

## JOURNAL OF SECURITY AND SUSTAINABILITY ISSUES

ISSN 2029-7017 print/ISSN 2029-7025 online

2018 June Volume 7 Number 4

[https://doi.org/10.9770/jssi.2018.7.4\(1\)](https://doi.org/10.9770/jssi.2018.7.4(1))

### DISCUSSION ABOUT PREEMPTIVE BAN ON LETHAL AUTONOMOUS WEAPON SYSTEMS

**Tadeusz Zieliński**

*War Studies University, Faculty of Management and Command,  
al. gen. A. Chruściela "Montera" 103, 00-910 Warsaw, Poland*

*E-mail: [t-zielinski@akademia.mil.pl](mailto:t-zielinski@akademia.mil.pl)*

*Received 20 January 2018; accepted 15 May 2018*

**Abstract.** Advanced studies on artificial intelligence increase the concerns of many non-governmental organizations regarding the possibility of their use in combat systems. Even now, there are many weapon systems that can interact with the enemy in an automated manner. However, the concerns are raised due to the potential emergence of autonomous combat systems, which will independently make decisions about life or death on the battlefield, both in relation to military personnel and civilians. This article provides a review of terminological issues associated with the concept of automation and autonomy in the context of combat systems. It was concluded that the absence of a commonly accepted definition of "lethal autonomous weapon system" within international law contributes to many interpretations in this field. On the basis of source literature, the second part of this article presents current proposals aimed at a preventive ban on the design and use of lethal autonomous weapons systems. The essential deduction leads to the conclusion that introduction of a total ban on the design and use of this type of weapon systems is impossible, nevertheless as much as possible should be done in order to make sure that the critical decisions, especially those about life and death, will remain in the scope of human control and that they will not be made by the machines.

**Keywords:** automation, autonomy, international humanitarian law, lethal autonomous weapon system, meaning human control, killer-robot

**Reference** to this paper should be made as follows: Zieliński, T. 2018. Discussion about preemptive ban on lethal autonomous weapon systems, *Journal of Security and Sustainability Issues* 7(4): 625-634. [https://doi.org/10.9770/jssi.2018.7.4\(1\)](https://doi.org/10.9770/jssi.2018.7.4(1))

**JEL Classifications:** O14, O33

## 1. Introduction

Technological development inseparably accompanies the social development affecting humanity in a significant manner. Groundbreaking technologies make it easier for people to operate in many aspects of everyday life, however on the other hand they may also constitute a threat. It seems that artificial intelligence belongs to this type of technologies, which may change the perception of modern world in the future, but it may also be a threat in the case of its application in combat operations. Currently, there is a discussion on various types of forums – in mass media, scientific centres, non-governmental organizations, as well as international agencies, regarding the use of artificial intelligence in various manifestations of social life. It may be assumed a priori that the emergence of intelligent machines (robots) capable of self-learning may result in many benefits. However, there are also statements that negatively refer to the future use of artificial intelligence, or even such statements that call for the cessation of studies in this field. Such authority as Stephen Hawking said that "Success in creating AI would be the biggest event in human history (...). Unfortunately, it might also be the last, unless we learn how to avoid the risks. In the near term, world militaries are considering autonomous-weapon systems that can choose and eliminate target (...). Humans, limited by slow biological evolution, could not compete and would

be superseded by A.I.” (Sainato, 2015). Elon Musk warns that artificial intelligence is „our greatest existential threat” (Sainato, 2015) and points to the fact that „(...) Alongside the benefits, AI will also bring dangers, like powerful autonomous weapons, or new ways for the few to oppress the many (...) and in the future, AI could develop a will of its own – a will that will conflict ours” (Martin, 2017). The threats articulated in relation to artificial intelligence mainly refer to combat applications of the systems equipped with artificial intelligence or simply autonomous systems. The above-mentioned Elon Musk and Mustafa Suleyman, who is responsible for artificial intelligence in Google, as well as 116 other specialists from 26 countries, signed a letter in 2017 calling the United Nations (UN) to ban the works on these technologies.

It should be assumed that the development of artificial intelligence is not a threat in itself. The problem concerns its use in the Lethal Autonomous Weapon Systems (LAWS), their possible development and application in armed conflicts. The possibilities of their use in combat activities have their supporters, as well as opponents. Widespread attention of the media was drawn by an open letter drafted in 2015 and signed by over 3 000 scientists involved in robotics and studies on artificial intelligence, which was aimed against the approaching arms race in the scope of artificial intelligence and robotics (Future of Life Institute, 2015). This is part of a wider “campaign to stop killer robots” led by a coalition of leading non-governmental organizations. This issue also has been appearing on the forum of UN agencies dealing with arms control since 2013, as well as on the forum of the Human Rights Council and General Assembly of the First Committee on Disarmament and International Security. In the meantime, the European Parliament also called for a total ban on the development, production and use of completely autonomous combat systems (European Parliament, 2014), and this matter was raised at the UN, where a decision was made in 2016 regarding the establishment of an informal group of government experts, whose task will be to develop a report on lethal autonomous weapon systems.

It should be clearly emphasized that there are no currently functioning fully autonomous weapon systems. Nevertheless, now is the time for an in-depth discussion regarding this matter. Terminology covering the issues of autonomy comes to the fore - there are no recognized and acceptable definitions of autonomous combat systems within international law. The second aspect, raised in this article, refers to the potential activities associated with the preventive limitation of the design and implementation of lethal autonomous weapon systems. It is necessary to clearly emphasize that this is just one of the approaches to this kind of armament, because there is a large number of supporters, who advocate their development and implementation in the scope of equipping the armed forces, and present sensible arguments in this regard.

## **2. Terminological context regarding lethal autonomous weapon systems**

The term “Lethal Autonomous Weapon System” is relatively new in the literature. For the first time, it appeared in the documents drawn up by the International Committee of the Red Cross in 2014 during a meeting of the experts from the field of autonomous technologies. In the report resulting from this meeting, the lethal autonomous weapon systems were defined as “weapons that can independently select and attack targets, i.e. with autonomy in the ‘critical functions’ of acquiring, tracking, selecting and attacking targets” (ICRC, 2014). This definition was maintained during a second meeting of the experts in 2016, with reference to the lethal autonomous weapon systems as an important issue of functioning of the Convention on Certain Conventional Weapons (ICRC, 2016).

Experts of the International Committee of the Red Cross indicate that this is a working definition (until this moment, there is no unequivocal definition of a lethal autonomous weapon system in the international law), referring to any weapon system that can perform functions autonomously. The proposed definition is aimed at supporting a better understanding of the problem and helping in discussions associated with lethal autonomous weapon systems, including in the legal and ethical application of this type of weapons. Some experts propose more precise definitions that would be able to show the differences between autonomous and highly automated systems. However, it seems that from a technical point of view this boundary gets blurred, while from a legal as well as ethical perspective the problems in both cases remain the same.

It should be mentioned that the term Lethal Autonomous Robotics (LAR) appeared in the early phase of discussion on autonomous weapon systems and it was understood as „weapon systems that, once activated, can select and engage targets without further human intervention” (United Nations, 2013). A term that appears very often in the public media discourse is “Killer Robots”, which is mainly associated with the science fiction movies. Currently, the prevailing view is that the name Lethal Autonomous Weapon Systems is more adequate. Therefore, it is necessary to clarify the concept of “autonomy” along with the related terms, which often lead to the terminological ambiguity. This is an important issue, e.g. from the viewpoint of legality of the use of lethal autonomous weapon systems in combat operations.

The fundamental problem in terminological issues is the distinction between autonomy and automation. Automatic systems are usually programmed to carry out specific activities in a defined and controlled environment. In turn, autonomous systems decide on their own whether to perform or not to perform the given task in a changing and unpredictable environment. The functioning of automated systems is based on specific principles, so it is deterministic and thus predictable. For example, the *Phalanx* system that has been used by the US Navy since 1980 and its ground defensive version - *Centurion*, as well as the Israeli *Iron Dome* are more advanced systems compared to the machines-automatic devices, but their principle of operation is based on the same model. Autonomous systems are much more independent, they have much greater freedom of action and thus, they are more unpredictable. Importantly, both categories - automatic and autonomous system - are not mutually exclusive and they are not homogeneous, because certain levels of automation and autonomy function in their case. There are no absolute criteria for distinguishing between automation and autonomy, but rather we may say that there is a continuum between them. It can be assumed that future combat systems will be multi-modal, hybrid: autonomous for certain roles (e.g. steering, launching) and automatic for the others (e.g. moving around).

In the case of many definitions of autonomy, it is indicated that the basis for the functioning of a given autonomous system is the lack of human intervention. However, in such case, the automated processes, in which certain sequences of actions will be performed without operator’s intervention, may also be considered as autonomous. Autonomous systems are independent in making decisions associated with the performance of a specific task. Thus, they can adapt to unpredictable situations and to a large extent this distinguishes them from automated systems, where during their programming various types of situations are anticipated and algorithms for proceeding in the case of their occurrence are developed. Comparing an autonomous and automated system, it may be concluded that the former is the final element of automation that is able to independently implement all or almost all of the actions in the form of a sequence of formalized operations. Therefore, it may be assumed that autonomy has its limits and a human may control the machine at any moment, even if such person is not in the decision loop (he/she is not an operator). In other words, such person is an element that has planned the whole process, along with individual sequences of its implementation (Kopeć, 2016). The above-mentioned approach describing the autonomy may be described as an extreme element of automation.

The second approach assumes the functioning of autonomy as an element different from automation. In this case, the behaviour of autonomous systems is not programmed and they have the ability to make independent decisions (Kopeć, 2016), i.e. in this case we can talk about artificial intelligence, which has the ability to make decisions and, moreover, to learn from gained experience, as well as to take into account the resulting conclusions in further activities.

In the case of defining autonomous systems, the most common approach includes the criterion of the degree of human control over a machine (human-machine relation). This categorization distinguishes semiautomatic systems (“human in the loop”), in which the automatic operation is possible until certain moment and then the human intervention is necessary. For example, weapon systems may independently affect the targets that have been previously selected and approved for combat by a human (operator). This type of armament includes various types of guided ammunition, also the “fire-and-forget” type (AIV/CAVV, 2015). The second group covers supervised systems (“human on the loop”), in which there is a possibility of uninterrupted autonomous operation, but with the possibility of human intervention at any given moment. Weapon systems from this group are

able to select and combat targets on their own, from the moment they were activated. However, the operator of such weapon system has the knowledge about what kind of objects can be targeted and the operator may intervene at any time by interrupting the process. In practice, these types of weapon systems (supervised) are used in defensive operations and in undemanding operational environment. They react directly to the defined threats, where the reaction of a human (operator) could be too slow, and in the case of doubtful situations the operator may react at any time (AIV/CAVV, 2015). The third group consists of fully autonomous systems (“human out of the loop”), without the possibility of human intervention. They refer to weapon systems that independently, without human participation, are able to select and combat targets in a previously defined geographical region, time interval and according to the adopted rules. The operator does not know what targets will be selected for combating, but it should be remembered that the types of combated objects have been previously defined by a human according to the specific criteria. In other words, a man decides earlier in what manner the autonomous combat system will carry out its tasks.

The above-mentioned considerations lead to the ascertainment that complexity of the system itself is manifested, when such expressions as: “automatic”, “automated” or “autonomous” are used for its description. Automatic systems implement simple mechanical functions that constitute a response to the environmental stimuli. Automated systems are more complex, because they shape their behaviours not only based on environmental stimuli, but they also take into account the previously defined rules. On the other hand, autonomous systems have the ability of self-direction, which does not result directly from the implemented software. Therefore, they are also able to shape their behaviour through the ability to self-learn. An interesting view on this issue is presented by A. Williams. He points out that in principle we should talk about the functions implemented by the combat system, which are performed in an automated or autonomous manner. According to him „Autonomous functioning refers to the ability of a system, platform, or software to complete a task without human intervention, using behaviours resulting from the interaction of computer programming with the external environment. Tasks or functions executed by a platform, or distributed between a platform and other parts of the system, may be performed using a variety of behaviours, which may include reasoning and problem solving, adaptation to unexpected situations, self-direction, and learning. Which functions are autonomous – and the extent to which human operators can direct, control, or cancel functions – is determined by system design trade-offs, mission complexity, external operating environment conditions, and legal or policy constraints. This can be contrasted with automated functions, which (although they require no human intervention) operate using a fixed set of inputs, rules, and outputs, the behaviour of which is deterministic and largely predictable. Automatic functions do not permit the dynamic adaptation of inputs, rules, or outputs” (Williams, 2015).

To sum up, it seems impossible to define a precise boundary between autonomous, automatic or automated weapon systems. Moreover, individual degrees of autonomy within the presented aspects may remain in multithreaded relations to each other. Simply put, at one end of the scale there are simple elements acting on the principle of stimulus-reaction, while on the other end there are complex systems capable of learning and drawing conclusions based on information obtained from the environment, in which they function. The distinguishing characteristic is the ability to adapt, independence in functioning and system’s ability to independently make decisions. The final conclusion is that currently it is not possible to talk about fully autonomous weapon systems (in reference to the existing weapon systems).

### **3. Proposals for limiting design and employment of lethal autonomous weapon systems**

Dilemmas associated with the future use of lethal autonomous weapon systems will deepen due to the imprecision of legal regulations, different perceptions by society and the military, as well as ethical and moral concerns. It should be assumed with high probability that the works on artificial intelligence will be continued and, as a consequence, implemented in the projects of modern weapon systems absorbing the autonomy. In this context, it is reasonable to formulate the following question: should the lethal autonomous weapon systems be subject to a preventive ban on the design and use? Despite the fact that such systems are not functioning currently in the armed forces of any country, the discussion regarding the use of artificial intelligence for the development of autonomous weapon systems, which may have a lethal effect, has been ongoing on various

forums for several years. Also, the propositions emerged in the scope of preventive restrictions regarding the development of autonomous combat systems, which cover various solutions.

It seems that one of the radical solutions is the introduction of a total ban on the design and use of lethal autonomous weapon systems. Certain non-governmental organizations are calling for the introduction of a ban on functioning of autonomous combat systems under an international agreement. The Human Rights Watch is particularly active in this area (Human Rights Watch, 2012), along with its Campaign to Stop Robot Killers initiative. It is a global coalition of over 64 international, regional and national non-governmental organizations from 28 countries supporting the above-mentioned ban (Campaign to Stop Robot Killers, 2017). Some experts claim that there is an ethical and legal obligation to prevent a future situation, in which autonomous combat systems could independently make decisions relating to human life or death (Asaro, 2012). The justification for the aforementioned statement is based on two key arguments. Firstly, it cannot be ensured that lethal autonomous weapon systems will be able to comply with the provisions of international humanitarian law. Secondly, further dehumanization of war would lower the threshold for starting armed conflicts even more. There may arise legal gaps in relation to the responsibility, which can only be solved by using a total ban on the design and use of lethal autonomous weapon systems. According to the Human Rights Watch, these types of problems arise not only in relation to autonomous weapon systems, but also in relation to weapons, which have such a large influence on the selection of targets that the actual human control over them is doubtful (Human Rights Watch, 2012).

Critics of a total ban point out that at such an early stage of development of the technologies associated with artificial intelligence, including lethal autonomous weapon systems, their banning would be irresponsible. According to them, this type of armament provides a chance to drastically reduce the risk borne by your own soldiers, which is certainly an important factor in the framework of international humanitarian law (Schmitt, Thurnher, 2013). Moreover, according to the opponents of this ban, it makes no sense, because development and implementation of such systems are already inevitable, since in principle the necessary technologies already exist (Anderson, Waxman, 2013). Proponents of a total ban strongly reject this argument (Asaro, 2012). Technological progress is not inevitable and there have been new types of weapons, such as blinding laser weapons (O'Connell, 2014), which had been banned before they were implemented. On the other hand, this is the only example of a combat system in regard to which such far reaching agreement has been successfully achieved (Geiss, 2014).

In his report for the UN Human Rights Council, the UN Special Rapporteur - Christof Heyns presented the view that the issue of a total ban on the design and use of autonomous combat systems requires an international negotiation process, in order to find answers to the questions arising from the essence of functioning of the autonomous combat systems. For this purpose, an interdisciplinary committee should be established at the level of the United Nations, whose aim should be to achieve a consensus regarding the international legal framework for the development and implementation of lethal autonomous weapon systems (United Nations, 2013). Until the introduction of such framework, moratoria at national levels should be developed in the scope of research, production, sales and implementation of such systems (United Nations, 2014). Bioethicist W. Wallach also calls for the development of such moratoria, however none of the countries have decided to take such a step so far.

The second method of preventive restriction of the lethal autonomous weapon systems refers to the use of the "meaning human control" requirement. It seems that an approach based on this requirement may constitute a substantive, constructive starting point for further debate. A view prevailing among many countries and experts is that "meaning human control" in relation to lethal autonomous weapon systems is a factor that makes it possible to exercise control over critical decisions made by autonomous combat systems, and primarily over the decision regarding the use of lethal force. What is supposed to be precisely understood as "meaning human control" remains an open question. Nevertheless, this initiative is gaining importance among many countries, constituting a base for deepened discussion associated with the use of lethal autonomous weapon systems, including the possible ban on their use in some areas. The "meaning human control" expression itself suggests that there is no room for full autonomy. Thus, the criterion of "meaning human control" is to some extent contradictory,

which does not change the fact that it is a constructive starting point for further substantive discussions about which critical decisions should always be subject to such control and how the “meaning human control” should be implemented in detail. Therefore, it should be assumed that the required level of control may refer to several factors: period of time from the moment when the last decision was made by a human and exertion of power by the machine; environment in which the machine implements its tasks, particularly in relation to the question whether civilians are present in this environment; orientation of implementation, so whether the machine is supposed to perform defensive or offensive tasks; issue of machine’s configuration regarding the use of lethal force; level of training of people engaged in exercising control over the machine; the issue of possibilities and scope of human interference in the machine’s activities and, if necessary, stopping of the mission; implementation of safeguards in relation to responsibility, e.g. through complete electronic registration of robot’s activities. These are just some of the aspects that must be considered when using the “meaning human control” concept.

It should be noted that the concept in question is not reflected in the provisions of normative documents under international law. This wording was described by non-governmental organization called “Article36”, which points out that “meaning human control” refers to the information held by operator or other supervisor of the weapon within the planned combat action (knowledge of: target itself, area of attack, why a specific target has been selected for combating, mission objectives, desired effects associated with combating a specific object), the need for active human activity during implementation of the attack (commencement of the attack requires a positive acceptance of the operator), as well as the responsibility of persons evaluating the possessed information and for results of the attack (Article36, 2013). Despite the lack of an acceptable international definition of “meaning human control”, both opponents and supporters of autonomous combat systems indicate the need of ensuring some control over functioning of the above-mentioned systems, even if only for the public acceptance. Thus, the debate on autonomous combat systems includes a clear voice of a significant part of the experts referring to the need for exercising human control over these systems. Very common is a public view that lethal use of weapon systems requires the involvement of a human in the decision-making process, who will be able to consider ethical, moral and legal aspects of the use of a particular weapon system. Consequently, the concept of “meaning human control” has its implications in relation to the responsibility for effects of conducted combat operations.

Another proposed solution aimed at preventive restriction of the functioning of LAWS is associated with interference in the robot’s software and it is referred to as the “ethical governor”. One of the supporters of this solution is R. Arkin, who is a scientist dealing in robotics. According to him, a ban on the design and use of autonomous weapon systems is not necessary, because in the medium-term perspective the technology will develop to a degree that will allow it to comply with provisions of the law of armed conflicts in a much better way than people. For this reason, the premature ban on these systems would be ethically irresponsible (Anderson, Waxman, 2013). In order to achieve the appropriate standard, Arkin proposes the “ethical governor” model, which is based on an algorithm located in software and used to control an autonomous combat system that either allows or prohibits the system to use weapons in specific cases (Arkin, 2007). In other words, the appropriate algorithm fulfilling the function of a security buffer would be programmed in the system, in order to assess compliance with the law before each use of lethal force, based on the principles of international humanitarian law. Arkin does not exclude that the use of “ethical governor” algorithm will not save the lethal autonomous weapon system from making mistakes, however he assumes that the error rate would be much lower than in case of the soldiers.

The US Department of Defense (DoD) applied an interesting approach to the considered issues. This approach is mainly pragmatic and it assumes that autonomous and semiautonomous combat systems are to be constructed in such a manner that there is always “an appropriate level of human assessment in relation to the use of force” (US DoD, 2013). The assumed standard is not completely unambiguous and it is probably weaker than the demand of “meaning human control”. The UN Special Rapporteur, Christof Heyns correctly noted that the word “appropriate” in some situations could mean not exercising any control over the use of a lethal autonomous weapon system. Basically, DoD focuses on restrictive requirements concerning the control of hardware and software of the systems checked during repeated tests of weapons. Operators of autonomous and semiautonomous systems are required to comply with the rules of armed conflicts and any other applicable treaties

of international law. It was clearly stated that only semiautonomous systems can be implemented to use lethal force. Autonomous systems cannot be used to select targets involving people, even if they are supervised by a human (US DoD, 2013). Nevertheless, DoD does not exclude the development of weapons that could be implemented outside these constraints, if they will meet more restrictive technical guidelines. W. Wallach correctly points out that in relation to the regulations regarding the tests for lethal autonomous weapon systems, military necessity can easily mean that such weapons will be used before the completion of required tests. For example, this happened in the case of the war in Iraq and Afghanistan, where new technologies were introduced, such as unmanned aircraft systems, before all scheduled tests were completed. A similar situation may occur in the case of lethal autonomous weapon systems (Wallach, 2013).

In the area of proposals associated with the preventive restriction of functioning of the autonomous combat systems, there are also those relating to the improvement of legal regulations associated with their use. Control mechanisms within international humanitarian law for the introduction of new weaponry technologies apply in themselves to the autonomous weapon systems. Article 36 of the First Additional Protocol and the Martens Clause result in the fact that new weapons cannot violate the fundamental principles of international humanitarian law and the applicable rules of international law. In this context, various countries, at the occasion of the expert meeting in Geneva in April 2015, advocated further development and determination of substantive requirements for the testing of newly developed weapons, within the meaning of Article 36 of the First Additional Protocol, especially in relation to the lethal autonomous weapon systems (US Delegation Opening Statement, 2015). Undoubtedly, a more detailed substantive development of the control mechanisms resulting from Article 36 of the First Additional Protocol would be helpful and desirable. However, it should be assumed that such controls in themselves are not able to properly assess various ethical dilemmas, issues of international law and problems arising from the functioning of lethal autonomous weapon systems.

The aforementioned ponderations leads to the conclusion that it is necessary to establish new rules and higher standards of protection in relation to the potential use of lethal autonomous weapon systems. This results from the fact that the currently existing legal regulations and regulatory systems were created only for people and their specific weaknesses and deficiencies. This generates a question about the extent to which battle robots, which are able to operate independently, can be covered by this legal system. Standards for conducting warfare are based on basic ethical assumptions created more than 100 years ago. Even if we will recognize that so far they have more or less fulfilled their task, then along with the development of new types of weapons, particularly in relation to the combat artificial intelligence, they do not seem to be too convincing. Supporters of lethal autonomous weapon systems point to their better (in comparison to human) capabilities in the scope of resistance to stress, accuracy and endurance. In other words, the elimination of real human weaknesses hypothetically is supposed to lead to "pure, ethical and lawful killing". On the other hand, both the deadly actions performed by machines and by people, as well as errors leading to death, are ethically equivalent. This leads to the ascertainment that it is not possible to easily assess the implementation of lethal autonomous weapon systems in the context of existing regulations, regardless of the fact which additional - legal or mechanical - security mechanisms will be used. Thus, there is a need to establish new and maybe much higher legal standards than those that are currently required by international humanitarian law. However, currently there are no indications for such approach in contemporary state practice. In the light of the current course of discussions on the occasion of expert meetings in Geneva, and above all in the context of current global political situation, such fundamental discussion is not in the immediate plans at the international level.

Scientists dealing with the issue of future functioning of lethal autonomous weapon systems also indicate other solutions ensuring a preventive restriction of this type of weapons. For example, J. Thurnher proposes the creation of four control mechanisms, which would always have to be taken into account before the implementation of autonomous weapon systems. Firstly, it is necessary to determine the clear rules of implementation, which will always be respected. Secondly, the scenarios of implementation should be limited in terms of time and space from the very beginning. Thirdly, autonomous systems should only be used in traditional armed conflicts between states, and not in civil wars or in order to suppress uprisings. Fourthly, human intervention (control) in the machine's activity must always be guaranteed (Thurnher, 2012).

On the other hand, Anderson and Waxman proposed a multilateral process implemented by experts, similar to the one that ultimately led to the preparation of (legally non-binding) “Tallinn Manual” in the scope of application of international law for conducting cyber wars (Shmitt, 2013). Within this process, it is necessary to examine in what manner the lethal autonomous weapon systems can be adequately regulated by applicable regulations of international law, and to develop appropriate (non-binding) guidelines (Anderson, Waxman, 2013).

In the context of a potential ban on the design and use of lethal autonomous weapon systems, an important voice in the discussion was provided by W. Wallach, who developed a plan consisting of 10 points, which strengthened the need for introduction of the ban. Basically, this plan is based on the introduction of the overriding moral principle on the basis of which the machines will never make decisions about life or death of people. Moreover, Wallach points to the fact that even the best technologies may fail, hence the constant control is necessary, and the process of possible implementation of autonomous combat systems should be transparent for all countries (Wallach, 2017).

## Conclusions

Lethal autonomous weapon systems should be recognized as prospective weapon systems. It may be assumed that in the longer perspective, they will change the quality of conducting combat operations, aspiring to be a turning point in the application of military technologies. Currently, no fully autonomous weapon systems are functioning, hence the ongoing debate regarding them is directed to the future, and therefore it includes uncertainties associated with their potential implementation and the resulting strategic advantages. Nevertheless, it should be assumed that far from the ongoing discussion regarding the combat robots, the importance of lethal autonomous weapon systems will increase in the longer perspective at all levels of military and political decision-making. This process will proceed gradually.

It must be emphasized that autonomy and automation in military systems are not a problem in themselves from the point of view of international law. Examples of their applications in the scope of disarming mines, such as sapper robots or autonomous (or automated) medical vehicles, do not raise any objections. It is also necessary to articulate the need to implement autonomous weapon systems in accordance with legal norms of the international humanitarian law, also in combat conditions. To what extent and when the technical development will allow to make it happen - especially in more complex conditions - remains unclear to date and we can only speculate. Nevertheless, the more precisely defined (in terms of time and space) scope of implementation and scope of tasks of lethal autonomous weapon systems, the greater probability that the legal provisions will be respected. In this context, it will be necessary to create appropriate control mechanisms based on the standards of international humanitarian law, which will allow to check whether lethal autonomous weapon systems are programmed in accordance with these regulations. On the other hand, the fundamental question that still remains current is whether it is ethically and legally justified to entrust decisions regarding the use of (potentially lethal) power to autonomous systems. Under international law, the soldiers and enemy fighters can be (in principle) attacked and killed as part of an armed conflict. However, the current law says nothing about whether they can be attacked and killed by autonomously operating machines.

It seems that the basis for further discussion and seeking solutions should be the introduction of the “meaning human control” requirement. This means that the starting point for a broader debate will be the basic assumption that critical decisions, so those concerning important legal interests, such as the right to life and the right to bodily integrity, cannot be delegated to fully autonomous systems in the legal (international law) and ethical context. In particular, the decisions regarding life and death must always be subject to final decision and control of a human. Therefore, any a priori demand for significant human control over critical decisions indirectly excludes the full autonomy of functioning of combat systems in these areas, so the ban on autonomy may only cover selected areas. In the case of going further, a more comprehensive ban on lethal autonomous weapon systems in the light of previous debates seems unrealistic at the international level. Undoubtedly, the inclusion of the “meaning human control” concept in detailed framework constitutes a challenge. Determining what decisions should be considered critical and what “meaning human control” really means require further



discussion with the participation of technical, military, legal experts, as well as experts on ethics. It is clear that the actual ability to exercise control, also in mission conditions, over critical decisions of the system must always be maintained.

To sum up, all initiatives aimed at increasing control over the decisions taken by lethal autonomous weapon systems have their justification. These initiatives should cover the following areas: technological, organizational, as well as legal area, and as a consequence they should prevent decision-making, particularly regarding life or death of people, by autonomous combat systems without human participation.

## References

- AIV/CAVV. 2015. Autonomous weapon systems. The need for meaningful human control. *Report No. 97 AIV/No. 26 CAVV*. Available on the Internet: <https://aiv-advice.nl/download/606cb3b1-a800-4f8a-936f-af61ac991dd0.pdf>
- Anderson, K.; Waxman, M.C. 2013. *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*. Available on the Internet: [http://media.hoover.org/sites/default/files/documents/Anderson-Waxman\\_LawAndEthics\\_r2\\_FINAL.pdf](http://media.hoover.org/sites/default/files/documents/Anderson-Waxman_LawAndEthics_r2_FINAL.pdf)
- Arkin, R.C. 2007. Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture. *Technical Report GIT-GVU-07-11*. Available on the Internet: <http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf>
- Article36. 2013. *Killer Robots: UK Government Policy on Fully Autonomous Weapons*. Available on the Internet: [http://www.article36.org/wp-content/uploads/2013/04/Policy\\_Paper1.pdf](http://www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf)
- Asaro, P.M. 2012. On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making, *International Review of the Red Cross* 94: 2012: 687–709. <https://doi.org/10.1017/S1816383112000768>.
- Campaign to Stop Robot Killers. 2017. *Who We Are*. Available on the Internet: <https://www.stopkillerrobots.org/coalition/>
- European Parliament. 2014. European Parliament resolution of 27 February 2014 on the use of armed drones (2014/2567(RSP)). Available on the Internet: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014IP0172&from=EN>
- Future of Life Institute. 2015. *Autonomous Weapons: An Open Letter from AI and Robotics Researchers*. Available on the Internet: <http://futureoflife.org/open-letter-autonomous-weapons/>
- Geiss, R. 2014. The Law of Weaponry from 1914 to 2014. Is the Law Keeping Pace with Technological Evolution in the Military Domain?, in Delbrück, J. et al. (Ed.). *Essays in Honor of the 100th Anniversary of the Walther Schücking Institute for International Law*. Berlin: Duncker & Humblot, 229–248.
- Gibbs, S. 2017. *Elon Musk leads 116 experts calling for outright ban of killer robots*. Available on the Internet: <https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war>
- Human Rights Watch. 2012. *Losing Humanity: The Case against Killer Robots*. Available on the Internet: [https://www.hrw.org/sites/default/files/reports/arms1112\\_ForUpload.pdf](https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf)
- ICRC. 2014. *Report of the ICRC Expert Meeting on “Autonomous weapon systems: technical, military, legal and humanitarian aspects”*. Available on the Internet: <https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf>
- ICRC. 2016. *Convention on Certain Conventional Weapons (CCW). Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 11-15 April 2016, Geneva. Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system, 11 April 2016*, Available on the Internet: <https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system>
- Kopeć, R. 2016. Autonomia systemów bojowych. *Przegląd Geopolityczny* 17: 133–147. Available on the Internet: <http://przegląd.org/wp-content/uploads/2016/09/08-Kope%C4%87.pdf>
- Martin, S. 2017. *When AI robots are in the streets killing us it will be too late, says tech guru Elon Musk*. Available on the Internet: <https://www.express.co.uk/news/science/830501/AI-robots-killing-too-late-Elon-Musk>
- O’Connell, M.E. 2014. Banning Autonomous Killing: The Legal and Ethical Requirement that Humans Make Near-Time Lethal Decisions, in Evangelista, M.; Shue, H. (Ed.). *The American Way of Bombing. Changing Ethical and Legal Norms, from B-17s to Drones*. Ithaca: Cornell University Press, 224–234.

Sainato, M. 2015. *Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence*. Available on the Internet: <http://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/>

Schmitt, M.N. (Ed.). 2013. *Tallinn Manual on the International Law Applicable to Cyber Warfare*. New York: Cambridge University Press.

Schmitt, M.N.; Thurnher, J.S. 2013. "Out of the Loop": Autonomous Weapons Systems and the Law of Armed Conflict, *Harvard National Security Journal* 4: 231–281. <http://harvardnsj.org/wp-content/uploads/2013/01/Vol-4-Schmitt-Thurnher.pdf>

Thurnher, J.S. 2012. No One at the Controls: Legal Implications of Fully Autonomous Targeting. *Joint Force Quarterly* 67: 77–84.

United Nations. 2013. *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns, 26 April 2013, United Nations, General Assembly, A/HRC/23/47*. Available on the Internet: [http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47\\_en.pdf](http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf)

United Nations. 2014. *General Assembly, 69th session, First Committee, statement by the ICRC*. Available on the Internet: <https://www.icrc.org/en/document/weapons-icrc-statement-united-nations-2014>

US Delegation Opening Statement. 2015. *CCW Informal Meeting of Experts on Lethal Autonomous Weapons Systems*. Available on the Internet: <https://geneva.usmission.gov/2015/04/15/u-s-opening-statement-at-the-ccw-informal-meeting-of-experts-on-lethal-autonomous-weapons-systems/>

US DoD. 2013. *Unmanned Systems Integrated Roadmap FY2013–2038*. Available on the Internet: <http://archive.defense.gov/pubs/DOD-USRM-2013.pdf>

Wallach, W. 2013. *Terminating the Terminator: What to Do about Autonomous Weapons*. Available on the Internet: <https://scienceprogress.org/2013/01/terminating-the-terminator-what-to-do-about-autonomous-weapons/>

Wallach, W. 2017. Toward a Ban on Lethal Autonomous Weapons: Surmounting the Obstacles, *Communications of the ACM* 60: 28–34. <https://doi.org/10.1145/2998579>

Williams, A. 2015. Defining Autonomy in Systems: Challenges and Solutions, in Williams, A. P.; Scharre, P. D (Ed.). *Autonomous Systems. Issues for Defence Policymakers*. Norfolk: Headquarters Supreme Allied Commander Transformation, 27–62.

**Col. Tadeusz ZIELIŃSKI** is Associated Professor at War Studies University in Warsaw, Poland. Currently he is a dean of Faculty of Management and Command. Crucial research interest: doctrinal and practical dimension of air power in contemporary conflicts, CSDP of the European Union, NATO capabilities in deterrence, modern technology in conflicts.