sciendo International Conference KNOWLEDGE-BASED ORGANIZATION
Vol. XXX No 2 2024

"TERMINATOR" FACTOR – AI TARGETING DECISION-MAKING AND THE LAW

Sabin GUȚAN

"Nicolae Bălcescu" Land Forces Academy, Sibiu, Romania
sabin_gutan@yahoo.com

Abstract: *The development of technologies based on artificial intelligence (AI) is a major leap of society towards a new era which, in theory, should be a progress for humanity. But there was no time to discern what are the possible advantages and areas in which AI could assist the development of human society, nor to carefully calculate the disadvantages and risks of its use, that we found ourselves in the midst of a raging arms race based on AI. This new technology is developing in two directions: one of assisting human decision in battle and another that receives its own decision in battle. If the first does not raise big problems, the decision-making belonging to the human factor, the second is a new element in human society and involves major risks for it. From the divine supreme right over life and death, we have gone through the regulation of the right to life and the abolition of the death penalty, and in a mad rush of development, we transcend to a world where a machine becomes "providence" through its ability to decide alone over the life and death of a human being. This article proposes a solution so that such devaluation of the human being never happens.*

Keywords: international humanitarian law, artificial intelligence, AI decision, right to life, AI regulation

1. Introduction

The newest definition of AI systems is found in art.3 par.1 of the EU AI Act [1]: "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments". Presumably, this definition will eventually become a globally accepted one.

The purpose of this article is focused on the legal implications of the use of AI in the sphere of means and methods of warfare. And from this sphere I do not analyze all situations, but only those related to the replacement of the human factor, as a decision, with a non-human one. Respecting, assuming and correctly applying international humanitarian law (IHL) are exclusive attributes of the human being with full legal capacity. New technologies must obey IHL, as effects (as stipulated in art.36 of Additional Protocol I of Geneva of 1977), and people must assiduously check this aspect before equipping the armed forces with new means of combat. In this regard, although many states have undertaken this process of verifying the compatibility of new weapons with IHL, few states have implemented such procedures. Even more difficult is for states to verify and certify this compatibility in the machine learning and

DOI: 10.2478/kbo-2024-0050

auto-update process of AI [2]. Moreover, in war, a multitude of ruses of war and treacherous situations are used, which can negatively alter the machine learning process and lead to auto-updates inconsistent with reality and IHL. Responsibility is purely human and should not be transferred to machines [2]. In this sense, for the establishment of liability, a fundamental element is the establishment of the causal link between the human factor (as will) and the effects produced (as violation). These rules must also be applied when using AI.

As a starting point, I draw three broad directions that the debate focuses on, with respect to the object of study of this material: enabling the development and free use of AI; allowing the development and broad use of AI, but under the control of the law; limiting or prohibiting the development and use of AI in all sectors that could seriously affect the human being and the absolute control of the human factor over the operation of AI.

The doctrine also differentiates between kinetic and non-kinetic targeting. The first refers to the use of means and methods of combat with lethal effects in the physical domain, while the second refers to those military and non-military actions that do not produce such effects [3].

2. AI with Own Decision and the Idea of Artificial Person

It is not difficult to manage the phenomenon of liability in the field of using AI – only the human factor is responsible. What is difficult and, from my point of view, impossible to accept is the creation of a new individuality: *the artificial person*, with human-like attributes - learning, planning, adaptation, own decision, independent action. Indeed, it is fun and exhilarating to see robots on the news that look like humans, talk, simulate emotions, give elevated responses, even tell jokes. The same AI software used in war hunts and kills people. To software, telling a joke or killing a human is the same thing – just an algorithm. Even though there is talk of the impossibility of humans ever leaving the high-level decision to AI systems and the development of such systems that can always be controlled by the human factor [4], without rigorous regulation, as has always happened, it won't take long for the "curious" to appear and press the "button"! The human factor must be absolutely protected, and its replication must not be allowed. Even human cloning, the creation of hybrids and chimeras, have produced major controversies, emerging international norms that combat them (for example, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine, Oviedo 1997; Universal Declaration of Bioethics and Human Rights 2005; The UNESCO Bioethics Program – which also generated recommendations on the Ethics of Artificial Intelligence [5]). Technological cloning/simulation of human consciousness and endowment of an object with the ability to kill by its own decision, to self-improve in this activity, is the most dangerous and unacceptable thing ever invented by mankind.

Own decision for AI? When we think about what effects it could generate, nuclear weapons are no longer a problem. But an AI system with its own decision and access to a nuclear weapon is probably the last thing humanity can do. The fact that "developers are often unable to explain how AI technology arrived at a decision because of its complex evolving internal processes", the so-called "the black-box effect", [6] is a sufficient reason not to leave a non-human process to make important or vital decisions related to people. But in a controlled environment and with all safety measures, specialists can experience, understand, improve and control this process, as is the case with The Project Convergence program [7]. The integration of advanced

technology in military operations brings an indisputable advantage to the state that uses it, both in terms of direct combat with the enemy, but also in terms of the protection of human resources - but only if the adversary does not have this technology. Cautious use, with patience and careful attention to identifying and correcting errors, without a direct and proper action of the technology on targets, is a positive way to integrate AI into military and security strategy, as in the case of Task Force Lima [8] or Project Maven [9] of the USA.

Important to mention in this context are also the experiments carried out in the context of the use of AI in the decision process. In such an experiment "in high-stakes military and foreign-policy decision-making", carried out by American researchers [10], it was observed that "models tend to develop arms-race dynamics, leading to greater conflict, and in rare cases, even to the deployment of nuclear weapons". In this context, the authors recommend continuing research to eliminate unknowns and risks and, above all, great caution when wanting to use these technologies "for strategic military or diplomatic decision-making".

Another experiment confirmed by the US authorities [11], - marginally reviewing an experiment not assumed by the US authorities and controversial with regard to the veracity of its realization, namely a simulation in virtual space of a situation in which a military drone with its own decision, in a given context, decides to attack its own operator [12] - is the first real-space simulation of an aviation confrontation between a human pilot and an AI pilot. The conclusions are, in all cases, that AI can learn, adapt, generate scenarios and decide how to act, but these decisions are not yet certain, potentially posing major risks to humans.

Even in situations where such technologies have the human factor as a decision-making element, but are used in large-scale military operations, the results are not always consistent with IHL. This fact can be seen - without being able to resort to confirmed official data and in-depth scientific analysis, but for now to investigations and journalistic sources - in the current conflict in the Gaza Strip, where the Israeli army has used two AI systems to identify human and nonhuman targets: Lavender and the Gospel [13]. A very large number of civilian casualties, especially women and children, are reported in this conflict [14]. In the mentioned context, it is difficult to appreciate whether this number of victims can be correlated with the use of AI systems. But what can be seen is the discrepancy between the speed of work and the large number of targets identified by AI technology and the limited ability of the human factor to keep up with the verification of the legality of these targets, the protected elements and the punctual calculation of other impediments drawn by IHL in the process of planning and executing the attack. This delay in the reaction of the human factor can lead to ignoring or skipping steps in the human decision-making process, to unjustifiably increasing the level of confidence in the information provided by AI systems, and thus to make decisions that produce undesirable results much more often. The ICRC recommends, in this sense, the development of AI in such a way that it folds on the speed of human reaction, so as not to force the decision of the human factor [15]. In the specialized doctrine, there are studies on the level of control and what are the principles on which it can be based for it to be effective and to avoid such errors [4]. Likewise, attention is drawn to the danger that, given the limitations and unpredictability of this technology, human decision-making may be altered by the information provided and lead to a violation of IHL [15].

3. AI and IHL

Responsibilities outlined by IHL address combatants, in general, and commanders, in particular, for the act of decision and for the conduct of subordinates. Objects have no obligations in IHL, they can be military objectives, means of combat or protected goods. Military operations and precautions in attack are human actions under the power of law, and means and methods of warfare are ways in which combatants carry out their missions. Combatant privilege is granted to a human person with legal combatant status, and never to an object. There is no way for a combatant to convey, delegate or mandate this quality to another person, and it is hard to imagine, from a legal point of view, giving the combatant privilege to an object [15]. However, IHL is not directed at making a concrete connection between the notion of combatant and the human person, but rather with the membership of an organization, conferred based on national laws. From here can come the danger of associating some AI technologies with the idea of a person or a citizen, especially since there are already examples of robots that have been given an identity specific to a human being (name, even citizenship) [16]. It would be very important for a future regulation to ban the use of the human form for AI software.

For a military AI technology to act properly, it must have the entire IHL with a multitude of practical examples transposed into the algorithm system. Naturally, in this case AI will be much more capable of identifying, from the multitude of norms and examples, the possible prohibitions, but their interpretation will always be algorithmic and devoid of the value of human consciousness. Human consciousness is based on inner processes of human experiences that we still do not fully understand, let alone replicate. Many rules of IHL raise problems of interpretation. From state to state, from doctrine to doctrine, to specialists, jurisprudence and combatants, the interpretation of norms is diverse and uneven, but it is still based on human consciousness that allows staying within certain limits. It will be much harder for the AI to apply legal presumptions in cases of uncertainty, to readjust and correct a decision when and if it finds that it was wrong, to apply the rules of protection, to assess an act as treacherous or not, not to generalize certain experiences, to interpret the notion of direct participation in hostilities or to decide on dual-use objects, the application of proportionality, etc. [6]. How will the AI deal with a person it has injured or surrendered? It is possible that the AI, over time, will be able to learn these things and even be able to decide better than a human and even help improve the protection rules. But how long and with how many failures? On the other hand, the doctrine also supports the possible advantage of AI's emotional non-involvement, which could lead to better compliance with IHL and thus to a considerable decrease in the number of victims and destruction [6].

Nor has the problem of cyber war (in virtual space) been solved, international regulations being late to appear, being in the drafting phase of the Tallinn Manual 3.0[17], that the unification of the two worlds has occurred – cyber war has been brought into the real world through AI. Law and society in general are not ready to accept a new intelligent entity that we can consider a *person*, especially since it is *artificial*. It is much more difficult and serious to accept and use in armed conflicts the idea of an *artificial combatant*. In this context, the problems of the use of AI by the criminal environment, in peacetime or wartime, which will create unprecedented problems at the national, regional and global level, must also be addressed. The idea that the *artificial terrorist* will be in every pocket, in every house, in every

institution, etc. raises security risks to an unprecedented level.

4. Banning AI in Armed Conflict

To accept the penetration into human society of machines with the ability to decide on the life and death of a human being, we must ensure all social guarantees by establishing clear legal norms and inserting a behavior according to these norms for AI, but it must also be clearly certified that technology respects them and cannot deviate from them [4]. Not all systems in this sphere pose a danger when operating autonomously. They are systems that have predetermined autonomy, especially defensive ones, or systems that work in the humanitarian area or assisting human actions [15]. These may be covered by existing IHL norms, with efforts being made when designed and tested to eliminate the possibility of error.

Problems are with systems that incorporate learning and self-adaptive AI. They start from a preset level that may be highly secure and IHL compliant, but along the way they rewrite themselves certain parameters of operation and action. For these there is no guarantee that through this process they will not become dangerous to civilians and civilian property, to other protected categories and, depending on the controlled arsenal and combat capability, may even pose a danger to large masses of people or even to humanity.

Limiting the development of AI must be done for any area that could harm the human being. Regarding military technologies, the international regulatory approach is to be welcomed, under the power of *The Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*. But the process of building a new protocol is arduous, agreeing on 11 general principles to govern future regulation, addressing, among others, the importance of the human factor in decision-making, compliance with IHL, identifying and combating risks. [2]

In the EU AI Act, in art.3, the following definition is found: "(49) 'serious incident' means an incident or malfunctioning of an AI system that directly or indirectly leads to any of the following:

(a) the death of a person, or serious harm to a person's health;

(b) a serious and irreversible disruption of the management or operation of critical infrastructure.

(c) the infringement of obligations under Union law intended to protect fundamental rights;

(d) serious harm to property or the environment" [1].

These incidents may occur in times of armed conflict, but are related to decision, intent, human error and the context of war, precautions in attack and military imperatives. LAWS regulation must contribute to the reduction of these incidents, not a justification for their multiplication.

The UN agenda for the regulation of AI also includes the assumption, in the form of a recommendation: "Building on the progress made in multilateral negotiations, conclude, by 2026, a legally binding instrument to prohibit lethal autonomous weapon systems that function without human control or oversight, and which cannot be used in compliance with international humanitarian law, and to regulate all other types of autonomous weapons systems" [18].

In this context, a new IHL principle must be born, namely *the principle of human decision*. This must be the foundation for the development and use of all means and methods of war, being, as the ICRC [15] states, "essential to preserve human control over tasks and human judgment in decisions that may have serious consequences for people's lives in armed conflict" and "AI and machine learning

systems must be used to serve human actors, and as tools to augment human decision-makers, not replace them".

A future regulation should aim both at limiting the use of AI technologies in armed conflicts, but also at prohibiting the development and use of certain AI applications, which are in direct connection with the level of decision-making and the actual capacity to respect IHL:

- direct human decision in selecting and attacking targets must be the only legal process;
- the innovation, production, possession, transmission or use of combat systems with AI decision should be prohibited;
- the prohibition of leaving the decision on the fate of a human being outside the human factor, including life and death, dignity, detention or the application of punishments;
- the technological simulation of humanism and the human factor must be prohibited;
- prohibiting the connection of AI technologies that can develop and act on their own to weapon systems and weapons;
- all military AI technologies must be provided with automatic shutdown or self-destruction systems for failure situations that are dangerous to the population and civilian property.

5. Conclusions

AI technologies, with their own decision, used to kill people, directly or indirectly, or that have a high potential to do so must be strictly prohibited, and any activity in this area must be considered a crime and regulated as such. AI can also bring many benefits to society, but the way this technology works must be strictly controlled. Mankind is making major efforts to create a peaceful society, to outlaw war of aggression, to disarm, and generally to secure life on Earth. It is a logical deduction, looking forward, that this technology will be widely used by states in the military field. This has happened with all technologies. One thing is clear: we all live on the same planet, and we have no other. If states do not find solutions to understand each other and strengthen international balance, peace and security, we will destroy ourselves. Why? Because we can, and from one day to the next these possibilities are diversifying more and more. The more technological is the war, the more it will destroy. AI will allow the combination and simultaneous use of all combat capabilities and, more than likely, the development of new ones that are much more effective in destroying and killing.

The question is whether IHL successfully covers the use of AI in armed conflict, or whether or not we should ever come to a practical answer to this problem. We can appreciate that IHL norms can cover a wide range of military actions, even those conducted with new weapons or in new spaces. But all these norms have a primary element, in terms of their application and interpretation: the human factor. This factor, even if it has a variable percentage of error, influenced by good faith, is generally accepted, sufficiently safe and balanced, but also controlled and burdened by the sanctioning system. Despite all the tragedies produced by wars or dictatorial regimes, we can appreciate that the level of security and balance of life on this planet is still quite high.

Even if we can create algorithms for AI to simulate decision-making, the human subjective element is irreplaceable. The doctrine supports the idea of innovation along the lines of developing *autonomous moral weapons* [4], capable of making ethical decisions about attacking a target. Until we get to talking about artificial wisdom, artificial reason, artificial consciousness, artificial morals and ethics, artificial prejudice, artificial assumption and responsibility, and many other factors that make the human mind and spirit work as a unified whole and discern right from wrong, and choose the good, it will take a

long time. This is AI, just an appearance, so dangerous that it should be considered the greatest threat to humanity. The benefits are too insignificant for the harm it can do. It's not something we can't live without, human intelligence is magnificent and can find solutions to all human needs. We just need patience and effort. AI is nothing more than a drug doping human society, it will create collective addiction and, ultimately, collapse.

After all, the idea of "artificial intelligence" is an aberration. It's not intelligence; this is a human characteristic and should be preserved as such. It is a right to human identity that we must not share with objects (software). As much as some may boast and as much as they try to convince us that soon people will be dumber than objects, this technology is not intelligent and should not be confused with it. It's just a data management system, with the ability to produce results in a very short time, which should be called something else, for example, "fast data processing system". In any case, it should not be confused with anything related to human beings. Accepting the idea of "artificial intelligence" we must also accept the idea of "artificial stupidity", and it can be seen that this technology has enough.

It is normal for there to be controversy regarding the use of AI in the military. In the domestic field, the situation is no different, the danger of using AI being confirmed by the new regulation of the European Union, which places special attention on the protection of human rights, democracy and the rule of law, on human control over technology, on the idea of trust and risk management, liability in the field, but also promotes responsible innovation. This document has the value of an international treaty and is open for signature by non-EU states as well [1]. However, these rules do not have direct applicability in the military domain, so the use of AI in times of armed conflict remains only under the power of IHL.

In fact, we must focus on finding relevant solutions to strengthen international peace and security. AI must remain only a tool to enhance human decision-making, not a substitute for it. We don't need killer robots, we need happy children!

References List

[1] EU AI Act 2024, https://artificialintelligenceact.eu/; https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law.

[2] House of Lords, AI in Weapon Systems Committee – Report of Session 2023-24, *Proceed with Caution: Artificial Intelligence in Weapon Systems*, published by the Authority of the House of Lords, UK, https://committees.parliament.uk/committee/646/ai-in-weapon-systems-committee/publications/

[3] Paul A. L. Ducheine, *Non-kinetic Capabilities: Complementing the Kinetic Prevalence to Targeting*, Chapter in book: *Targeting: The Challenges of Modern Warfare* edited by Paul A.L. Ducheine, Michael N. Schmitt, Frans P.B. Osinga, Published by t.m.c. asser press, The Hague, The Netherlands www.asserpress.nl, 2016, pp. 201-230, DOI 10.1007/978-94-6265-072-5;

[4] Mateus De Oliveira Fornasier, *THE REGULATION OF THE USE OF ARTIFICIAL INTELLIGENCE (AI) IN WARFARE: between International Humanitarian Law (IHL) and Meaningful Human Control*, May 2021, Revista Jurídica da Presidência 23(129):67, DOI: 10.20499/2236-3645.RJP2021v23e129-2229;

[5] https://www.unesco.org/en/artificial-intelligence/recommendation-ethics?hub=387;
[6] Anastasia Roberts and Adrian Venables, *The Role of Artificial Intelligence in Kinetic Targeting from the Perspective of International Humanitarian La*w, 2021 13th International Conference on Cyber Conflict (CyCon), Tallinn, Estonia, 2021, pp. 43-57, doi: 10.23919/CyCon51939.2021.9468301;
[7] https://www.defense.gov/Spotlights/Project-Convergence-Capstone-4/;
[8] https://www.dds.mil/akdlskandkanddnknajdjkajdksjskjdskjnnfeujksks;
[9] https://www.bloomberg.com/news/newsletters/2024-02-29/inside-project-maven-the-us-military-s-ai-project;
[10] Juan-Pablo Riveraa, Gabriel Mukobib, Anka Reuelb, Max Lamparthb, Chandler Smithc, Jacquelyn Schneider, *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*, arXiv:2401.03408v1 [cs.AI] 7 Jan 2024, https://arxiv.org/pdf/2401.03408, https://doi.org/10.48550/arXiv.2401.03408;
[11] https://www.airandspaceforces.com/usaf-first-ai-against-human-dogfights/;
[12] https://defensescoop.com/2023/06/14/what-the-pentagon-can-learn-from-the-saga-of-the-rogue-ai-enabled-drone-thought-experiment/;
[13] [https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets; https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes;
[14] https://www.ohchr.org/en/press-releases/2024/05/onslaught-violence-against-women-and-children-gaza-unacceptable-un-experts;
[15] ICRC, *Artificial intelligence and machine learning in armed conflict: A human-centred approach*, REPORTS AND DOCUMENTS, International Review of the Red Cross (2020), 102 (913), 463–479., Digital technologies and war, doi:10.1017/S1816383120000454;
[16] Sati, Mustafa Can, Attributability of Combatant Status to Military AI Technologies under International Humanitarian Law (September 11, 2023). Global Society, 2023, Available at SSRN: https://ssrn.com/abstract=4607136;
[17] https://ccdcoe.org/research/tallinn-manual/;
[18] UN, *Our Common Agenda Policy Brief 9, A New Agenda for Peace*, JULY 2023, https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf.