General Legal Limits of the Application of the Lethal Autonomous Weapons Systems within the Purview of International Humanitarian Law

Roman Dremliuga¹

¹ School of Law, Far Eastern Federal University, Vladivostok, Russia

Correspondence: Roman Dremliuga, School of Law, Far Eastern Federal University, Vladivostok, Russia. Tel: 7-914-707-1474. E-mail: dreamluck@yandex.ru

Received: April 14, 2020      Accepted: May 9, 2020      Online Published: May 25, 2020
doi:10.5539/jpl.v13n2p115
URL: https://doi.org/10.5539/jpl.v13n2p115

The study was carried out with the financial support of the Russian Foundation for Basic Research in the framework of research project No. 18-29-16129

Abstract
This article focuses on the problem of regulation of the application of the autonomous weapons systems from the perspective of the norms and principles of international humanitarian law. The article discusses the question of what restrictions are imposed on the application of such weapons in the international humanitarian law. The article presents a number of principles that must be met by both the weapons and their method of their application: distinction between civilians and combatants, military necessity, proportionality, prohibition on causing unnecessary suffering, and humanity.

The author concludes that from the perspective of the principles of the international humanitarian law, it is doubtful if autonomous systems would be able to comply with these principles. Weapons that hit targets without human intervention have been applied for a long time, but they have never had the independence that they have now. The issue of compliance of autonomous weapons systems with the international humanitarian law can be considered if sufficient experience of application of such weapons in real conditions is accumulated.

This study demonstrates that it is impossible to say that autonomous weapons systems do not comply with the principles of humanitarian law in general. The paper provides policy recommendations and assessments for each of the principles under consideration. The author also concludes that it would be necessary not to prohibit autonomous weapons, because they do not comply with the principles of international humanitarian law, but to develop rules for their application and for human participation in their functioning. A significant challenge to the development of such rules is the opacity of these autonomous weapons systems, if we look at them as at the complex intelligent computer systems.

1. Statement of the Problem

Artificial intelligence systems have already been applied in almost all parts of our lives. Autonomous systems compose poems and lyrics, issue loans, diagnose diseases, and teach children. Like any promising technology, artificial intelligence did come into the sphere of interests of the armed forces around the globe. Intelligent systems can be used by the armed forces in various applications ranging from improvement of the effectiveness of military training to analysis of strategic risks, but the application of artificial intelligence as the “digital brain” of autonomous weapons has attracted the greatest attention of the global community (Asaro, 2012). The main focus of the discussion here is on the ability of existing autonomous weapons systems to meet the legal requirements of international humanitarian law (IHL) and on the predictions that future technologies may or may not meet these requirements.

Autonomous weapons have been adopted in many countries around the world, and their presence determines the strategic military capacity of these countries (Gill, 2019). The advantages of autonomous weapons are obvious, as they are often more accurate, more effective, and are not subjected to “the human factor”. Such weapons can be cheaper to operate, and they can be easily improved by the software updates. At the same time, there are
concerns about the non-compliance of such weapons with the international law.

Red Cross defines an autonomous weapons system as “Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.” Such weapons are often referred to as lethal autonomous weapons or lethal autonomous armed systems (LAWS) (Egeland, 2016).

Many countries define lethal autonomous weapons in their own ways. For example, the Netherlands defines “LAWS” as “a weapon that, without human intervention, selects and engages targets matching certain predetermined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention.” France defines lethal autonomous weapons as “fully autonomous systems ... [that] implementing a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command ... [as well that] would be capable of moving, adapting to its land, marine or aerial environments and targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation.” The US Department of Defense on the other hand gives a definition of LAW as a system that is capable “once activated, to select and engage targets without further intervention by a human operator.” Similar definitions are given in some other countries.

Based on the definition, the key characteristic of such systems is their autonomy, which could be understood as their ability to act independently of human actions. Autonomy can be complete, when the system makes all decisions independently from the moment of launch, and partial, when a human operator takes part in making some decisions in one form or another. However, based on the definitions given above, such systems (LAWS) are meant to be fully autonomous.

Here decision-making refers to the choice of a target and the application of striking military force against this target. That is, first of all, an automated target selection, and not the management of movement or other activities of the weapon system. For example, an autonomous vehicle equipped with a machine gun that is remotely focuses on a target by an operator will not be considered to be an autonomous weapon.

Complete autonomy is a property characterized by an independent functioning and behavior. Despite the fact that such systems are developed by humans, it is quite difficult to predict how they will behave at one time or another. Moreover, some authors suggest that autonomous weapons are guaranteed to behave unpredictably in difficult situations of the real combat (Egeland, 2016). Thus, autonomy implies the possibility of action without human participation and a certain degree of unpredictability.

Such systems can operate on land (Nguyen, 2009), in the air (Wingo, 2018), and at sea (Wirtz, 2020) in the conditions that are not suitable for humans (zones of radioactive contamination, high temperatures, overloads, etc.). Currently, there are many examples of how the application of the autonomous weapons systems has increased the effectiveness of solving combat tasks. For instance, the US and Israel jointly developed and commissioned the Iron Dome system, which protects against ground-to-ground weapons such as mortar mines and rockets fired at Israel. The Iron Dome consists of three subsystems: anti-missile, artillery-mortar, and close-range air defense. The system automatically intercepts up to 90% of all missiles launched from the territories surrounding Israel (Grudo, 2016).

The lethal autonomous weapon is actively used to secure borders. Such systems are already deployed by Israel and South Korea. Another example of promising lethal autonomous weapon is an automated gun turret Super

---

6 https://www.wired.com/2007/06/for-years-and-y/
7 https://singularityhub.com/2010/12/16/south-koreas-robot-machine-gun-turret-can-see-you-coming-3-km-away/
aEgis 2. It is developed in South Korea and can detect and lock onto human targets from kilometers away.\(^8\) The turret is able to operate without any intervention of an operator. The weapon is exported in many countries but with human-in-the-loop regime.\(^9\) Above examples do not rise legal questions because these weapons can only fire on targets that are encroaching a well-delimited area (Johnson, 2013).

Despite the fact that autonomous weapons systems demonstrate high performance in testing and operation, there are concerns that the application of such weapons may violate the norms and principles of the international humanitarian law (Garcia, 2018; Egeland, 2016). Some authors argue that the application of autonomous weapons systems in general threaten the world order (Sharkey, 2010; Rosert, Sauer, 2019).

2. Existing Norms of the International Humanitarian Law That Apply of the Autonomous Weapons

Rules relating to the legality of the application of one weapon or another are contained in many international agreements, but most notably the Geneva Conventions of 1949\(^10\) and their protocols\(^11\). The assessment of the legitimacy of the application of autonomous weapons is based on a number of principles that must comply with both the weapon and the method of its application, which include distinction between civilians and combatants, military necessity, proportionality, prohibition on causing unnecessary suffering, and humanity.\(^12\) The objective that brings together these principles is to minimize casualties, suffering, and material losses of the civilian population. The ideal war from the standpoint of the international humanitarian law, which is probably impossible in reality, is an armed conflict that does not cause any inconvenience to the civilian population.

Making a Distinction Between Civilians and Combatants. According to this principle, belligerents must always distinguish between the civilian population and combatants and between civilian and military infrastructure. Participants of an armed conflict have to direct their actions only against military targets. This principle does not mean that any loss of civilian life or damage to civilian objects is a violation of international humanitarian law, but the belligerents are obliged to minimize such situations. Article 48 of Additional Protocol I to the Geneva conventions provides that belligerents must always be able to distinguish between combatants and civilians, as well as between military and civilian installations and infrastructure. Thus, the validity of the application of the new weapons systems should be checked in terms of the ability to target a military forces and military facilities.

Autonomous weapons systems can successfully cope with this task, according to some authors (Kellenberger, 2011). In some cases, the deployment of remote-controlled weapons or robots may result in fewer accidental civilian casualties and less damage to the civilian population as compared to the situations of application of conventional weapons (Kellenberger, 2011). Both research and practice demonstrate that artificial intelligence in pattern recognition is superior to humans in both speed and quality.\(^13\) Thus, autonomous weapons systems can more accurately distinguish a combatant from a non-combatant by visual characteristics. Intelligent systems are

---

already able to detect emotions and hostility by facial expressions (Wang, 2020) and to distinguish people even with a partially covered face. All of the facts mentioned above allow proponents of the application of autonomous weapons systems to argue that the principle of distinction between civilians and combatants will be less likely to be violated with the more extensive application of such systems.

Despite promising results in the development of cognitive abilities of the machines, there are still reasons for concerns. First, the tactics of warfare have changed. Wars involving insurgent or terrorist groups have now become more common. Such units often mix with civilian population, do not wear insignia, and form temporary military groups that are assembled for a specific task. Such groups do everything possible not to be distinguished from the civilian population. Second, some research demonstrates that despite having better results in pattern recognition, emotion detection, etc., autonomous machines can still be deliberately misled (Eykholt, 2018). This type of deliberate attack is called “adversarial example” and is essentially a small change in that input data that leads to the unpredictable behavior of intelligent systems. For example, in a well-known pattern recognition experiment, a photo with a panda was recognized correctly by an intelligent software piece with the 57.7 percent confidence. After the researchers have applied some invisible noise to the original image (just 0.04 percent of the total number of the pixels in the image), it was recognized by the deep neural network as a gibbon with the 99.3 percent confidence. Although scientists have made considerable efforts to solve this problem (Carrara, 2018), the final solution is still far away (Papernot, 2017). Thus, the decision on each individual weapon system should be made separately, taking into account how such a system is resistant to provocations and conditions of the anti-terrorism or guerrilla warfare.

Military Necessity and Proportionality. These two principles are linked inseparably. In an armed conflict, military necessity may require the infliction of damage to the civilian population and civilian infrastructure to achieve the goal of defeating the enemy. In this case, a balance must be maintained between the intended goal and the collateral damage caused. Thus, civilian casualties should not be “excessive” in relation to the specific military advantage gained as a result of the assault. This principle is not quantifiable, which means that there is no formula or a specific proportion that justifies the achievement of a specific military result.

Multiple authors express reasonable doubts that autonomous weapons systems would be able to comply with this principle (Grimal, 2018; Asaro, 2012). To comply with this rule, the intelligent weapons system must independently assess the proportionality of the achieved results of the application of the military force with potential civilian casualties. In order for an autonomous weapons system to comply with this norm, it must have a clear understanding of when the expected damage to the civilian population will be excessive in relation to the military advantage obtained. This requires an intelligent system to understand military strategy, operational issues, and tactics (Egeland, 2016). At the moment, intelligent systems cannot independently understand such a multi-context environment with many connections and dependencies, and therefore autonomous weapons will not be able to comply with this principle without human intervention.

Although the autonomous system cannot independently understand the proportionality of the intended military result with potential civilian casualties, it can do so as a part of a joint decision-making shared with the human operator. In this case, it is the operator who will determine the boundaries of the application of autonomous weapons in advance, and it will act independently in the real world within these established boundaries. Thus, the autonomous system will not have to address a dilemma, which would be too difficult for it, and at the same time, the preliminary identification of the boundaries will not reduce its effectiveness in carrying out the combat task. An example of this could be a situation when a human operator makes a decision to apply autonomous weapons in a densely populated area by evaluating potential civilian casualties against the expected achievement of the military objective. In the case of such “human-machine” combination, autonomous weapons may well comply with the principles of “military necessity” and “proportionality.”

Prohibition on Causing Unnecessary Suffering and Humanity. These two principles are also better considered together, since they both intend to minimize the suffering caused during military operations. The right to humane treatment is absolute and applies not only to prisoners of war, but also to the civilian population of the occupied territories. The prohibition of unnecessary suffering, in turn, implies that “prohibited to employ weapons,

projectiles and material and methods of warfare of a nature to cause superfluous injury or unnecessary suffering.”16

Both prohibitions imply that both biological and psychophysical, as well as moral characteristics of the person will be taken into account. For example, shooting at the threat of death of the fellow villagers is not associated with physical suffering or torture, but is associated with a lack of humanity, as it causes significant moral suffering. In this sense, to comply with this prohibition, it is necessary not only to understand the human essence and the aspects that cause the suffering, but also to understand the essence of humane behavior.

Many authors agree that currently and in the near future robots will not be able to understand the basics of humane behavior and the application of autonomous weapons will lead to dehumanization of armed conflicts (Szpak, 2020; Sharkey, 2010; Asaro, 2012). We should agree with these authors because the complex context of the term “humanity” is understandable only for the yet-to-be-created strong multi-skil artificial intelligence.17 Currently, artificial intelligence systems can only perform narrow specific tasks, and therefore it is difficult for them to make choices based on the principle of “humanity.”

3. Opacity of Lethal Autonomous Weapons Systems

The problem with autonomous weapons is also that it is difficult to assess whether a particular system can comply with the principles and rules of international humanitarian law in combat before the first violations of these principles are reported. Despite the fact that the development and operation of weapons is strictly controlled by the government, there is no independent public scrutiny and oversight. Thus, a system declared to comply with the principles of international humanitarian law may in practice pose a threat to the conduct of war within the given rules.

Contemporary intelligent systems, which are the core of autonomous weapons systems, are characterized by opacity, which can be divided into three types. The first type is legal opacity, when an algorithm or autonomous system is protected by law. In the case of autonomous systems, access to their weapon devices is restricted not only by intellectual property law and trade secret law (Wexler, 2017), but also by laws protecting military and strategic secrets. Access to any device of such a system by an unauthorized person can be considered as espionage, and corporations engaged in its development, even if they wanted, could not publish data, which is classified.

The other two types of opacity are related to the technical complexity of the algorithm or of the system, which makes the decisions. First, the system can be difficult to understand without the required knowledge in mathematics and computer science (Burrell, 2016). Thus, even if the materials that explain the principles of the autonomous weapons system get published, no one outside the immediate circle of experts would be able to understand them. Second, the system can be so complex that understanding it can be beyond human capabilities (Burrell, 2016). This problem is often referred to as a “Black Box problem”, when even the developer does not fully understand what is happening inside the system. Moreover, the authors agree that “deciphering the black box has become exponentially harder and more urgent. The technology itself has exploded in complexity and application.” (Castelvecchi, 2016) In other words, the problem of opacity is becoming more acute every day.

It turns out that the question of compliance of autonomous weapons systems with international humanitarian law can be considered if sufficient experience of using such weapons in real conditions is accumulated. Obviously, an autonomous system designed to function in complex field conditions must be quite complex from the point of view of the device. Thus, even in the case of a long test in a real battle, the system would not be immune to unpredictable behavior due to the high complexity of its device.

4. Conclusion

To summarize, it would be incorrect to say that autonomous weapons systems do not comply with the principles of the international humanitarian law in general. Some principles, such as humanity, are not really intended to be fulfilled by modern machines, but that does not mean that the application of the autonomous weapons is impossible in the framework established by the international humanitarian law. More likely, it means that human operator should be playing a more significant role in the in the application of such weapons in the foreseeable future. At the same time, multiple authors share opinion that the principle of distinction between civilians and

16 Article 35, Protocol I Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, of 8 June 1977

17 Judgements on when this kind of AI would be created vary across the experts. Bill Gates for instance, believes that this will happen in several decades https://www.theguardian.com/technology/2015/jan/29/artificial-intelligence-strong-concern-bill-gates
combatants can be executed by autonomous weapons even better than by the weapons under human control. This assumption however does not imply that autonomous weapons can be used without any misgivings.

The author assumes that it is important not to prohibit autonomous weapons, because they do not comply with the principles of international humanitarian law, but to develop rules for their application and for human participation in their functioning. Literature review in the field demonstrates that oftentimes scholars do not take into account multiple other factors, ranging from the progress in the artificial intelligence technologies to the changes in the conditions at the battlefield. Weapons that can hit the targets without any human intervention have already been applied for a long time but never before had they had such an independence as they do now. To develop non-controversial and effective system of rules, which could be applied to the autonomous weapons there is a need to bring together experts from the different fields ranging from law to technology. Only interdisciplinary approach can help consider all factors required. In each particular case considered, the issue of utmost importance is how the autonomous weapons system makes its decisions. A significant challenge to the development of such rules is the opacity of the autonomous weapons systems, if they are perceived as complex intelligent computer systems.

References


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).