General Data Protection Regulation, Article 22.

⁷¹ Human Rights Committee, General Comment No. 36, § 13.

⁷² *Ibid.*, § 14.

⁷³ *Ibid.*, § 15.

⁷⁴ *Ibid.*, § 21.

⁷⁵ *Ibid.*, § 27.

other relevant norms of international law has been established.⁷⁶ This formulation appears to be influenced by the legality assessment process found in Article 36 of the First Additional Protocol to the Geneva Convention.

It is interesting to note that the Human Rights Committee singled out, in paragraph 65, several problematic features in the operation of autonomous weapon systems: lack of human compassion and judgment, and questions of legal responsibility. These issues relate to all three levels of criticism discussed in section two: proper law application (lack of judgment), additional humanitarian considerations (compassion) and other legal and ethical concerns (legal responsibility). Whereas under IHL, *lex lata* focuses only on the capacity to properly apply the law, the IHRL framework is, as explained above, broad enough to capture more abstract notions of human dignity, humanitarian protection, accountability and transparency. Still, even under IHRL, the Committee did not call for an outright ban on LAWS, but rather for extra-caution in their development, deployment and use. It cannot be excluded that, once sufficient empirical data has been gathered concerning the ability of future versions of LAWS to comply with IHRL (and IHL), and especially after adequate safeguards concerning transparency, *ex ante* supervision, real-time control and *ex-post* accountability have been put in place, they could be regarded as IHRL-compatible.

The ability of bodies such as the Human Rights Committee to continuously monitor states' record in developing, deploying and using military AI during periodic reviews of state reports under relevant human rights instruments⁷⁷ provides these bodies with a unique opportunity to fine-tune the interpretation and application of specific IHRL norms governing military AI. A similar contribution can be made by the work of UN special procedures operating under the auspices of the Human Rights Council. One question that lies beyond the scope of the present discussion—but which might nonetheless be considered in the future by IHRL-applying bodies—is whether the growing reliance on military AI increases the propensity to resort to military force in ways that violate the prohibition against the use of force in international law, and, by implication, Article 6 of the ICCPR.⁷⁸

5. Conclusion

Military AI is already changing how armed forces operate, prompting a growing reliance on machines to replace humans in decision-making. While this development raises difficult ethical and legal issues—especially given doubts about the quality of machine performance, aversion to machines making fateful decisions for human beings, and the

⁷⁶ *Ibid.*, § 65.

⁷⁷ See e.g., ICCPR, Article 40.

⁷⁸ See Human Rights Committee, General Comment No. 36, § 75.

chronic problems of transparency and accountability afflicting the use of AI—military AI might over time also improve the quality of decisions in and around the battlefield, potentially resulting in better compliance with IHL and IHRL.

As a result, decision-makers might sooner or later face the dilemma of whether—after appropriate impact and risk assessments have been conducted—to develop, deploy and use LAWS as a cost-effective method to improve compliance with IHL and enhance humanitarian protections. Even then, the IHRL framework appears more conducive than the existing IHL framework to consolidating specific normative expectations relating to human dignity, transparency and accountability, possibly directing the field's development towards patterns of machine–human interaction that provide safeguards against violations of applicable IHRL and IHL standards.

References

- Amoroso, D. & Giordano, B. (2019) 'Who Is to Blame for Autonomous Weapons Systems' Misdoings?' in: Carpanelli, E. & Lazerini, N. (eds.) *Use and Misuse of New Technologies*. Springer.
- Article 36 (2019) Policy Note: Targeting People – Key issues in the regulation of autonomous weapons systems, <<https://article36.org/wp-content/uploads/2019/11/targeting-people.pdf>>.
- Asaro, P. (2012) 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making', 94 *International Review of the Red Cross* 687.
- Atherton, K. (2022) 'Understanding the Errors Introduced by Military AI Applications', *Brookings Tech Stream*, 6 May <<https://www.brookings.edu/techstream/understanding-the-errors-introduced-by-military-ai-applications/>>.
- Barrett, R.C. (2019) 'Humanising the Law of Targeting in Light of a Child Soldier's Right to Life', 27 *The International Journal of Children's Rights* 3.
- Bellotti, M. (2021) 'Helping Humans and Computers Fight Together: Military Lessons from Civilian AI', *War On The Rocks*, 15 March, <<https://warontherocks.com/2021/03/Helping-Humans-And-Computers-Fight-Together-Military-Lessons-From-Civilian-Ai/>>.
- Brenneke, M. (2018) 'Lethal Autonomous Weapon Systems and their Compatibility with International Humanitarian Law: A Primer', *Yearbook of International Humanitarian Law* 59.
- Brewster, T. (2021) 'Project Maven: Startups Backed By Google, Peter Thiel, Eric Schmidt And James Murdoch Are Building AI And Facial Recognition Surveillance Tools For The Pentagon', *Forbes*, 8 September.
- Cabitza, F. (2019) 'Biases Affecting Human Decision Making in AI-Supported Second Opinion Settings' in: Torra, V., et al. (eds.) *Modelling Decisions for Artificial Intelligence*. Springer.
- Center for Strategic and International Studies (2021) Maritime Security Dialogue: The Aegis Approach with Rear Admiral Tom Druggan, 22 November.
- Cook, M.L. & Hamann, P.A. (1994) 'The Road to Basra: A Case Study in Military Ethics', 14 *The Annual of the Society of Christian Ethics* 207.
- Department of Defense (March 2022) Summary of the Joint All-Domain Command and Control (CJADC2) Strategy.
- Geiss, R. (2016) Autonomous Weapons Systems: Risk Management and State Responsibility, submission to Third CCW meeting of experts on lethal autonomy.

- us weapons systems (LAWS) Geneva, 11–15 April, <[https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Informal_Meeting_of_Experts_\(2016\)/Geiss-CCW-Website.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Informal_Meeting_of_Experts_(2016)/Geiss-CCW-Website.pdf)>.
- Geiss, R. & Zimmermann, A. (2017) ‘The International Committee of the Red Cross: A Unique Actor in the Field of International Humanitarian Law Creation and Progressive Development’ in: Geiss, R., Zimmermann, A. & Haumer, S. (eds.) *Humanizing the Laws of War: The Red Cross and the Development of International Humanitarian Law*. Geneva: International Committee of the Red Cross.
- Gibson, J. (2021) ‘Death by Data: Drones, Kill Lists and Algorithms’, *E-International Relations*, 18 February, <<https://www.e-ir.info/2021/02/18/death-by-data-drones-kill-lists-and-algorithms/>>.
- Heller, K.J. (2013) “One Hell of a Killing Machine” Signature Strikes and International Law’, 11 *Journal of International Criminal Justice* 89.
- Heller, K.J. (2023) ‘The Concept of “The Human” in the Critique of Autonomous Weapons’, 15 *Harvard National Security Journal* 1.
- Hollings, A. (2021) ‘America’s Loitering Radar-Hunting Missile Is Due For A Comeback’, *Sandboxx*, 14 December, <<https://www.sandboxx.us/Blog/Americas-Loitering-Radar-Hunting-Missile-Is-Due-For-A-Comeback/>>.
- Hua, S.-S. (2019) ‘Machine Learning Weapons and International Humanitarian Law: Rethinking Meaningful Human Control’, 51 *Georgetown Journal of International Law* 117.
- Human Rights Watch (2015) *Mind the Gap: The Lack of Accountability for Killer Robots*.
- Human Rights Watch and International Human Rights Clinic – Harvard Law School (December 2021), *Crunch Time on Killer Robots Why New Law Is Needed and How It Can Be Achieved*, <https://www.hrw.org/sites/default/files/media_2021/11/Crunch%20Time%20on%20Killer%20Robots_final.pdf>.
- Ibrahim, A. (2022) *Loitering Munitions as a New-Age Weapon System*, Centre for Strategic and Contemporary Research, 5 December, <<https://cscr.pk/explore/themes/defense-security/loitering-munitions-as-a-new-age-weapon-system/>>.
- ICRC (2021) *Position On Autonomous Weapon Systems*.
- ICRC (2022) *What you need to know about autonomous weapons*, <<https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons>>.
- Jensen, E.T. (2020) ‘The (Erroneous) Requirement for Human Judgment (and Error) in the Law of Armed Conflict’, 96 *International Law Studies* 26.

- Korteling, J.E. (Hans), *et al.* (2021) 'Human versus Artificial Intelligence', *Frontiers Artificial Intelligence* (online edition), <<https://www.frontiersin.org/articles/10.3389/frai.2021.622364/full>>.
- Kwik, J. (2022) 'A Practicable Operationalisation of Meaningful Human Control', 11 *Laws* 43.
- Laitinen, A. & Sahlgren, O. (2021) 'AI Systems and Respect for Human Autonomy', *Frontiers in Artificial Intelligence* (online edition, 26 October), <<https://www.frontiersin.org/articles/10.3389/frai.2021.705164/full>>.
- Makridakis, S. (2017) 'The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms', 90 *Futures* 46.
- McFarland, T. (2015) 'Factors Shaping the Legal Implications of Increasingly Autonomous Military Systems', 97 *International Review of the Red Cross* 1313.
- Manuel, R. (2022) 'French Military Approves Final Phase of Big Data and AI Platform Artemis', *The Defence Post*, 15 July, <<https://www.thedefensepost.com/2022/07/15/french-military-big-data-ai/>>.
- Milmo, D. (2024) "'Godfather of AI' Shortens Odds of the Technology Wiping Out Humanity Over Next 30 Years', *The Guardian*, 27 December, <<https://www.theguardian.com/technology/2024/dec/27/godfather-of-ai-raises-odds-of-the-technology-wiping-out-humanity-over-next-30-years>>.
- Mimran, T., Pacholska, M., Dahan, G., & Trabucco, L. (2024) 'Beyond the Headlines: Combat Deployment of Military AI-Based Systems by the IDF', *Articles of War*, 2 February, <<https://lieber.westpoint.edu/beyond-headlines-combat-deployment-military-ai-based-systems-idf/>>.
- Min, R. (2022) 'Israel deploys AI-powered robot guns that can track targets in the West Bank', *Euronews*, 17. October <<https://www.euronews.com/next/2022/10/17/israel-deploys-ai-powered-robot-guns-that-can-track-targets-in-the-west-bank>>.
- Morgan, F.E., *et al.*, (2020) *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World*. RAND.
- Nasu, H. (2021) 'The Kargu-2 Autonomous Attack Drone: Legal & Ethical Dimensions', *Articles of War*, 10 June, <<https://lieber.westpoint.edu/kargu-2-autonomous-attack-drone-legal-ethical/>>.
- Pacholska, M. (2023) 'Military Artificial Intelligence and the Principle of Distinction: A State Responsibility Perspective', 56 *Israel Law Review* 3.
- Price, R. (2016) 'In Defence of Killer Robots', *Insider*, 24 June, <<https://www.businessinsider.com/in-defence-of-killer-robots-interview-with-expert-dr-william-boothby-about-laws-2016-6?r=US&IR=T>>.

- Renic, N.C. & Schwartz, E. (2023) 'Inhuman-in-the-loop: AI-targeting and the Erosion of Moral Restraint', *Articles of War*, 19 December, <<https://opiniojuris.org/2023/12/19/inhuman-in-the-loop-ai-targeting-and-the-erosion-of-moral-restraint/#:~:text=The%20IDF%20has%20at%20least,data%20from%20other%20surveillance%20sources>>.
- Rosengrün, S. (2022) 'Why AI is a Threat to the Rule of Law', 1 *Digital Society* (online version) Article 10.
- Runkle, B. (2015) 'The Obama Administration's Human Shields: How the Obama administration is using the threat of civilian casualties to hold its fire against the Islamic State', *Foreign Policy*, 30 November, <<https://foreignpolicy.com/2015/11/30/the-obama-administrations-human-shields/>>.
- Sassóli, M. (2014) 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified', 90 *International Law Studies* 308.
- Schmitt, M.N. (2013) 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics', *Harvard National Security Journal* (online edition), <<https://harvardnsj.org/2013/02/05/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/>>.
- Schmitt, M.N. (2013) 'Wound, Capture, or Kill: A Reply to Ryan Goodman's "The Power to Kill or Capture Enemy Combatants"', 24 *European Journal of International Law* 855.
- Schuller, A.L. (2019) 'Artificial Intelligence Effecting Human Decisions to Kill: The Challenge of Linking Numerically Quantifiable Goals to IHL Compliance', 15 *I/S: A Journal Of Law And Policy* 105.
- Schwartz, E. (2018) 'The (Im)possibility of Meaningful Human Control for Lethal Autonomous Weapon Systems', *Humanitarian Law & Policy*, 29 August, <<https://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/>>.
- Shany, Y. (2023) 'Human Rights Norms Applicable in the Situation of Armed Conflict: Beyond the Lex Generalis/Lex Specialis Framework', 66 *Japanese Yearbook of International Law* 3.
- Sharkey, N.E. (2012) 'The Evitability of Autonomous Robot Warfare', 94 *International Review of the Red Cross* 787.
- Statman, D., et al. (2020) 'Unreliable Protection: An Experimental Study of Experts' *In Bello* Proportionality Decisions', 31 *European Journal of International Law* 429.

- Swoskin, E. (2024) 'Israel has built an 'AI Factory' for War. It has unleashed it in Gaza', *Washington Post*, 29 December, <<https://www.washingtonpost.com/technology/2024/12/29/ai-israel-war-gaza-idf/>>.
- Taddeo, M. & Blanchard, A. (2022) 'A Comparative Analysis of the Definitions of Autonomous Weapons Systems', 28(5) *Science and Engineering Ethics* 37.
- Talbot Jensen, E. & Alcalá, R.T.P. (2019) *The Impact of Emerging Technologies on the Laws of Armed Conflict*. Oxford: Oxford University Press.
- Tversky, A. & Kahneman, D. (1974) 'Judgement Under Uncertainty: Heuristics and Biases', 185 *Science* 1124.
- UAS Vision (2019) DARPA Reveals Details of CODE Program. Wagner, M. (2014) 'The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems', 47 *Vanderbilt Journal of Transnational Law* 1371.
- Walker, P. (2021) 'Leadership Challenges from the Deployment of Lethal Autonomous Weapon Systems: How Erosion of Human Supervision Over Lethal Engagement Will Impact How Commanders Exercise Leadership', 188 *The RUSI Journal* 10.
- Winter, E. (2022) 'The Compatibility of Autonomous Weapons with the Principles of International Humanitarian Law', 27 *Journal of Conflict and Security Law* 1.
- Work, R.O. (2021) Principles for the Combat Employment of Weapon Systems with Autonomous Functionalities. Center for a New American Security.
- Xiao, B., *et al.* (2016) 'Computational Analysis and Simulation of Empathic Behaviors: A Survey of Empathy Modelling with Behavioral Signal Processing Framework', 18 *Current Psychiatry Reports* 49.
- Zerilli, J., *et al.* (2019) 'Transparency in Algorithmic and Human Decision Making: Is there a Double Standard?' 32 *Philosophy & Technology* 661.

Znanstveni članek

DOI: 10.51940/2024.1.163-184

UDK: 341.3:342.7:004.8

623.09:004.8

Yuval Shany

Uporabiti umetno inteligenco ali ne? Avtonomni orožni sistemi in njihovo zapleteno razmerje s pravico do življenja

Razširjenost tehnologij umetne inteligence, razvite ali prilagojene za vojaško uporabo, odpira zahtevna vprašanja o skladnosti teh novih tehnologij z mednarodnim pravom nasploh ter še zlasti z mednarodnim pravom človekovih pravic. Odbor za človekove pravice, strokovno telo, ki je zadolženo za spremljanje izvajanja Mednarodnega pakta o državljanskih in političnih pravicah, je leta 2018 podal svoje mnenje o razmerju med pojavom nove vojaške umetne inteligence in spoštovanjem pravice do življenja. Članek preučuje razprave v okviru mednarodnega prava človekovih pravic v zvezi z uvajanjem tehnologij umetne inteligence v vojaške kontekste in njihovim razmerjem s pravico do življenja. V prvem delu na kratko predstavi nekatere dejanske in možne uporabe umetne inteligence v vojaških okoljih. V drugem delu obravnava tri glavne ugovore zoper uvajanje umetne inteligence v bojna območja: zmožnost avtonomnih ali polavtonomnih sistemov umetne inteligence, da delujejo skladno s pravili mednarodnega humanitarnega prava, pomisleke glede dejanskega znižanja standardov humanitarne zaščite ter etične in pravne posledice prenosa nekaterih odločitev o življenju ali smrti z ljudi na stroje. V tretjem delu – ob upoštevanju teh treh načelnih ugovorov – avtor preuči konkretne predloge Mednarodnega odbora Rdečega križa za omejitev uporabe umetne inteligence v vojaških okoljih (omejitev področja in načina uporabe avtonomnih orožnih sistemov ter izključitev nepredvidljivih in smrtonosnih sistemov). V četrtem delu so glavna vprašanja, ki jih obravnava ta članek, preučena z vidika pravice do življenja po mednarodnem pravu človekovih pravic, kot jo pojasnjuje Splošni komentar št. 36.

Ključne besede

avtonomni orožni sistemi, pravica do življenja, mednarodno humanitarno pravo, človekovo dostojanstvo, odgovornost, preglednost, smiselni človeški nadzor, Mednarodni odbor Rdečega križa, vojaška umetna inteligenca.

Scientific Article

DOI: 10.51940/2024.1.163-184

UDC: 341.3:342.7:004.8

623.09:004.8

Yuval Shany

To Use AI or Not to Use AI? Autonomous Weapon Systems and Their Complicated Relationship with the Right to Life

Abstract

The increased prevalence of AI technology developed or adapted for military use raises difficult questions about the compatibility of this new technology with international law in general, and international human rights law (IHRL) in particular. The Human Rights Committee, the expert body entrusted with monitoring the application of the International Covenant on Civil and Political Rights, expressed its view in 2018 on the relationship between the emergence of new military AI and respect for the right to life. The article reviews the terms of the IHRL debate surrounding the introduction of AI technology into military contexts and its relationship to the right to life. Section one briefly reviews some actual and potential applications of AI in military contexts. Section two deals with three principal objections to introducing military AI to battlefield environments: the capacity of autonomous or semi-autonomous AI systems to properly apply international humanitarian law (IHL), concerns about *de facto* lowering of standards of humanitarian protection, and the ethical and legal implications of transferring certain life-and-death decisions from humans to machines. Section three reviews, in light of these three principled objections, specific proposals by the ICRC to limit the use of AI in military contexts (limiting the scope and manner of use of autonomous weapon systems, and excluding unpredictable and lethal systems). Section four reviews the main issues discussed in this article from the vantage point of the right to life under IHRL, as elaborated in General Comment No. 36.

Key words

autonomous weapon systems, right to life, international humanitarian law, human dignity, accountability, transparency, meaningful human control, ICRC, military AI.