Marcin Marcinko

ORCiD: 0000-0001-6495-9606

DO AUTONOMOUS WEAPON SYSTEMS NEED SPECIFIC RULES OF ENGAGEMENT?*

CZY AUTONOMICZNE SYSTEMY UZBROJENIA WYMAGAJĄ SZCZEGÓLNYCH ZASAD PODEJMOWANIA DZIAŁAŃ PRZY UŻYCIU SIŁY (ROE)?

There is no doubt that future combat will see AWS as new players in battlespace.

Markus Wagner

Streszczenie

Wykorzystanie w pełni autonomicznych systemów uzbrojenia (*autonomous weapon systems* – AWS) w operacjach wojskowych rodzi istotne wyzwania natury prawnej i etycznej. Wyzwania te dotyczą również kwestii technicznych związanych z tym, czy takie systemy będą w stanie przestrzegać obowiązujących zasad użycia siły (*rules of engagement* – ROE), które są tradycyjnie opracowywane z myślą o ludziach. Operacje militarne realizowane są w bardzo złożonym kontekście prawnym, wojskowym i politycznym, co budzi wątpliwości, czy autonomiczne systemy sterowane sztuczną inteligencją będą w stanie uwzględnić liczne zmienne niezbędne do zapewnienia zgodności z zakazami i ograniczeniami wynikającymi z misji bojowych. Niniejszy artykuł wyjaśnia, czy AWS powinny być objęte tymi samymi ROE, które mają zastosowanie do konwencjonalnych rodzajów broni, czy też wymagają odrębnych ROE dostosowanych do ich unikalnych cech operacyjnych. Kluczowe w tym kontekście jest pytanie, czy AWS mogą skutecznie funkcjonować w nieprzewidywalnym i dynamicznym środowisku wojennym, biorąc pod uwagę złożoność zadań oraz cykliczny charakter procesu *targetingu*. Poza tym, problemy takie jak syndrom „czarnej skrzynki" podkreślają nieprzewidywalność podejmowania decyzji przez sztuczną inteligencję, co dodatkowo komplikuje dostosowanie ROE do AWS. Chociaż empiryczne przykłady zastosowania AWS są ograniczone, analiza teoretyczna oparta na elementach ROE oraz definicyjnych cechach AWS – autonomii, sztucznej inteligencji i algorytmicznym podejmowaniu decyzji – może zaowocować istotnymi spostrzeżeniami i wnioskami dotyczącymi omawianej materii.

* This research was funded in whole by the Polish National Science Center, grant number 2021/43/B/HS5/02324.

słowa kluczowe: zasady użycia siły; ROE; autonomiczne systemy uzbrojenia; AWS; operacje wojskowe; sztuczna inteligencja; uczenie maszynowe; algorytmy

1. Introduction

The rapidly advancing development of military technologies and the changing face of modern wars inevitably lead to the "robotization" of military operations, in which humans will increasingly play the role of passive observers. From a historical perspective, fighting at a distance has always been an inseparable part of military operations, and this fact will remain the same in the future. However, it will lead to a qualitative change—the 20th-century challenges of distance of war will be replaced by challenges associated with "disconnecting" soldiers from the battlefield.[1] Conducting armed conflicts at a distance, using remotely controlled platforms aimed at gathering intelligence or transferring weapons, has already become commonplace in the most developed armies in the world.[2] The means of warfare heralding the next step in this direction will be all kinds of combat robots, which, according to the forecasts of experts, will become a common means of warfare on land, in the air and on water, and possibly also in outer space.[3]

Based on the achievements of robotics to date, it can be assumed that machines will take over some of the tasks performed so far by soldiers because they do not have negative psychophysical features typical of humans (fatigue, discouragement, succumbing to emotions and fear, fear of losing life or damaging the body). In other words, they will never act out of panic, vengeance or hatred, thus, the risk of excessive behavior in emotionally stressful situations is non-existent. Additionally, due to their lack of emotions or physical exhaustion, they would be much more capable of carrying out 3D (dull, dirty and dangerous) work. Furthermore, machines have divisible/shared intelligence and can calculate and operate at digital speed (so they are assumed to be far more capable of collecting and processing new in-

---

[1] B. Sajduk, *Problem walki na odległość w perspektywie historycznej, społecznej i etycznej* [*The problem of remote warfare in a historical, social and ethical perspective*], [in:] K. Kowalczewska, J. Kowalewski (eds), *Systemy dronów bojowych. Analiza problemów i odpowiedź społeczeństwa obywatelskiego* [*Combat drone systems. Problem analysis and civil society response*], Wydawnictwo Naukowe Scholar, Warszawa 2015, p. 17.

[2] K. Kowalczewska, *Sztuczna inteligencja na wojnie. Perspektywa międzynarodowego prawa humanitarnego konfliktów zbrojnych* [*Artificial intelligence at war. The perspective of international humanitarian law of armed conflict*], Wydawnictwo Naukowe Scholar, Warszawa 2021, p. 24; J. Kwik, T. van Engers, *Algorithmic fog of war: When lack of transparency violates the law of armed conflict*, "Journal of Future Robot Life" 2021, vol. 2, № 1-2, p. 3.

[3] W. Bieńkowski, *Roboty wojskowe w zastosowaniach militarnych. Zastrzeżenia natury prawnej i etycznej* [*Military robots in military activities. Legal and ethical objections*], [in:] M. Szuniewicz (ed.), *Automatyzacja i robotyzacja współczesnego pola walki wyzwaniem dla prawa międzynarodowego* [*Automation and robotization of the modern battlefield as a challenge to international law*], Akademia Marynarki Wojennej, Gdynia 2015, p. 30.

formation than humans), and they do not forget the orders. Machines may also expand the battlefield by facilitating penetration beyond enemy lines and maintaining a presence in the operational theatre for significantly longer durations than humans are capable of. It is also noted that combat robots will not endanger civilian lives by exposing them to hostile fire, as there is no need for self-defense, and will minimize their own losses, thereby contributing to the mitigation of the negative impacts of armed conflicts[4].

Indeed, "the deployment of weapon systems with increasingly autonomous modes or functions, particularly small armed drones and loitering munitions[5], is a fact of contemporary conflicts."[6] However, the use of combat robots that can independently carry out tasks as part of military operations, and in particular use force, raises legal and ethical controversies. There are also doubts as to whether, from a technical point of view, such robots will be able to cope with the rules of using force adopted for combat missions, whether appropriate changes will be required to adapt the details of their operation to specific, but also situation-oriented Rules of Engagement (ROE)[7], and if so, in what direction these changes should go. Contemporary military operations take place in a very complex legal, military and political environment – the question arises whether fully autonomous weapons systems, equipped with artificial intelligence and guided by an algorithm, will be able to make decisions in a way that takes into account dozens of variables while ensuring compliance with the prohibitions and restrictions adopted for the implementation of a combat mission.[8]

---

[4] *Ibidem*, p. 33; R. Geiß, H. Lahmann, *Autonomous weapons systems: a paradigm shift for the law of armed conflict?*, [in:] J.D. Ohlin (ed.), *Research Handbook on Remote Warfare*, Edward Elgar Publishing, Cheltenham 2017, p. 372–373; C. Gutiérrez Espada, M.J. Cervell Hortal, *Autonomous Weapons Systems, Drones and International Law*, "Revista del Instituto Español de Estudios Estratégicos" 2013, Núm. 2, pp. 2–3; K. Kowalczewska, *op. cit.*, p. 162; K. Kowalczewska, B. Kijewska, *War Algorithms in Modern Deliberative Democracies: Parliamentary Technology Assessment as a Public Conscience Discovery Tool?*, "Teoria Jurídica Contemporânea" 2022, V. 7, pp. 6–7.

[5] This is "a kind of aerial weapon that can hover over, detect and dive onto targets, and detonates on impact" (ICRC, *International Humanitarian Law and the Challenges of Contemporary Armed Conflict*s, Report of September 2024, 34IC/24/10.6, p. 51, footnote 179.

[6] *Ibidem*, p. 51.

[7] That is, "directives to military forces, including individuals, that define the circumstances, conditions, degree, and manner in which force, or actions which might be construed as provocative, may be applied" (NATO, *AAP-06(2021) – NATO Glossary of Terms and Definitions*, December 15, 2021, p. 113).

[8] P. Łubiński, *Ku zwiększonej autonomiczności. Postępująca autonomiczność kluczowych procesów decyzyjnych w systemach bojowych* [*Towards increased autonomy. Progressing autonomy of key decision-making processes in combat systems*], [in:] K. Kowalczewska, J. Kowalewski (eds), *Systemy dronów bojowych. Analiza problemów i odpowiedź społeczeństwa obywatelskiego* [*Combat drone systems. Problem analysis and civil society response*], Wydawnictwo Naukowe Scholar, Warszawa 2015, pp. 184–185.

The international community recognizes potential problems and challenges related to the military use of the so-called autonomous weapon systems (AWS), and numerous initiatives aimed at regulating this issue have been undertaken by states and various international bodies for over a decade. It can be assumed that initial discussions at the international level were stimulated by the reports of two successive Special Rapporteurs on extrajudicial, summary, or arbitrary executions to the UN General Assembly and the Human Rights Council in 2010 and 2013 respectively, focusing on the legal, ethical, and policy implications of emerging lethal robotic technologies used in armed conflicts and law enforcement contexts.[9] The parties to the Convention on Conventional Weapons (CCW)[10] in turn initiated informal expert meetings. During their fifth Review Conference in 2016, they made the decision to form an open-ended Group of Governmental Experts (GGE) dedicated to examining emerging technologies within the realm of lethal autonomous weapon systems (LAWS). The GGE was assigned the responsibility of investigating and reaching a consensus on potential recommendations and options within the framework of the Convention, which encompassed not only assessing the compliance of LAWS with international humanitarian law (IHL) and international human rights law (IHRL) but also considering broader factors such as their impact on regional and global security, effects on the threshold for armed conflicts, risks of an arms race, proliferation risks, and the threats posed by cyber operations related to LAWS.[11]

Despite such widespread interest in the issue of autonomous weapon systems[12], the actions taken so far have not led to the development of an international legal instrument that would regulate the use of these systems on the battlefield. It is worth emphasizing, however, that in 2019, as a result of the GGE's efforts, the CCW member states approved 11 guiding principles in the

[9] See: UN General *Assembly, Interim Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Philip Alston*, August 23, 2010, UN Doc A/65/321; UN General Assembly, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns*, April 9, 2013, UN Doc A/HRC/23/47.

[10] 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 (and Protocols) (As Amended on December 21, 2001), October 10, 1980, 1342 UNTS 137.

[11] M. Kunz, S. Ó Héigeartaigh, *Artificial Intelligence and Robotization*, [in:] R. Geiß, N. Melzer (eds), *The Oxford Handbook of the International Law of Global Security*, Oxford University Press, Oxford 2021, pp. 631–632.

[12] In this context, it is worth mentioning that currently, about 80 States participate in the work of the GGE. Among them are the permanent members of the UN Security Council, member states of the European Union, as well as numerous non-governmental organizations and academic institutions, including Amnesty International, the International Committee of the Red Cross, and the United Nations Institute for Disarmament Research (UNIDIR). See G. Garcia Vera, *Funkcjonowanie autonomicznych systemów zbrojnych w kontekście regulacji multilateralnych* [*Functioning of Autonomous Weapons Systems in the context of multilateral regulations*], "Journal of Modern Science" 2023, vol. 1/50, p. 484.

field of autonomous weapon systems. The GGE was also given a mandate to develop recommendations for clarifying, considering, and developing normative and operational frameworks for AWS.[13] This initiative, like many others, primarily focuses on the legal aspects of the use of AWS in military operations. There is no doubt that IHL must be central to this debate, and there is consensus that AWS can be only deployed insofar as they abide by the IHL principles and rules of conducting hostilities, namely, distinction, proportionality and precautions. Nevertheless, there remains a lack of in-depth research on whether, and to what extent, the use of AWS could be regulated within the framework of the ROE. Given that AWS might be employed in combat missions in the future, their actions should comply not only with humanitarian law but also with the ROE, which constitute the primary and most important document regulating the use of force in military operations. Perhaps it may be deemed premature to discuss ROE methodology in relation to weapons that do not yet exist. However, the need to integrate ROE into the programming of such weapons suggests that these considerations should be addressed even at this stage, enabling the anticipation of such requirements in current research and development efforts.[14]

Taking the above into account, it is necessary to consider the research problem—whether it will be feasible to develop ROE that account for the specific nature of AWS and if the answer is affirmative, to what extent the use of AWS could be regulated within the framework of the ROE. The research hypothesis is based on the belief that the ROE formulated thus far, being addressed solely to humans, are not adequate for application to machines such as AWS. Therefore, there is a need to demonstrate whether future "fully" autonomous weapon systems should be treated like other types of weapons to which so-called weapon-specific ROE apply, or whether they would require ROE specifically dedicated to them (and distinct from those issued for the human agents within the force they are a part of). Should the latter be the case, a series of further questions arise, such as: will these systems be capable of adhering to ROE established for military opera-

[13] See *Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, September 25, 2019, CCW/GGE.1/2019/3. See also International Committee of the Red Cross, *Commentary on the 'Guiding Principles' of the CCW GGE on 'Lethal Autonomous Weapons Systems'*, Geneva, July 2020, https://documents.unoda.org/wp-content/uploads/2020/07/20200716-ICRC.pdf (accessed: 6/20/2023).

[14] J.F.R. Boddens Hosang, *Control Through ROE in Military Operations: Autonomous Weapons and Cyber Operations as Reasons to Change the Classic ROE Concept?*, [in:] R. Bartels, J.C. van den Boogaard, P.A.L. Ducheine, E. Pouw, J. Voetelink (eds), *Military Operations and the Notion of Control Under International Law: Liber Amicorum Terry D. Gill*, T.M.C. Asser Press, The Hague 2021, p. 416.

tions? In particular, it is essential to consider that military operations are conducted in a complex and dynamic environment characterized by a certain degree of unpredictability and uncertainty, which may pose challenges to the application of AWS. Additionally, the analysis must take into account the nature and complexity of tasks within operations and the targeting procedure, which, contrary to common belief, is not a binary process but rather a multi-phase and cyclical one.[15] Furthermore, there is the issue of how autonomous drones make decisions, including the so-called "black box" syndrome and the resulting unpredictability of AI-driven machines. Thus, what changes will be necessary to adapt situation-oriented ROE to the automated decision-making processes upon which both existing and potential future machine-learning-based systems rely?

To address the above questions, a desk research methodology was employed, utilizing a qualitative method. A range of legally binding and non-binding documents issued by international governmental organizations such as the UN, NATO, and the EU were analyzed, alongside reports, recommendations, and opinions from non-governmental organizations as well as international and national expert bodies. This analysis was complemented by a review of the literature concerning both the definition and components of ROE, as well as various aspects (primarily legal and military) of the potential use of AWS in military operations.

It is important to note that studying ROE presents certain challenges, as generally States and international organizations involved in establishing ROE regard them as sensitive information, often classified as confidential. Consequently, access to these rules is restricted to a select group, and their disclosure to external parties is prohibited. The reason is quite obvious—since ROE outline the situations where force may be used or actions that could be viewed as provocative, disclosing these regulations, including both standing and long-term ROE, would provide a considerable advantage to an adversary in an armed conflict or any other military operation.[16] In other words, the confidential nature of ROE primarily stems from the concern that the knowledge of applicable ROE could be used against the armed forces adhering to these rules. Thus, in principle, limited access to the source material of the subject matter impedes any significant examination of ROE. Fortunately, secondary sources significantly facilitate the study of these rules,

[15] See, e.g., NATO, *AJP-3.9 Allied Joint Doctrine for Joint Targeting*, Edition B, version 1, November 2021.

[16] J.F.R. Boddens Hosang, *Rules of Engagement and the International Law of Military Operations*, Oxford University Press, Oxford 2020, p. 3; M. Faix, *Rules of Engagement – Some Basic Questions and Current Issues*, "Czech Yearbook of International Law" 2010, Vol. 1, p. 134.

namely, a considerable number of conceptual articles (and in recent years also monographs) devoted to ROE and related issues. The value of these publications is particularly high because in most cases, the authors of these works are military lawyers who possess not only theoretical knowledge in the field of ROE but also have practical experience in this matter, as they often had the opportunity to develop ROE for the armed forces in which they served as legal advisors.[17] It is also worth mentioning that in the case of NATO, ROE for specific operations are developed based on a standardized (and unclassified) set of ROE set out in the Military Committee Decision MC 362/1 *NATO Rules of Engagement*.[18] Furthermore, very helpful, both for scientific researchers and military practitioners, is the *Sanremo Handbook on Rules of Engagement*, prepared under the auspices of the International Institute of Humanitarian Law in 2009.[19]

As for the literature on AWS, an abundance of studies on the subject has emerged in recent years. However, from the perspective of the presented research problem, there is a notable lack of detailed publications addressing the possibility of applying or formulating ROE for military operations involving AWS. The only publication that touches upon this issue (albeit at a general level, serving as a starting point for further research) is the 2021 article by J.F.R. Boddens Hosang, which is cited in this work. Furthermore, neither examples nor practical experience are available to assist in this matter, as such examples have yet to materialize. Nevertheless, certain observations and conclusions can be drawn based on theoretical considerations, grounded on the one hand in the components of ROE and their role in military operations, and on the other hand in the developed definitions of AWS and the aspects characterizing these systems—namely autonomy, artificial intelligence, and the concept of algorithms.

2. Rules of Engagement as a *sui generis* code of conduct in military operations

The implementation of tasks entrusted to the armed forces as part of a military operation requires not only detailed training and appropriate equipment but also a complex and multi-faceted system of planning, involving

[17] Among the authors worth mentioning are J.F.R. Boddens Hosang, C. Guldahl Cooper, G.P. Corn, R. McLaughlin, and G.D. Solis. See also R. McLaughlin, *Protecting Civilians in Armed Conflict Through Rules of Engagement*, [in:] I. Primoratz, D.W. Lovell (eds), *Protecting Civilians During Violent Conflict: Theoretical and Practical Issues for the 21st Century*, Routledge, Farnham 2012, p. 94 and the literature cited there.

[18] North Atlantic Military Committee, *NATO Rules of Engagement*, Military Decision, MC 362/1, June 30, 2003.

[19] A. Cole, Ph. Drew, R. McLaughlin, D. Mandsager (Drafting Team), *Sanremo Handbook on Rules of Engagement*, International Institute of Humanitarian Law, Sanremo 2009.

many elements and resulting in a number of operational documents.[20] The previously mentioned ROE are just one component of the overall operational planning process and one factor influencing the conduct of the operation, including the use of force. Although they are the primary and most important document regulating the use of force within the operation, they are not the only ones. In addition to the ROE and their associated instruction cards, various operational directives can influence how the ROE are implemented or complement them in directing and controlling the use of force. The most significant of these documents include targeting and tactical directives, as well as special instructions.[21]

Focusing exclusively on ROE—which are the subject of consideration in this article—they are specific directives that define the degree and manner in which force may be used and are designed to ensure that the use of force is controlled and legal. The ROE is therefore intended to provide commanders at the operational and tactical level with greater control over the implementation of the military operation by their units.[22] They are not, however, tactical rules—ROE do not instruct soldiers on how to perform missions, although the tactics and ROE undoubtedly complement each other. ROE are "designed to provide boundaries and guidance on the use of force that are neither tactical control measures nor substitutes for the exercise of the commander's military judgment."[23] They inform commanders of the restrictions imposed and the degree of freedom they have in the implementation of the mission.[24]

The definitions of "Rules of Engagement" may, of course, differ in some details, depending on who formulates these rules and for what needs. This fact was noticed by the authors of the ROE Manual, published by the International Institute of Humanitarian Law in Sanremo in 2009.[25] According to the authors of the manual, "ROE are issued by competent authorities and assist in the delineation of the circumstances and limitations within which military forces may be employed to achieve their objectives."[26] The authors of the handbook also explain that ROE appear in a variety of forms in national military doctrines, including execute orders, deployment orders, operational

[20] J.F.R. Boddens Hosang, *Rules of Engagement…*, p. 22.

[21] *Ibidem*, p. 24.

[22] G.S. Corn, G.P. Corn, *The Law of Operational Targeting: Viewing the LOAC Through an Operational Lens*, "Texas International Law Journal" 2012, vol. 47, Issue 2, p. 354.

[23] G.D. Solis, *The Law of Armed Conflict: International Humanitarian Law in War, 3rd ed.*, Cambridge University Press, Cambridge 2022, p. 373.

[24] B.F. Klappe, *International Peace Operations*, [in:] D. Fleck (ed.), *The Handbook of International Humanitarian Law, 2nd ed.*, Oxford University Press, Oxford 2008, para. 1320.

plans, or standing directives. Whatever their form, they provide authorization for and/or limits on, among other things, the use of force, the positioning and posturing of forces, and the employment of certain specific capabilities. In some countries, ROE have the status of guidance to military forces, in other countries, ROE are lawful commands.[27] In general, however, we may distinguish two key elements of ROE, defined both explicitly or implicitly: that the rules for the use of force are directives issued by specific supreme bodies and addressed to their armed forces, and that they contain, at least, rules regarding the use of force.[28]

Properly constructed ROE allow for a harmonious combination of the needs resulting from military activity and limitations resulting from the application of the law of armed conflict (LOAC).[29] Being a kind of conglomerate of prohibitions and permits, the principles of the use of force should balance political aspects that are to ensure the achievement of the assumed goal, legal considerations responsible for the compliance of the developed procedures with the norms of armed conflict law and the requirements of military necessity aimed at minimizing own losses. Importantly, the rules for the use of force are structured to be applied in a specific military operation—if it is carried out in an armed conflict, ROE also take into account the specificity of the conflict, including its international or non-international character. Moreover, in modern military operations, ROE are more restrictive than the law of armed conflict, which results from the need to take into account the political aspects related to the use of military force. This is particularly important in the case of non-international armed conflicts, where military actions may be intertwined with law enforcement activities, and the purpose of the opera-

[25] The objective of the authors of the said manual was to create possibly unified, comprehensive, realistic and open guidelines that each State can use both for training and teaching purposes, and as a model when developing ROE for specific military operations (D. Mandsager, *Preface*, [in:] A. Cole *et al.* (Drafting Team), *op. cit.*, p. v). As a result, the content of the handbook reflects the best practices in the application of ROE by various countries of the world (M. Moreno, *Foreword*, [in:] *ibidem*, p. ii). It is worth adding that in 2022, the second edition of this handbook was released, incorporating the experience and lessons learned from the application of the first edition over more than a decade. The so-called Newport ROE Handbook builds on the successes of the first edition of the handbook and the revision includes several significant changes as reflected in the summary of changes (see D. Mandsager, A. Cole, Ph. Drew, R. McLaughlin (Drafting Team), *Newport Rules of Engagement (ROE) Handbook*, "International Law Studies" 2022, Vol. 98).

[26] A. Cole *et al.* (Drafting Team), *op. cit.*, Part I, para. 3.

[27] *Ibidem*.

[28] R. McLaughlin, *Protecting Civilians in Armed Conflict Through Rules of Engagement*, [in:] I. Primoratz, D.W. Lovell (eds), *Protecting Civilians During Violent Conflict: Theoretical and Practical Issues for the 21st Century*, Routledge, Farnham 2012, p. 95.

[29] A. Makowski, *Zasady podejmowania działań przy użyciu siły w operacji morskiej w świetle podręcznika ROE* [*Rules of Engagement in naval operation in the light of ROE Manual*], "Międzynarodowe Prawo Humanitarne" 2012, Vol. III, p. 21.

tion is not necessarily to physically defeat the enemy, but rather to win the support of the local population. In such circumstances, properly designed and selected ROE can contribute to the use of force in a manner appropriate to the situation and take into account the goals of the military operation conducted in the context of an armed conflict. In other words, ROE, formulated for the needs of each military operation, are directives specifying the circumstances, conditions, degree, and manner of using force, both by military units and by any individual soldier.

The ROE, developed at the senior commander level, are ultimately limited to a simplified soldier card, taking into account the situational context (the so-called ROE card), which are used by members of the armed forces participating in a given military operation, to the extent necessary to perform the tasks assigned to them.[30] For example, according to the ROE card, when completing a mission, a soldier may be required to use the minimum necessary force only, and if enemy units want to surrender, not to shoot them, but to disarm them and hand them over to the commander; and in the case of the use of weapons, shoot in such a way as not to unnecessarily damage property and cease-fire as soon as the situation allows.[31]

Therefore, on the one hand, ROE serve to simplify the procedures for the implementation of combat tasks and missions, and on the other hand, they require a quick—but often multi-faceted—assessment of the situation in order to properly apply a given rule in specific circumstances, taking into account the political, military and legal context. Such an assessment may be subjective and although ROE are often formulated categorically, it is up to the commander/soldier what decision to make, relying on their knowledge and experience, as well as on intuition.

It should be underlined, however, that ROE are often broader than some of their definitions would indicate. Besides authorizing or restricting the use of force, including delineating the various levels of force applicable in diverse situations, ROE frequently encompass directives on the positioning and posturing of forces, the employment of specific capabilities, the manage-

---

[30] A.J. Carswell, *Converting treaties into tactics on military operations*, IRoRC 2014, Vol. 96, № 895/896, p. 923; M. Szuniewicz, *Aspekty prawne NATO ROE – MC 362/1 (ze szczególnym uwzględnieniem komponentu morskiego)* [*Legal aspects of NATO ROE – MC 362/1 (with particular emphasis on the marine component)*], "Międzynarodowe Prawo Humanitarne" 2012, Vol. III, p. 69.

[31] Several examples of ROE cards are provided in: G.D. Solis, *op. cit.*, pp. 393–396. However, as J.F.R. Boddens Hosang observes, although these cards are often referred to as "ROE cards" (or simply ROE), they do not encompass ROE in the conventional sense. Instead, serving as simplified summaries of the actual ROE, they are intended to provide clear and comprehensible instructions to the personnel they are issued to regarding the use of force within their decision-making scope (*idem*, *Rules of Engagement*…, p. 44).

ment and disposition of captured or detained individuals, and any delegation or withholding of authorities concerning approval of these actions. Ultimately, ROE serve as a pivotal command and control (C2) tool, intended to afford military and political leadership heightened oversight over the conduct of operations, whether combat-related or otherwise, by subordinate forces. As directives guiding operations, ROE fall under the purview of the commander and his or her operations staff element.[32] An absence of appropriate ROE may not only cause a general failure of the operation but may also endanger the deployed forces.[33]

3. Autonomous Weapon Systems—searching for a path in the maze of definitions

In contrast to the inherent nature of war, which remains unchanged as a unique contest of wills between belligerents, and as Clausewitz's "continuation of politics by other means"[34], the methods of conducting warfare are subject to continuous and sometimes very rapid changes. These changes are influenced by a variety of factors, primarily the development of new technologies that can be applied on the battlefield to establish or enhance military superiority over a potential (or actual) adversary.[35] Historically, there has been a general trend in the development of these technologies aimed at distancing danger from one's own soldiers while simultaneously increasing the damage that can be inflicted on the enemy from ever greater distances.[36] This tendency is especially pronounced in contemporary warfare, which is characterized by this very remoteness, giving rise to the term "remote warfare" in the academic literature. Although scholars do not fully agree on the precise scope of this term, some contend that it should be understood liter-

---

[32] G. P. Corn, *Developing Rules of Engagement: Operationalizing Law, Policy, and Military Imperatives at the Strategic Level*, [in:] G.S. Corn, R.E. van Landingham, S.R. Reeves (eds), *U.S. Military Operations: Law, Policy, and Practice*, Oxford University Press, Oxford 2016, p. 212.

[33] M. Faix, *op. cit.*, p. 141. *Cf.* K.B. Sandvik, *Regulating War in the Shadow of Law: Toward a Re-Articulation of ROE*, "Journal of Military Ethics" 2014, Vol. 13, No. 2, p. 118 ("In all circumstances, appropriately drafted and implemented ROE are of central importance to force protection and mission accomplishment.")

[34] *Cf.* G.S. Corn, K. Watkin, J. Williamson, *The Law in War: A Concise Overview, 2nd ed.*, Routledge, London-New York 2023, p. 2; H. Strachan, *The Direction of War: Contemporary Strategy in Historical Perspective*, Cambridge University Press, Cambridge 2013, p. 208.

[35] M.C. Waxman, *Introduction: The Future Law of Armed Conflict*, [in:] M.C. Waxman, T.W. Oakley (eds), *The Future Law of Armed Conflict*, Oxford University Press, Oxford 2022, p. 1.

[36] B. Sajduk, *op. cit.*, p. 18. *Cf.* I.S. Henderson, P. Keane, J. Liddy, *Remote and autonomous warfare systems: precautions in attack and individual accountability*, [in:] J.D. Ohlin (ed.), *Research Handbook on Remote Warfare*, Edward Elgar Publishing, Cheltenham 2017, pp. 335–336 ("From the early firearms that ended the reign of armoured knights through to intercontinental ballistic missiles, humankind has employed great ingenuity in developing means of warfare to attack adversaries remotely. Modern remote weapon systems, including vehicles that can be operated remotely, are but the latest element in this continuum of development.")

ally as the conduct of hostilities from a distance. This is facilitated by three modern categories of means of warfare: remotely controlled vehicles (such as unmanned aerial vehicles—UAVs), cyber weapons, and autonomous weapon systems. Each of these means allows the attacker to inflict losses on the enemy and weaken their military capability, while the weapon operators remain safe, far removed from the theatre of operations.[37]

Among the aforementioned means of warfare, remotely operated systems have been utilized for quite some time, with early instances dating back to the late 19th century. In recent years, there has been a heightened focus on these systems, especially due to the rise in attacks conducted by UAVs, including operations referred to as targeted killings. While in terms of their remote use, UAVs and AWS show some similarities, they are actually distinct from each other. Unlike remotely operated systems, AWS do not require a human operator to be in the loop. Rather, by design they operate without direct human input.[38] However, when it comes to the legal definition of AWS, it proves to be a significant challenge. Despite numerous diplomatic debates, including those led by the CCW member states, there is still a lack of consensus on such a fundamental issue as the development of a common, universally accepted definition for AWS. This does not imply the absence of a definition *per se*. On the contrary, many States and international organizations (both governmental and non-governmental) have adopted their own definitions of AWS, whether in the form of domestic legal acts or in their guidelines, directives and reports.[39]

The first State to regulate its strategy concerning AWS (specifically, guidelines for the development and use of autonomy in weapon systems) and to formulate a definition of this system, thereby providing others with a basis for further considerations on the nature and characteristics of AWS, was the United States. According to the U.S. Department of Defense directive of 2012, AWS is "a weapon that, once activated, can select and engage targets without further intervention by a human operator."[40] This includes "human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation."[41] As regards entities other

[37] J.D. Ohlin, *Remoteness and reciprocal risk*, [in:] J.D. Ohlin (ed.), *op. cit.*, p. 15.

[38] M. Wagner, *Beyond the Drone Debate: Autonomy in Tomorrow's Battlespace*, "ASIL Proceedings" 2012, Vol. 106, № 1, p. 81.

[39] See G. Garcia Vera, *op. cit.*, p. 484.

[40] United States Department of Defense, *Autonomy in Weapon Systems*, Directive No. 3000.09, November 21, 2012, Part II, p. 13.

[41] *Ibidem*, Part. II, p. 13–14.

than States, the International Committee of the Red Cross (ICRC), for example, emphasizes in its definition that AWS may be "any weapon system with autonomy in its critical functions—that is, a weapon system that can select (search for, detect, identify, track or select) and attack (use force against, neutralize, damage or destroy) targets without human intervention."[42] In a similar vein, the European Parliament formulates its definition, indicating that the term LAWS[43] refers to "weapon systems without meaningful human control over the critical functions of selecting and attacking individual targets."[44]

Regarding the definition of AWS/LAWS, China presents an interesting approach. In a position paper submitted during the meeting of the GGE in 2018, the Chinese delegation primarily emphasized that LAWS should be understood as fully autonomous lethal weapon systems, and it explained that LAWS should "include but not be limited to the following 5 basic characteristics:

1. the first is lethality, which means sufficient pay load (charge) and for the means to be lethal;
2. the second is autonomy, which means the absence of human intervention and control during the entire process of executing a task;
3. thirdly, impossibility for termination, meaning that once started there is no way to terminate the device;
4. fourthly, indiscriminate effect, meaning that the device will execute the task of killing and maiming regardless of conditions, scenarios and targets;
5. fifthly evolution, meaning that through interaction with the environment, the device can learn autonomously, and expand its functions and capabilities in a way exceeding human expectations."[45]

Certain elements from the above list can be found in the definition formulated by France, which defines LAWS as "a lethal weapon system programmed to be capable of changing its own rules of operation particularly as regards target engagement, beyond a determined framework of use, and ca-

---

[42] International Committee of the Red Cross, *Views of the ICRC on autonomous weapon systems*, paper submitted to the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems, April 11, 2016, p. 1.

[43] It should be explained that the difference between AWS and LAWS, as noted in the doctrine, is that lethal autonomous weapon systems are anti-personnel AWS, i.e. they may target human beings (M. Pacholska, *Autonomous Weapons*, ASSER Research Paper 2023-03, p. 10). Nevertheless, for the sake of simplicity, both terms (i.e., AWS and LAWS) are used interchangeably in this study.

[44] European Parliament resolution of September 12, 2018 on autonomous weapon systems, (2018/2752(RSP)), Official Journal of the European Union C 433/86, 12/23/2019, para. B.

[45] *Position Paper Submitted by China to Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons*, April 11, 2018, CCW/GGE.1/2018/WP.7, p. 1.

pable of computing decisions to perform actions without any assessment of the situation by human military command."[46] It should be added that France also limits the understanding of LAWS to fully autonomous weapons, thereby presenting "a restrictive approach avoiding those weapons being confused with remotely operated or supervised weapons systems, which always involve a human operator."[47]

Given the thematic scope of this article, it is essential to include the definition formulated in the second edition of the *Sanremo ROE Handbook*, published in 2022. The so-called *Newport ROE Handbook* includes a glossary of terms in which a definition of AWS appears, absent in the 2009 edition of the handbook. According to this definition, an autonomous weapon system is "a weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but which can select and engage targets without further human input after activation."[48] The handbook also defines the mentioned "human-supervised autonomous weapon system", which is "an autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagement."[49] Interestingly, the authors of the handbook also decided to define what a semi-autonomous weapon system is, stating that it is "a weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator."[50]

These are, of course, just examples of definitions of AWS/LAWS, as there are now numerous such definitions and they continue to expand[51]. However, this brief overview is sufficient to—following the example of the Chinese proposal—highlight certain recurring elements in AWS definitions that are crucial for our further analysis. Specifically, the aim is to identify the characteristic features of AWS that will help determine whether and to what extent such systems will be able to cope with ROE adopted for military operations, and what kind of changes will be required to adapt the situation-oriented

[46] J.-B. Jeangène Vilmer, *A French Opinion on the Ethics of Autonomous Weapons*, June 2, 2021, https://warontherocks.com/2021/06/the-french-defense-ethics-committees-opinion-on-autonomous-weapons/ (accessed: 9/07/2023).

[47] *Ibidem*.

[48] D. Mandsager *et al.* (Drafting Team), *op. cit.*, Annex D, p. 103.

[49] *Ibidem*, Annex D, p. 105.

[50] *Ibidem*, Annex D, p. 108.

[51] An interesting and comprehensive analysis of these definitions can be found in: M. Taddeo, A. Blanchard, *A Comparative Analysis of the Definitions of Autonomous Weapons Systems*, "Science and Engineering Ethics" 2022, Vol. 28(5), № 37.

ROE to automated decision-making processes on which both existing and potential future-machine-learning based systems rely on. In the author's opinion, to answer these questions, first, it is necessary to refer to the concepts and phrases the explanation of which is indispensable in order to understand the analyzed problem. Thus, it is necessary to address the concept of autonomy and autonomous functioning, as well as the concepts of artificial intelligence (AI) and algorithms.

3.1. Autonomous, automated, automatic—what is the difference?

There are already a number of automated processes or types of weapons on the modern battlefield. This is due to the fact that in today's world, the problem is usually not a lack of information, but the excess of it. In connection with the above, we often resort to instruments designed to screen information to the extent that we can make the most reasonable decision. The introduced algorithms prevent some activities from being undertaken by humans. An example is the automatic pilot in airplanes or ships. The challenge arises when a system or algorithm decides to apply lethal force.[52] It is true that some modern weapon systems already automate aspects of attack decision-making. The data from sensors can be fused with a view to presenting the pilot with a coherent picture of what is taking place on the battlefield, thereby simplifying the attack decision-making process.[53] There are already so-called automated systems, such as the American *Phalanx* system or the Russian *Kashtan* system, installed on navy ships that independently locate and destroy incoming rockets or missiles. This means that a person programs such a device and precisely defines the target, e.g. an incoming missile, and the device's response. However, such devices do not take any action other than what is programmed.[54]

According to M. Wagner, "the different types of unmanned systems can be usefully grouped into three different categories […]: remotely operated systems, automated systems, and systems that operate autonomously. The distinctions among the categories serve an important purpose, namely to separate the existing weapon systems—which are either automated or remotely operated—from those future systems that will function in an autono-

[52] P. Łubiński, *op. cit.*, p. 178.
[53] W. Boothby, *New Technologies in Means and Methods of Warfare*, [in:] J. Wouters, Ph. De Man, N. Verlinden (eds), *Armed Conflicts and the Law*, Intersentia, Cambridge 2016, p. 345.
[54] P. Łubiński, *op. cit.*, p. 179.

mous manner."[55] Therefore, the concept of "autonomous" should not be confused with the term "automated" or "automatic", and in particular should these terms not be used interchangeably. Generally speaking, future actions are foreseeable in automatic systems due to the inability to react differently to a given factor, in automated systems the ability to operate is enhanced but at the same time limited to the types of situations that have been pre-programmed in a limited way, while in the case of autonomous systems, the effect of actions is difficult to predict since the system is given a high level of decision-making in a changing and complex environment.[56] Until recently, there was still a division into different levels of autonomy, but this division turned out to be impractical and focused excessively on machines and not on the relationship between the machine and a human being.[57] In general, autonomy denotes a system capable of operating independently for some period and making its own decisions without direct human supervision.[58] It can therefore be assumed that, from a technical point of view, autonomous systems are "systems that can operate without direct human control or supervision in dynamic, unstructured, open environments based on feedback information from a variety of sensors."[59] Thus, an autonomous weapon system is best understood as "being composed of disparate soft- and hardware elements that work together – including sensors, algorithmic targeting and decision-making mechanisms, and the weapon itself."[60]

According to T. Vestner and A. Rossi, autonomy in weapon systems can be categorized according to three different traits, namely:

1. the human-machine command-and-control relationship;
2. the sophistication of the machine's decision-making process; and
3. the types of decisions or functions being made autonomous.[61]

As regards to the first trait, systems can be classified based on whether they receive inputs by a human operator to perform their functions in:

1. "human-in-the-loop" of the targeting decision, referring to systems that select targets and deliver force upon human command;

---

[55] M. Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems*, "Vanderbilt Journal of Transnational Law" 2014, Vol. 47, № 5, pp. 1379–1380.

[56] K. Kowalczewska, *op. cit.*, p. 33; L.A. Letendre, *Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers*, "International Law Studies" 2020, Vol. 96, p. 278.

[57] See T. McFarland, *Factors Shaping the Legal Implications of Increasingly Autonomous Military Systems*, IRoRC 2016, Vol. 97, issue 900, pp. 1319–1321.

[58] L.A. Letendre, *op. cit.*, p. 278.

[59] Article 36, *Structuring debate on autonomous weapons systems. Memorandum for delegates to the Convention on Certain Conventional Weapons (CCW)*, Briefing Paper, November 2013, p. 1.

[60] *Ibidem*.

[61] T. Vestner, A. Rossi, *Legal Reviews of War Algorithms*, "International Law Studies" 2021, Vol. 97, p. 529.

2. "human-on-the-loop," capable of selecting targets and delivering force without human interaction but remaining under the oversight of humans so that humans retain the power to override the machine's action; and
3. "human-out-of-the-loop," which refers to systems that select targets and deliver force autonomously without humans being able to intervene during the process.[62]

The major difference between manned systems ("human-in-the-loop") and autonomous systems ("human-on-the-loop" and "human-out-of-the-loop") concerns the decision-making process in the targeting cycle, i.e. an overall process involving the making of decisions preceding or related to the moment of the use of force.[63] Through the targeting process, "armed forces seek to match potential targets with appropriate lethal and nonlethal actions, the ultimate aim being to achieve specific and measurable *effects* that advance the commander's objectives."[64] In the most extreme case, it is the system itself that makes targeting decisions based on its observations, perceptions, and evaluations rather than human operators. Accordingly, AI-driven weapon systems that conduct targeting decisions need to apply and comply with the entire spectrum of so-called targeting law, namely, those rules under the law of armed conflict (LOAC) that determine who and what can be targeted in an armed conflict and how.[65]

Thus, the lack of "human control/human intervention" is a crucial element of the definitions of AWS. In fact, many semi-autonomous weapons in use today rely on autonomy for certain parts of their system but have a communication link to a human that will approve or make decisions. In contrast, a fully autonomous system could be deployed without any established communication network and would independently respond to a changing envi-

---

[62] *Ibidem*, pp. 529–530. *Cf.* V. Boulanin, M. Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, SIPRI, November 2017, pp. 5–7.

[63] M. Ekelhof, *Human control in the targeting process*, [in:] ICRC Report, *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*, Expert Meeting, Versoix, Switzerland, March 15–16, 2016, p. 54. As M. Wagner notes, "[t]here are at least two characteristics that define the notion of autonomy in the specific context of AWS. First is the ability to operate independently and engage targets without being programmed to specifically target an individual object or person. This includes the capability to react to a changing set of circumstances, and requires that the rules of IHL be 'translated' into code. The second, interrelated, aspect is the capability to make discretionary decisions. The actions of AWS are therefore, in contradistinction to automated systems, predictable only within the range that they were programmed" (M. Wagner, *The Dehumanization of IHL…*, p. 1383).

[64] M.N. Schmitt, J. Biller, S.C. Fahey, D.S. Goddart, Ch. Highfill, *Joint and Combined Targeting: Structure and Process*, [in:] J. Ohlin, L. May, C. Finkelstein (eds), *Weighing Lives in War*, Oxford University Press, Oxford 2017, p. 1.

[65] T. Vestner, A. Rossi, *op. cit.*, p. 530.

ronment and decide how to achieve its pre-programmed goals. Nowadays, autonomy is present in many military applications that do not raise concerns such as take-off, landing, and refueling of aircraft, ground collision avoidance systems, bomb disposal and missile defense systems. However, advances in autonomy have the potential to move a human soldier further and further out of the control loop and to leave more and more decisions up to the machine. The problem is that ROE include, *inter alia*, principles and norms of LOAC, and are created with the intention of their application and implementation by humans, not machines. Translating those legal and other requirements of ROE cannot be an afterthought in the development of autonomous weapon systems—it must be built-in from the beginning.[66]

There is another aspect of the issue concerning autonomy in weapon systems that should be mentioned here. According to the report prepared within the framework of the Multinational Capability Development Campaign 2013–2014 (MCDC)—a collaborative program between 19 nations, NATO and the European Union—"autonomy" literally means "self-governing". In the context of defense systems, the term is typically used to describe how machines perform certain functions—to varying extents—independently of human control. However, machines are not autonomous in a literal sense but may exhibit "autonomous-like" functions, relative to a particular level of human control and situational context. Labeling a function as autonomous implies a certain level of ability to adapt to complex or unanticipated situations, which is different from an automatic function occurring according to defined inputs, rule sets and outputs.[67] It is worth adding that the common feature of currently used machines is that "autonomous" robots operate in strictly limited environments, and their level of autonomy can be described as low (rather, they should be referred to as "automated"). In addition, these robots do not interact directly with people other than their operators, and even if they do, such robots cannot distinguish whether the obstacle they encounter is a human or an object.[68] Therefore, it is better to use the term "autonomous functioning" instead of the term "autonomy", which refers to the ability of a system, platform, or software, to complete a task without human intervention, using behaviors resulting from the interaction of computer programming with the external environment. Tasks or functions executed either by a platform, or distributed between a platform

[66] See L. A. Letendre, *op. cit.*, p. 275.
[67] Multinational Capability Development Campaign (MCDC) 2013-2014, *Policy Guidance: Autonomy in Defence Systems*, October 29, 2014, p. 8.
[68] K. Kowalczewska, *op. cit.*, p. 38.

and other parts of the system, may be performed using a variety of behaviors, which may include reasoning and problem-solving, adaptation to unexpected situations, self-direction, and learning.[69]

3.2. Artificial intelligence and algorithm as essential elements of autonomous functioning

Understanding of "autonomous functioning" depends, among others, on related concepts such as artificial intelligence (AI) and decision-making algorithms. However, there is no universal legal definition of AI, and even in academia, there is no standard definition of AI (although references to human intelligence are most common). Various definitions are particularly presented in international organizations (e.g., the European Union, OECD, or UNESCO).[70] For example, in a joint report of the Innovation Centre of the International Criminal Police Organization (INTERPOL) and the Centre for AI and Robotics of the United Nations Interregional Crime and Justice Research Institute (UNICRI), published in 2019, the term "Artificial Intelligence" is defined as "an intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals."[71] This term is "applied particularly when a machine mimics *cognitive* functions that are associated with human minds, such as *learning* and *problem solving*."[72] A definition of an AI system is proposed by the OECD Council in the Recommendation on Artificial Intelligence dated May 22, 2019, stating that it is "a machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."[73] The AI system is defined similarly in the EU Regulation 2024/1689 of the European Parliament and of the Council—this

---

[69] R. Arnold, *The legal implications of the use of systems with autonomous capabilities in military operations*, [in:] A.P. Williams, P.D. Scharre (eds), *Autonomous Systems: Issues for Defence Policymakers*, NATO, The Hague 2015, p. 84–85; MCDC, *Policy Guidance…*, p. 9. Indeed, the term "autonomous weapons" is a commonly adopted shorthand for weapon systems with autonomous functionalities (M. Pacholska, *op. cit.*, p. 1). *Cf.* T. McFarland, *Minimum Levels of Human Intervention in Autonomous Attacks*, "Journal of Conflict & Security Law" 2022, Vol. 27, № 3, p. 389.

[70] See M. Świerczyński, Z. Więckowski, *Sztuczna inteligencja w prawie międzynarodowym. Rekomendacje wybranych rozwiązań* [*Artificial Intelligence in International Law. Recommendations of Selected Solutions*], Wydawnictwo Difin, Warszawa 2021, pp. 35–39.

[71] *Artificial Intelligence and Robotics for Law Enforcement*, Report prepared by the United Nations Interregional Crime and Justice Research Institute and the International Criminal Police Organization, Torino-Lyon 2019, Annex 1, p. 27.

[72] *Ibidem*.

[73] OECD Legal Instruments, *Recommendation of the Council on Artificial Intelligence*, May 22, 2019, OECD/LEGAL/0449, para. I.

document emphasized that a key characteristic of AI systems is "their capability to infer."[74] This capability "refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms, or both, from inputs or data. The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved."[75] The regulation also clarifies that AI systems are designed to "operate with varying levels of autonomy, meaning that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention. The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use."[76]

Despite some conceptual similarities, there is, admittedly, no universally accepted definition of "artificial intelligence", and numerous utility definitions are divided into those that focus on the ability to map rational human reasoning/thinking and those that focus on simulating rational human behavior.[77] However, certain main characteristics of AI systems can be identified, which are:

1. perception of the environment, including consideration of the complexity of the real world;
2. information processing, encompassing the collection and interpretation of input data;
3. action-taking, which involves task execution (including adaptation and response to environmental changes) while maintaining a certain level of autonomy;
4. achievement of specific goals.[78]

For the purposes of our considerations, it is sufficient to assume that AI is "a set of computational techniques that enables machines to solve complex

[74] *Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, OJ L, 2024/1689, 07/12/2024, para. 12.
[75] *Ibidem*.
[76] *Ibidem*.
[77] Ch. Bartneck, Ch. Lütge, A. Wagner, S. Welsh, *An Introduction to Ethics in Robotics and AI*, Springer, Cham 2021, pp. 5–11; J.N. Kok, E.J. W. Boers, W.A. Kosters, P. van der Putten, *Artificial Intelligence: Definition, Trends, Techniques, and Cases*, Encyclopedia of Life Support Systems 2009, http://www.eolss.net/Eolss-sampleAllChapter.aspx, pp. 1–3 (accessed: 6/20/2023).
[78] M. Świerczyński, Z. Więckowski, *op. cit.*, p. 40.

and abstract problems that are naturally performed through human intelligence."[79] AI can provide a system with "cognitive capabilities" to independently undertake functions such as observation or learning. Autonomy is therefore a constitutive trait of AI[80].

According to existing research, AI is typically grouped into the following three categories:

1. Artificial Narrow Intelligence (ANI)—this is an AI that can execute one particular decision type; ANI is also described as Weak AI; currently, all AI methods we use are examples of ANI;
2. Artificial General Intelligence (AGI)—this is an AI that can execute multiple functions and is comparable to human intellectual abilities; AGI is also known as Strong AI; no AI system has yet achieved general intelligence;
3. Artificial Super Intelligence (ASI)—AI that surpasses human intelligence; ASI is also categorized as Strong AI and also does not exist yet.[81]

The distinction between Strong AI and Weak AI essentially boils down to the difference between being intelligent and acting intelligently. Strong AI denotes a "mind" that is truly intelligent and self-aware. Weak AI refers to what is currently available, namely systems that exhibit intelligent behaviors despite being "ordinary" computers.[82]

Taking the above aspects into account and considering the military application of AI, it is also worthwhile to mention the definition of AI included in the NATO Science & Technology Organization report of 2020. In light of this definition, AI refers to "the ability of machines to perform tasks that normally require human intelligence—for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action—whether digitally or as the smart software behind autonomous physical systems."[83] The authors of the report emphasize that "AI has the potential for revolutionary impact on NATO operations and capabilities"[84] and "integration of AI into combat models & simulation, enterprise systems, de-

[79] T. Vestner, A. Rossi, *op. cit.*, p. 513.
[80] *Ibidem*, pp. 513–514.
[81] Ch. Bartneck *et al.*, *op. cit.*, p. 10; M. Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*, Penguin Random House, New York 2017, p. 55.
[82] M. Świerczyński, Z. Więckowski, *op. cit.*, p. 44–45.
[83] NATO Science & Technology Organization, *Science & Technology Trends 2020-2040: Exploring the S&T Edge*, Brussels 2020, p. 50.
[84] *Ibidem*, p. 14.

cision support systems, cyber defence systems and autonomous vehicles will allow for rapid and more effective human-machine decision making."[85]

The second of the above-mentioned elements of autonomous functioning —the algorithm— is one of the key concepts to understand how AI systems work. From a technical point of view, it can be defined as a precisely specified computational procedure that, based on certain values provided (input), produces certain values (referred to as output). Algorithms allow, among others, to make a "decision" or "choice" based on programmed parameters and specific inputs, and therefore replace human judgment or decision-making processes.[86] It should be noted, however, that typical AI algorithms that are data-driven "are inherently brittle, which means that such algorithms cannot generalize and can only consider the quantifiable variables identified early on in the design stages when the algorithms are originally coded."[87] Furthermore, until the level at which ideally autonomous systems are created—i.e. those created by other autonomous systems—is achieved, the algorithm will always be the work of a human, and therefore its functioning will be subject to some extent to human control.[88]

The concept of an algorithm is, of course, a general one, referring broadly to the functioning of AI systems. However, researchers from the Harvard Law School, within the framework of the Program on International Law and Armed Conflict, sought to "focus" this phenomenon and introduce a new concept—war algorithms—that elevates algorithmically-derived "choices" and "decisions" to a central concern regarding technical autonomy in war. The result of their research is the report "War-Algorithm Accountability", published in 2016, in which the authors limit the types of algorithms to those that fulfill three conditions:

1. they are expressed in computer code;
2. they are effectuated through a constructed system; and
3. they are capable of operating in relation to armed conflict.[89]

Thus, the authors of the report focus on capability, or an effect-based approach, presenting a broader perspective than the concept of intended use, or a design-based approach.[90] According to the authors of the report, the

[85] *Ibidem*, p. 15.
[86] K. Kowalczewska, *op. cit.*, p. 31.
[87] M.L. Cummings, *Artificial Intelligence and the Future of Warfare*, International Security Department and US and the Americas Programme Research Paper, January 2017, p. 7.
[88] K. Kowalczewska, *op. cit.*, p. 31; T. McFarland, *op. cit.*, pp. 1327–1328.
[89] D.A. Lewis, G. Blum, N.K. Modirzadeh, *War-Algorithm Accountability*, Harvard Law School Program on International Law & Armed Conflict (PILAC), August 2016, p. 16.
[90] See K. Kowalczewska, *op. cit.*, p. 31.

definition of war algorithms, relevant to armed conflict, should be based on their capability to be used in such contexts, even if not originally designed for it. This inclusive approach encompasses all potentially adaptable algorithms, rather than just those intended for warfare. The rationale is that modern technology is often modular and adaptable for military use regardless of initial intent. Additionally, focusing on capability rather than intent broadens accountability to include not only those who deploy or operate these algorithms but also those involved in their design and development across various sectors. This ensures that a wide range of actors can be held responsible for the use of war-related algorithms.[91] Although the concept of a war algorithm has not yet received widespread attention, it can be helpful in understanding the complexity of new algorithm-based technologies. This understanding can consequently allow for a conscious assessment of the legality of the means of warfare that use algorithms in so-called critical functions, such as targeting, tracking, selecting, and attacking a target.[92]

In summary, considering the aforementioned aspects of the functioning of an autonomous system and deeming them significant for further consideration within the scope of the subject research, one can attempt to formulate a simplified definition of the "autonomous weapon system" that includes the most essential components. Thus, it can be concisely stated that if the machine is intended for use on the battlefield and is driven by artificial intelligence, which allows it to function autonomously, using an algorithm, it can be defined as the widely used term "autonomous weapon system" (AWS).

4. Can AWS comply with ROE?

As already indicated above, ROE are written as a series of prohibitions and permissions applicable to activities in a wide range of military operations.[93] They are developed separately for each military operation and are therefore tailored to the specifics of each of them.[94] Although certain standards are usually developed, *inter alia*, through the experience gained during

[91] D.A. Lewis *et al.*, *op. cit.*, pp. 16–17.

[92] K. Kowalczewska, *op. cit.*, p. 32. As T. Vestner and A. Rossi rightly notice, "developers need to introduce targeting law into the AI system by translating law into standards that an algorithm can understand […]. The main challenges are [however] the translation of the nature of the law into the algorithm's language and the associated risks of unintended consequences" (T. Vestner, A. Rossi, *op. cit.*, pp. 532–533). Questions would include, for instance, whether the data provided are sufficient for the algorithm to distinguish between a person who is *hors de combat*, a combatant, or a civilian directly participating in hostilities.

[93] North Atlantic Military Committee, *NATO Rules of Engagement*, Military Decision, MC 362/1, June 30, 2003, Part V, para. 15.

[94] D.R. Bugajski, *Morskie, stałe zasady podejmowania działań przy użyciu siły* [*Maritime Standing Rules of Engagement*], "Międzynarodowe Prawo Humanitarne" 2012, Vol. III, p. 40.

previous operations, each time the preparation of ROE requires taking into account both the tasks that will be carried out by the armed forces and the environment in which these forces will operate. It should be emphasized that both factors, i.e. tasks and the environment, may determine which rules will apply during a given operation. For example, during military operations conducted on land, the tasks of the armed forces may include not only armed combat but also search and rescue (SAR) operations. During a maritime operation, the activities of warships may include the inspection, seizure, and destruction of property. When assessing ROE from a task perspective only, the armed forces may be involved in a peace operation during which their main task will be the evacuation of non-combatants or humanitarian assistance, so the use of force will be a last resort and the escalation of force (EOF) procedures will apply in this respect.

As for the targeting process itself, consisting of selecting and prioritizing targets and matching the appropriate response to them,[95] it requires taking into account operational requirements and capabilities, applicable ROE, and LOAC. In particular, if the ROE allow for the use of lethal force, the so-called law of targeting must be considered, which is based on the principles and norms of LOAC regarding the conduct of hostilities. Specifically, this refers to the fundamental principles of IHL—the principle of distinction and the principle of proportionality. These are supplemented by detailed obligations concerning the exercise of precautions during attacks and against the effects of attacks, as well as equally detailed regulations regarding the means and methods of warfare. The relationship between ROE and targeting may be summarized in three guidelines: first, forces may target only those military objectives permitted to be targeted in the relevant ROE; second, ROE may impose policy restrictions on targeting that go beyond the requirements of the LOAC; third, ROE must never permit targeting that is not consistent with LOAC. Targeting directives for a mission may set out limitations such as restricted target lists and no-strike lists. Additionally, commanders may be restricted from taking certain actions by their ROE.[96]

All these aspects—the level of complexity of the environment, the task at hand, and broadly understood mission goals to achieve—may influence the level of difficulty of making effective decisions while implementing ROE[97]

[95] More on this topic – see: North Atlantic Treaty Organization, *AJP-3.9 Allied Joint Doctrine for Joint Targeting*, Edition B, version 1, November 2021, Chapter 1, paras. 1.1–1.7. See also M.N. Schmitt, *International Humanitarian Law and the Conduct of Hostilities*, [in:] B. Saul, D. Akande (eds), *The Oxford Guide to International Humanitarian Law*, Oxford University Press, Oxford 2020, p. 147–174.

[96] A. Cole *et al.* (Drafting Team), *op. cit.*, Appendix 6 to Annex A, paras. 6.2–6.3.

[97] Department of Defence, *Research & Engineering…*, p. 2.

and may present an extremely demanding challenge for the use of ROE-driven AWS on the battlefield. Obviously, some of the tasks already carried out by AI systems are related to the decision-making process within ROE—they are used to a great effect for intelligence, surveillance and reconnaissance (ISR), and taking over functions such as verification or targeting.[98] However, such systems would need to comply with those rules of targeting law that are relevant to the functions they are entrusted with. For instance, ISR would need to be assessed in light of the principle of distinction. It would also need to be checked to determine if it can properly report that a person is *hors de combat*. While the system would not need to be able to conduct a proportionality assessment itself, it would need to be able to provide the relevant and correct information for assessing the proportionality of an attack and deciding on feasible precautionary measures.[99]

Despite the significance of the discussed issue concerning the requirement to develop appropriate ROE for the use of AWS in military operations, this topic has yet to garner widespread interest among States, international organizations, or scholars. This lack of attention is somewhat understandable given the ongoing debate about whether AWS can adhere to the principles and norms of IHL and the complexities surrounding legal responsibility for the use of these machines in combat. Indeed, since ROE must comply with existing laws, particularly IHL/LOAC[100], finding an appropriate formula or model that enables AWS to adhere to these laws will be a significant step toward developing ROE for AWS.[101] However, it should be underlined that ROE are the result of three "groups of interests"—political, military, and legal. In essence, ROE should take into account the political objectives of the military operation, consider the strategy and tactics of the armed forces employed, and maintain compliance with relevant principles and norms of international and national law. Theoretically speaking, creating an algorithm

---

[98] J. Kwik, T. van Engers, *op. cit.*, p. 1. More on the use of so-called Artificial Intelligence Decision Support Systems (AI DSS) in military decision-making – see: ICRC and Geneva Academy, *Expert Consultation Report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts*, Geneva, March 2024.

[99] T. Vestner, A. Rossi, *op. cit.*, p. 531.

[100] See, e.g., D. Mandsager *et al.* (Drafting Team), *op. cit.*, Part III, para. 16.

[101] It is worth mentioning that currently some efforts are being made to incorporate the fundamental principles of IHL into the structure of military AI. Such an attempt was made by T. Żurek, J. Kwik and T. van Engers, who based their model on the military targeting cycle, integrating operational and legal considerations to ensure practical usability and legal compliance. This alignment ensures that the hypothetical AI can effectively replace a human-controlled system, addressing objectives, desired effects, military advantage, and humanitarian concerns like incidental harm/collateral damage (see T. Zurek, J. Kwik, T. van Engers, *Model of a military autonomous device following International Humanitarian Law*, "Ethics and Information Technology" 2023, Vol. 25, № 15).

that considers all three groups of interests would be significantly more challenging than merely incorporating legal principles and norms.

To date, J.F.R. Boddens Hosang is probably the only researcher who has addressed the issue of the effects of weaponized autonomous technology on the traditional ROE concept and doctrine. According to Boddens Hosang, two of the elements of ROE doctrine are particularly relevant in the context of autonomous weapons: "the degree of system-level specificity of the ROE […] and the level to which the ROE should be disseminated."[102] As regards the first element, ROE can authorize or regulate the use of specific weapons (so-called weapon-specific ROE), defining the circumstances and purposes for their deployment in military operations. For automatic weapon systems, due to their predictable responses based on pre-programmed parameters, the ROE can be straightforward, specifying when and where these systems may be activated and whether weapon release authority should reside with the firing unit or a higher command level. In contrast, fully autonomous weapons are not confined to predictable uses or specific target types, making it challenging to regulate them with traditional ROE. Limiting the activation of AWS to specific times and places is analogous to controlling a soldier with a rifle only by deployment parameters, leaving all other decisions to the soldier. This approach may be feasible in operations during extremely large-scale armed conflicts but is not a realistic or viable option in limited military operations (with clearly defined goals). Given the nature of autonomous weapons, it is debatable whether they should be governed by weapon-specific ROE or treated as machine counterparts to human operators. This implies that rather than regulating the machines with traditional ROE, the systems themselves should be capable of interpreting and applying the ROE. This shift does not alter the fundamental doctrine of ROE but necessitates a redesign in how they are formulated and implemented. Drafting ROE for autonomous systems involves leveraging technical capabilities and operational experience, similar to automatic systems and loitering munitions. However, true AWS require ROE to be written in a way that the systems can independently interpret and apply them. This approach not only complicates the drafting process but also demands software capable of integrating legal constraints with specific operational requirements. Consequently, there would likely be two sets of ROE: one for human operators adhering to "traditional"

[102] J.F.R. Boddens Hosang, *Control Through ROE…*, p. 406.

doctrines, and another tailored for electronic "operators", accommodating the unique demands of autonomous systems.[103]

Regarding the level of dissemination of ROE, in case of a future fully AWS the interaction with ROE can follow two models: either the system is deployed with ROE guidance directed at the human decision-maker, who considers the system's capabilities and limitations, or the system is designed to independently interpret and apply the ROE. The first model presents inherent uncertainties and risks, suggesting that decisions to deploy such systems would likely not be made at the tactical level and might require higher command approval, consistent with current doctrine. This approach is improbable except in extreme armed conflict scenarios. The second model, where the ROE are integrated into the autonomous system itself, appears more viable (apart from questions as to whether that is achievable from a technical point of view). This model would transform the autonomous systems into operators that need ROE dissemination at the tactical level, similar to combat pilots. Thus, increasing the autonomy of weapon systems necessitates a corresponding dissemination of ROE at lower command levels to ensure appropriate use of force.[104]

However, the issue of developing appropriate ROE that can be integrated with autonomous systems to ensure these systems are capable of correctly interpreting and applying the ROE remains unresolved. Neither the ROE formulated by States and international organizations nor model manuals on these principles provide clear answers. An exception is the updated 2022 ROE Manual, which in Annex B (the Compendium of ROE) contains the specific rules to be used for drafting ROE. It classifies AWS under the rules pertaining to the use of weapon systems, including specially regulated weapons systems. However, these are merely general guidelines intended as a framework for more detailed and concrete rules—indeed, according to Series 84, the use of an autonomous weapon system is prohibited unless authorized by a specific rule.[105]

---

[103] *Ibidem*, pp. 406–408.

[104] *Ibidem*, pp. 408–409.

[105] D. Mandsager *et al.* (Drafting Team), *op. cit.*, Annex B, p. 68. It should be underlined, however, that ROE are not used to provide instructions on how to operate a specific weapon or weapon system. In other words, ROE can authorize or regulate the use of specific weapon systems, such as crew-served weapons or riot control agents, by defining the circumstances or purposes for their deployment. However, these rules do not provide technical instructions for the actual use of those weapons. Such specifics are left to commanders to address through other means, such as directives (J.F.R. Boddens Hosang, *Control Through ROE…*, pp. 401–402).

Even if it were possible to develop special ROE for AWS, overcoming the technical challenges and enabling the decision to use such weapon systems at the operational level, and even if specific "algorithmic" ROE cards were created for these systems to be used at the tactical level, there would still be no certainty that AWS, at their current level of technological development, could properly interpret and apply the ROE in battlefield conditions. It should be noted that the main purpose of ROE is to "control actions and behaviour which (directly) relate to or influence the behaviour of (potential) hostile forces and thereby (attempt to) maintain control over, or influence, the overall conduct of the parties and the use of force in the theatre of operations."[106] However, these are not tactical rules—ROE do not instruct soldiers on how to execute missions, although tactics and ROE undoubtedly complement each other.[107] They inform commanders about "imposed restrictions" and the "degree of freedom" they have during mission execution.[108] In other words, ROE do not normally dictate how a result is to be achieved but will indicate what measures may be unacceptable. This means that both commanders and soldiers have a certain degree of discretion in assessing and deciding whether the use of force is permissible under given circumstances.[109] Obviously, guidelines regarding the use of force must be appropriately tailored to the circumstances and operational environment, but due to the uncertainty and unpredictability of combat situations, targeting remains, in fact, a dynamic process characterized by situation-specific decision-making.[110] It appears, therefore, that certain restrictions that could hinder the application of ROE in the case of AWS include the environment in which the military operation takes place and the tasks that must be accomplished to achieve the objectives of the operation.

4.1. The first restriction—the environment

Taking into account the universal application of ROE by modern armed forces, these rules are being issued and implemented in a wide spectrum of

[106] J.F.R. Boddens Hosang, *Rules of Engagement…*, p. 21.

[107] See G.D. Solis, *op. cit.*, p. 373 ("ROE are designed to provide boundaries and guidance on the use of force that are neither tactical control measures nor substitutes for the exercise of the commander's military judgment.") *Cf.* A. Cole *et al.* (Drafting Team), *op. cit.*, para. 3 ("ROE are not used to assign missions or tasks nor are they used to give tactical instructions.")

[108] B.F. Klappe, *op. cit.*, para. 1320.

[109] For example, in case of weapon use, a soldier must remember to aim when firing, to use only as much ammunition as necessary, to employ weapons in a manner that avoids unnecessary destruction of property, and to cease fire as soon as the situation allows (based on the ROE card of the PKW KFOR – on file with author).

[110] *Cf.* ICRC and Geneva Academy, *Expert Consultation Report…*, p. 12.

operations, comprising not only cases of international armed conflict but also non-international armed conflicts and an entire spectrum of other types of operations involving the potential use of force.[111] In fact, the type of operation has a decisive impact on the scope and content of the ROE. In the case of operations conducted during peacetime (e.g., humanitarian operations) or in the period preceding the outbreak or escalation of armed conflict, force can only be used within the authorization provided by ROE (while observing two important principles—restraint and legitimacy), and the only exception to this rule is the use of force in self-defense.[112] In situations of armed conflict, ROE serve to specify and clarify the norms of the LOAC, typically used to limit the lawful use of force to achieve operational or political objectives. Special ROE are also prepared depending on the geographical environment in which the military operation is to be conducted—whether it is on land, at sea, or in the air. In each case, the ROE must take into account the specific characteristics of the environment, including applicable norms of IHL.

In the context of using AWS, since AI systems operate and learn based on the data they receive, data is central to both the development of the system and the determination of its operational behavior. Belligerents must ensure that the system is trained with appropriate data. If the system autonomously collects data and learns, it must be programmed to collect only relevant data and use it appropriately. Therefore, the system must be trained with a focus on the environment in which it will be deployed and under conditions representative of that environment.[113] However, limiting the discussion solely to issues related to conducting hostilities, it should be noticed that the decision to use lethal force is made differently in international armed conflict when the AWS operates over the area occupied by regular armed forces (some autonomy in searching for the target is then allowed), and differently in a situation of an internal conflict, when the AWS independently searches for targets in an urban area where insurgents take refuge among the civilian population.[114] Indeed, complex battlefields, particularly urban environments, are characterized by numerous civilian objects, agents, and dense infrastructures that must be understood and navigated. In urban areas, conducting military operations in a manner that effectively protects civilians is significantly more challenging due to the presence of civilians and the interconnectedness of

[111] M. Faix, *op. cit.*, p. 136.
[112] See North Atlantic Military Committee, *NATO ROE*, Part II, para. 9, Part III, paras. 10–12.
[113] T. Vestner, A. Rossi, *op. cit.*, p. 533.
[114] P. Łubiński, *op. cit.*, p. 185.

military objectives and civilian objects, including critical infrastructure.[115] Furthermore, considering the global trend towards urbanization, it appears that future conflicts are likely to increasingly occur in densely populated battlefields, which are conducive to high levels of uncertainty.[116] The uncertainty and dynamic nature of these environments require systems to make accurate predictions and timely actions. Modern battlefields present highly adversarial multi-agent environments, necessitating engagement with cooperative, neutral, and competitive agents simultaneously, including allied agents, civilians, third parties, and enemy agents. While air and maritime environments are relatively simple with few navigational obstacles,[117] land navigation presents numerous challenges such as irregular terrain, vegetation, unstable surfaces, and a higher number of agents (both human and automated). Due to these difficulties, current autonomous capabilities are typically deployed only in "controlled" environments.[118] Therefore, it is reasonable to agree with the opinion of V. Boulanin and M. Verbruggen, who aptly observed that "[a]dvances in machine perception are key to the progress of machine autonomy. In many respects, it is the limitation of perceptual intelligence that is today the most important obstacle to the development and use of robotic technologies outside simple, predictable or well-controlled environments."[119]

4.2. The second restriction—tasks

Turning to the issue of tasks, it is important to clarify initially that ROE do not allocate specific tasks within an operation; they cannot be used as authority to perform activities beyond those assigned by higher command or outside the scope of the mandate. In essence, while the ROE outline the authorizations (or restrictions) related to the use of force within the context of assigned tasks and duties, they do not impose an inherent obligation or authorization to create or seek out situations where such force might be used. The assignment of tasks or duties necessary to execute the operation and achieve its overall objectives is determined by the operational chain of command, in accordance with the mandate and relevant strategic instructions and directives.[120] ROE can therefore assist in fulfilling the tasks assigned to

[115] *Cf.* ICRC and Geneva Academy, *Expert Consultation Report…*, pp. 12–13.
[116] See M.N. Schmitt, M. Schauss, *Uncertainty in the Law of Targeting: Towards a Cognitive Framework*, "Harvard National Security Journal" 2019, Vol. 10, p. 151.
[117] An exception to this, however, is submarine operations.
[118] J. Kwik, T. van Engers, *op. cit.*, p. 7.
[119] V. Boulanin, M. Verbruggen, *op. cit.*, p. 15.
[120] J.F.R. Boddens Hosang, *Control Through ROE…*, p. 402.

operational forces, including both combat-related tasks and those not directly involving the use of force.

In the case of combat operations, ROE are usually permissive in nature—they empower commanders with the authority to use all essential measures to defend their units and generally permit the use of any lawful weapon or tactic for mission accomplishment. This permissive framework grants commanders significant discretion in crafting ROE tailored to their specific mission. At times, ROE also outline escalation-of-force procedures. Additionally, most ROE include other recurring elements such as identifying enemy hostile acts and intent, dealing with enemy forces declared hostile, and a positive identification requirement.[121] Besides authorizing or restricting the use of force, including delineating the various levels of force applicable in diverse situations, ROE frequently encompass directives on the positioning and posturing of forces, the employment of specific capabilities, the management and disposition of captured or detained individuals, and any delegation or withholding of authorities concerning approval of these actions. Ultimately, ROE serve as a pivotal command and control tool, intended to afford military and political leadership heightened oversight over the conduct of operations, whether combat-related or otherwise, by subordinate forces.[122] There are instances where ROE grant substantial discretion to commanders and decision-makers, not only due to legal allowances or requirements but also because such discretion is indicated or necessary for operational purposes.

Such a broad range of activities regulated by ROE does not fall within the scope of tasks that can theoretically be performed by autonomous systems, and particularly by AWS. Referring to the classification formulated by V. Boulanin, it is possible to distinguish five different "task areas" into which automation can be implemented:

1. Mobility (platform movement and navigation),
2. Health (survival management, e.g. refueling or self-repair functions),
3. Interoperability (communication with other agents in the environment),
4. Intelligence (analysis of tactical or strategic battlefield data) and
5. Force (detection, identification and engagement of targets).[123]

[121] G.D. Solis, *op. cit.*, p. 379. See also J.F.R. Boddens Hosang, *Rules of Engagement…*, p. 42.
[122] G.P. Corn, *op. cit.*, p. 212.
[123] V. Boulanin, *Mapping the development of autonomy in weapon systems: A primer on autonomy*, SIPRI, December 2016, p. 7.

The first three tasks (Mobility, Health and Interoperability) are often referred to as "operational functions" while the latter two (Intelligence and Force) are often referred to as "critical functions". Boulanin accurately asserts that it is frequently the critical task areas that are deemed more problematic, or at least which raise more concerns,[124] since critical functions are those that trigger a substantial likelihood of incompliance with IHL.[125] Importantly, given the intended purpose of AWS, it is precisely these functions that they are designed to perform. Since critical functions are those related to targeting, this implies that the tasks assigned to AWS are limited to combat operations where the use of force is essentially indispensable. Consequently, the ROE designated for AWS—as mentioned earlier—will either be directed to the military operation commander, taking the form of weapon-specific ROE, or they will be developed with the intention of being implemented by the autonomous systems themselves in the context of targeting. However, it is essential to clearly define the objectives of the operation and the specific tasks of AWS. If AWS are granted a certain degree of decision-making autonomy, which is inherent in the ROE, in situations characterized by poor situational awareness and significant battlefield dynamics—conditions that entail a high degree of uncertainty and unpredictability—there is a risk that AWS may fail, thereby violating the ROE, and potentially even leading to breaches of IHL.[126] In other words, AWS could be applicable in situations of deliberate targeting operations that may commence weeks, months, or even years before an attack on a target occurs, allowing ample time to employ a strategic approach. The use of AWS should be avoided in dynamic targeting situations, where the level of uncertainty and unpredictability increases. Although dynamic targeting operations follow similar decision-making steps, the timeline for deciding to attack a target is significantly shortened, allowing military forces to respond more swiftly and effectively to rapidly changing situations.[127]

The challenges associated with the proper functioning of autonomous systems when the level of uncertainty increases have been clearly explained by M.L. Cummings, who analyzed "the stages of reasoning that any agent

[124] *Ibidem*, pp. 7–8.

[125] See N. Hayir, *Defining Weapon Systems with Autonomy: The Critical Functions in Theory and Practice*, "Groningen Journal of International Law" 2022, Vol. 9, № 2, p. 265.

[126] Such a situation would trigger the entire procedure concerning the commander's responsibility for violations of humanitarian law, since the commander is responsible for ensuring the legality of the attack and, when delegating certain tasks to autonomous systems, must ensure that the necessary legal assessments are properly conducted (*cf.* T. Zurek, J. Kwik, T. van Engers, *op. cit.*, p. 3).

[127] See ICRC and Geneva Academy, *Expert Consultation Report…*, p. 8.

must possess in order to deal with increasingly complex decision-making scenarios, including those of autonomous weapons operation."[128] Drawing on J. Rasmussen's SRK (skills, rules, and knowledge-based behaviors) taxonomy, she developed a model that explicitly represents expertise and uncertainty. According to Cummings, skill-based tasks are the easiest to automate due to their repetitive nature and reliance on feedback loops, provided the necessary sensors are in place. Rule-based behaviors are also suitable for automation but become challenging as uncertainty increases, requiring a shift to knowledge-based reasoning, which necessitates expertise to manage uncertainty effectively. The author uses navigation as an example, noting that while drone navigation[129] is relatively simple due to low uncertainty, car navigation is more complex due to higher uncertainty and less reliable sensors. Finally, in highly uncertain situations, where knowledge-based reasoning is required, algorithms may struggle to identify and apply the correct set of rules, making it difficult to find viable solutions.[130]

It is widely acknowledged that AI has not yet advanced to the point where it can operate fully autonomously throughout the entire targeting phase, specifically in making independent decisions regarding whom to target and when to deploy weapons.[131] However, the complexity and range of tasks assigned to AI will inevitably continue to expand, necessitating a transition from skill- and rule-based tasks to those based on knowledge and goals.[132] As AI takes on more dynamic and complex functions, such as target classification, prioritization, and engagement, there will be a growing need for knowledge-based reasoning and deep learning models that facilitate flexible decision-making. On the other hand, the increased complexity of the envisaged tasks introduces a greater degree of uncertainty[133]—and ROE have many provisions which implicitly or explicitly entail an assumption of foreseeability and predictability during the execution of assigned tasks. Likewise, since many concepts in IHL rely on notions of foreseeability and predictability, the high degree of uncertainty presents further challenges to the effective implementation of IHL obligations.[134]

In theory, the solution seems to align the functions and tasks of autonomous systems so that they can perform them in accordance with the ROE.

[128] M.L. Cummings, *op. cit.*, p. 5.
[129] This refers to navigation in the aerial environment.
[130] *Ibidem*, p. 6–7.
[131] T. Vestner, A. Rossi, *op. cit.*, p. 517.
[132] J. Kwik, T. van Engers, *op. cit.*, p. 5.
[133] *Ibidem*.
[134] *Ibidem*, p. 2.

Significantly, if a system is tasked with critical functions related to targeting, the level of predictability required must be higher, depending on the environment in which it will be deployed. For example, a system responsible for targeting functions must achieve a very high level of confidence in the legality of a target before it can engage, even when operating alongside a human operator. The acceptable level of predictability for each system-scenario combination would then serve as the benchmark for evaluating the algorithm in order to meet legal review requirements. Alternatively, the system could be deployed in very restricted operational contexts, thereby limiting its potential, or implementing safety measures that might restrict its functions to non-critical ones.[135]

5. Machine learning—opportunity or challenge?

Although AWS are not yet standard equipment in modern armies and are absent from contemporary battlefields, their future deployment will face limitations related to the environments in which they operate and the tasks/functions they are expected to perform. Consequently, ROE developed for these systems will need to account for these limitations and consider that the complexity and dynamics of modern military operations may lead to unpredictable behavior and inexplicable responses from AI-driven AWS. Indeed, as evidenced by various applications, the characteristics of current AI systems significantly impair our ability to predict the system's outputs and understand the reasons behind them.[136] According to T. Vestner and A. Rossi, the unpredictability of AI systems "is of a different nature and degree depending on the AI techniques employed, namely hand-coded programming or machine learning."[137] In regards to the first technique, it involves programmers developing a model of the world, including its logical rules, which are embedded into the system to enable autonomous operation. Since this model is created by humans, the system's internal processes are understandable and transparent. By providing specific inputs, one can observe the system's responses and trace the reasoning behind its outcomes. However, this handcrafted approach is limited to environments that can be reduced to clear mathematical rules. Given the unpredictable and constantly changing nature of armed conflicts, this approach is generally unsuitable for such scenarios. As a result, handcrafted AI is only applicable in highly controlled and predictable environments, such as underwater military operations. Besides, it is

[135] T. Vestner, A. Rossi, *op. cit.*, p. 541.
[136] ICRC and Geneva Academy, *Expert Consultation Report…*, p. 16.
[137] T. Vestner, A. Rossi, *op. cit.*, p. 535.

impossible to test an AI system for every potential input it may encounter during operation, limiting the range of outcomes that can be observed and assessed. Consequently, the system remains unpredictable and prone to failure when faced with untested or novel situations.[138]

Regarding machine learning, this technique enables AI systems to improve their performance by expanding their knowledge. Machine learning is defined by a dual purpose: 1) scientific (understanding and mechanically generating phenomena related to temporal changes and adapting reasoning) and 2) practical (automatically acquiring foundational knowledge based on examples). Machine learning can be defined as the improvement of outcomes through experience, which is closely related to generalization. Machine learning is a technique that allows algorithms to extract correlations from data with minimal supervision. The goals of machine learning can be quite diverse but often involve attempting to maximize the accuracy of the algorithm's predictions.[139]

Machine learning enables a system to learn how to perform tasks by analyzing large amounts of data, identifying correlations, and building a representation of the world. Unlike traditional programming, it does not require explicit instructions; instead, developers must design a structure that allows the system to learn and adapt to changing conditions, providing it with extensive, well-selected data relevant to its operating environment. Learning capabilities indicate that, unlike handcrafted systems, the system alters its structure, program, or data (based on inputs or external information) in a way that enhances its anticipated future performance. For instance, when deployed in armed conflict settings, a machine learning system might develop its own criteria for applying a set of rules based on observations made on the battlefield. As regards learning processes, they can take place either offline or online. In the offline scenario, the algorithm is trained during the development phase. In the online scenario, the system continues to learn and adapt while in deployment. Due to these technical characteristics, machine learning systems are particularly well-suited for operating in complex environments where hand-coded programming would be inadequate. In the military domain, machine learning is already utilized for tasks such as target recognition.[140]

However, the machine learning process is not without flaws and involves at least two challenges that undermine the reliability of decisions made by

[138] *Ibidem*, pp. 535–536.
[139] M. Świerczyński, Z. Więckowski, *op. cit.*, p. 43.
[140] T. Vestner, A. Rossi, *op. cit.*, pp. 537–538.

the AI system. The primary issue with machine learning is its dependence on training data for learning. The system's effectiveness relies on data that accurately represents the environment in which it will operate to prevent biases or significant flaws. In complex scenarios, the range of inputs is far greater than in handcrafted systems, potentially altering the system's parameters and increasing uncertainty about what the system will learn and how it will respond to new inputs.[141]

Another challenge is inherently tied to the process by which machine learning systems derive their conclusions from given inputs, often referred to as the "black box" or opacity problem. The term "black box" refers to complex algorithms whose implementation and operation are opaque, meaning that the internal functioning of the method is difficult to comprehend. Deep neural networks are typically so intricate that it is impossible to determine specific weights for individual features globally, that is, across all possible predictions.[142] In other words, researchers are unable to fully explain the functioning or the outcomes of AI—neither programmers nor users can fully understand how, and why, a given output was produced, which entails certain risks, such as actions that may not align with the original intentions of the AI system creators.[143] This indicates that the system is unable to provide justifications for its estimations or decisions, lacking what is known as explainability. When the learning process of a system is confined to the offline phase, the system remains unchanged once it is operationally deployed by the military. While this may enhance the predictability of the system's future behaviors, it does not necessarily mean that the system operates deterministically—producing identical outputs when given the same inputs. Even in the case of deterministic AI systems, predictability is still limited. It is unrealistic to expect that these systems will encounter exactly the same inputs in real-world scenarios as they did during the pre-deployment assessment.[144]

This lack of predictability in AWS decision-making calls into question its ability to operate in accordance with the ROE established for a given operation. Since ROE determine when, where, how and against whom force may be used and consist of both permissions for, and restrictions on, the use of force and other provocative actions, they enable commanders to better un-

[141] *Ibidem*, p. 538–539.
[142] M. Świerczyński, Z. Więckowski, *op. cit.*, p. 41.
[143] See K. Kowalczewska, op. cit., p. 22, fn. 29; M. Pacholska, *op. cit.*, p. 5.
[144] T. Vestner, A. Rossi, *op. cit.*, p. 539–540. *Cf.* M.L. Cummings, *op. cit.*, p. 8.

derstand, anticipate, and customize force actions.[145] Moreover, ROE serve as tools for command and control, and "by controlling the use of force, ROE can influence the actions taken by other parties (including opposing forces)."[146] However, how can the use of force be effectively controlled when one component of the armed forces, driven by AI, is inherently unpredictable in its actions? Naturally, at various levels of military command, preceding the final decision to deploy an AI system, steps should be taken to determine whether the involvement of AI would facilitate or hinder the achievement of operational objectives and whether this can be accomplished in compliance with IHL and ROE. It is also essential to anticipate the risks associated with the use of AI systems and to develop appropriate parameters and constraints to mitigate these risks and ensure adherence to the relevant norms and rules.[147] However, given the inherent unpredictability of complex and dynamic environments like armed conflicts, it is highly doubtful that developers and their organizations could certify that a system would respond safely or appropriately to any input or conditions it might encounter.[148] This means that until the "black box" issue and the associated unpredictability and lack of explainability in decision-making by machine-learning AI systems are resolved, it will be impossible to implement effectively enforceable ROE for AWS. Consequently, the use of such systems may be significantly limited, likely restricted to handcrafted AI, applicable only in highly controlled and predictable environments.

6. Concluding remarks

There is no doubt that the question of whether AWS will be used on the battlefield is now obsolete—it should rather be asked when this will occur. As technologies continue to progress, States' interest in using AI to manage military operations, particularly targeting cycles, is likely to grow. Developments in machine learning have prompted politicians and military officials to focus their research and development efforts on enhancing the autonomy of weapon systems through AI.[149] Supporting this thesis are, for example, the plans of the US Department of Defense, which has officially announced its

[145] M. Faix, *op. cit.*, p. 138.

[146] J.F.R. Boddens Hosang, *Rules of Engagement and Targeting*, [in:] P.A.L. Ducheine, M.N. Schmitt, F. Osinga (eds.), *Targeting: The Challenges of Modern Warfare*, T.M.C. Asser Press, The Hague 2016, p. 164.

[147] See A. Holland Michel, *The Black Box, Unlocked: Predictability and Understandability in Military AI*, United Nations Institute for Disarmament Research, Geneva 2020, p. 15.

[148] ICRC and Geneva Academy, *Expert Consultation Report*…, p. 18.

[149] T. Vestner, A. Rossi, *Legal Reviews*…, p. 518.

intentions to develop and deploy advanced AWS before the year 2038.[150] On the other hand, the question of how to utilize this technology in a way that ensures it functions in compliance with legal and operational requirements in a military context remains unresolved. One must agree with the opinion of A. Wyatt that "[d]eveloping a definition for a complete lethal autonomous weapon system […] is arguably one of the major stumbling blocks to developing an effective international response to the emergence of increasingly autonomous military technology, whether regulation or a developmental ban".[151] However, it is not only about defining these systems but also about determining the extent of their autonomy—whether we will allow them to make decisions regarding critical functions, or if we will decide when and how they should engage. In the context of developing ROE for AWS, as J.F.R. Boddens Hosang aptly noted, an analysis must first determine how the autonomous weapons were intended to be used and integrated into military operations, followed by an assessment of their actual performance versus their intended function. Once this understanding is established within an operational context, the applicability of ROE to these technologies can be tested. This evaluation will determine whether standard ROE are sufficient or if specific or specialized ROE are necessary for their use.[152]

Given the characteristics of AWS, including the fact that they are driven by artificial intelligence and rely on algorithms, one could attempt a—still theoretical—assessment of the extent to which AWS would be capable of adhering to prohibitive rules and correctly interpreting the rules permitting the use of force. This assessment, however, is rather unfavorable. ROE are designed for various types of operations conducted in diverse environments and aimed at achieving various, often demanding tasks. Yet, the current level of technological advancement would limit AWS to operating in controlled, predictable environments and within the framework of deliberate targeting. Consequently, ROE—whether in the form of weapon-specific rules or directly "addressed" to autonomous systems—would need to be adapted to these conditions and circumstances. The use of AWS equipped with machine-learning AI systems is further constrained by the unpredictability and lack of explainability of their decisions—and ROE are created with the inten-

[150] US Department of Defense, *Unmanned Systems Integrated Roadmap FY2013-2038*, 2013, https://dod.defense.gov/Portals/1/Documents/pubs/DOD-USRM-2013.pdf (accessed: 6/20/2023).

[151] A. Wyatt, *So Just What Is a Killer Robot?: Detailing the Ongoing Debate around Defining Lethal Autonomous Weapon Systems*, Washington Headquarters Services, June 8, 2020, https://www.whs.mil/News/News-Display/Article/2210967/so-just-what-is-a-killer-robot-detailing-the-ongoing-debate-around-defining-let/ (accessed: 6/20/2023).

[152] J.F.R. Boddens Hosang, *Control Through ROE…*, p. 417–418.

tion of ensuring effective control, which implies operating within the boundaries set by these rules, even when autonomous decision-making is involved.

Therefore, changes will be needed in the approach to the current ROE doctrine, which must evolve alongside the AI-driven systems they govern during military operations to ensure that the overall purpose of ROE remains effective. Achieving this necessitates open dialogue aimed at fostering mutual understanding between ROE specialists and technologists, as well as between political and military commanders and the operators tasked with executing military operations.[153] At the operational and tactical levels, a potentially more effective approach than assigning tasks solely to AWS could be collaboration in the form of "cobotics", a term derived from "cooperation" and "robotics"—referring to the partnership between humans and robots. The goal of these cobots is to automate a broad range of tasks and work more closely with people.[154] In this scenario, appropriate ROE would also need to be developed, taking into account human-machine interaction and cooperation. Therefore, do autonomous weapon systems need specific ROE? At their current level of development, these systems would likely struggle to properly apply these rules, although such rules will undoubtedly be necessary when the deployment of AWS becomes a reality. Certainly, the task of formulating these rules so that they can be strictly adhered to and properly interpreted by these systems will be a true challenge for their designers.

[153] *Ibidem*, p. 418.
[154] See K. Kowalczewska, *op. cit.*, p. 21.