Autonomous Weapon Systems and Ethical Issues. A Focus on Targeted Killings

Cybersecurity and Cybercrime course. February, 2021
Irene Parodi

Abstract

Artificial Intelligence (AI) is a new technology with many applications. Its development would eventually change the world technologies in all domains. The military field is highly interested in implementing AI technology. Machines' ability to operate autonomously would develop a new way of using weapon systems. However, there are huge uncertainties about the increment of autonomy of weapon systems. Governments are trying to regulate it. Unmanned Aerial Vehicles (UAVs) – armed drones that are capable of conducting an operation completely autonomously – are quite discussed. They have multiple ethical and moral implications with regards to the International Humanitarian Law (IHL).

Introduction

Military is driven by the need of change. It is always on the pursuit of faster, better, and stronger weapons or technologies. Artificial Intelligence (AI) has advanced quickly in the last few years, and it has the potential to impact all the domains and the level of warfare.[1] There is not a specific definition of Artificial Intelligence, but it can be said that "AI is the capability of a computer system to perform tasks that normally require human intelligence, such as visual perception, speech recognition and decision-making".[2]

Though AI has been developing since 1956, since 2010 the interest in the field started to increase. This occurred particularly because of three important technological developments, which accelerated AI evolution: the availability of "big data" sources, the progress in machine learning approaches, and the increases in the computer processing power. AI technology has important and useful applications (in both civil and military) that can enhance human life. Yet, it can be an issue because of the level of autonomy that this technology can reach, particularly in the military field. The human-machine relationship is another important issue that has also legal implications involving the international community and the Law of War.[3]

This work tries to analyse the debate on ethical and legal implication of AI in the battlefield. In the first part, it will address the relationships between humans and machines and how the international community is trying to address the issue in legal and ethical terms. Then, a second section will explain autonomy in weapon systems in different domains accordingly to 2017 Stockholm International Peace Research Institute dataset. Finally, it will underline the implications of the target detection and targeted killing from an ethical perspective.

[1] Gloria Shkurti Özdemir, "Artificial Intelligence application in the military: the case of United State and China", in *SETA*, n. 51 (2019), 8. https://www.setav.org/en/analysis-artificial-intelligence-application-in-the-military-the-case-of-united-states-and-china/. (05/02/2021).
[2] M. L. Cumming, "Artificial Intelligence and the Future of Warfare", in *Chatham House* (2017), 2. https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings.pdf . (05/02/2021).
[3] Shkurti Özdemir, "Artificial Intelligence application in the military", 9.

Human-machine relationships and their ethical and legal implications

The first step is to understand the differences between an automated system and an autonomous system. The former has a rule-based structure that organise the information in an if-then-else reasoning. It is a deterministic system; thus, for each input the output will always be the same, unless there are malfunctions. Instead, an autonomous system analyses a set of input in a probabilistic prospective. It makes guesses about the best output from the given inputs, and it produces a set of behaviours, so the outputs are not always the same. AI is programming to emulate the human sequence that generally follows the perception-cognition-action information processing loop. As humans perceives something around them, they think about what to do and then act. On the contrary, the computer processes the information coming from the world around it through optimization and verification algorithms, and then acts. The main differences between humans and machines rely in how AI processes the input coming from the world. Autonomous systems – that interact with a dynamic environment – need to create a world model that has to be continually updated. This means that the world is sensed through cameras, microphones, and tactile sensors, and then reconstructed in order to give the machine an accurate view of the world before it makes decisions. The more accurate and update the model, the better the system works.[4]

Applying these concepts to the military field, autonomous weapon systems use this technology. A weapon system is the combination of a weapon and the items associated with its employment. An autonomous weapon system is identified as a weapon system that, once activated, can target objectives without any further human interference. Before analysing the degree of autonomy, it is important to understand the degree of capability that autonomous systems can have compared to humans.[5]

Automated and autonomous systems can address a different degree of uncertainty that can occur in a situation of risk – such as weapons release. An automated system may be useful for tasks requiring *skills-based behaviours*, that are basically sensory-motor reactions. An example can be flying an aircraft. But, as the cognitive continuum increases in complexity, the need for *rules-based behaviours* arises. Procedures can

[4] Cumming, "Artificial Intelligence and the Future of Warfare", 3, 4.
[5] Michael N. Schmitt, Jeffrey S. Thurnher, ""Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict", in *Harvard National Security Journal*, vol. 4 (2013), 234,235. https://harvardnsj.org/wp-content/uploads/sites/13/2013/01/Vol-4-Schmitt-Thurnher.pdf. (05/02/2021).

help to manage the complexity of various tasks, and rules-based reasoning may assist the decision-making process in determining possible courses of actions. However, with a high level of uncertainty, it is difficult to understand which set of rules need to be applied. In this case, automated systems are not capable of addressing the situation. Indeed, *knowledge-based reasoning* is needed when an established set of rules does not match the situation. Although autonomous systems use cognitive representations of the external world, in time-critical situations – that are by definition ambiguous and vague – algorithms may not be able to understand and achieve feasible solutions. *Expert behaviour* reasoning is necessary to address uncertain scenarios and to find a suitable solution. "The key question for any autonomous system engaged in a safety-critical task (such as weapons release) is whether that system can resolve ambiguity in order to achieve acceptable outcomes […] The power of human induction – i.e., the ability to form general rules from specific pieces of information – is critical in a situation that requires both visual and moral judgment and reasoning".[6] Currently, experts believe that decades need to pass until an autonomous system reaches this kind of capability. Nevertheless, the debate about AI autonomy is an important issue in the human-machine relationship. [7]

There are three different degrees of relationship between humans and machines, regarding human's position in the decision-making process. When the human is in the loop, the machine has control on the environment, but the human takes the final decision. This relationship is defined as a *semi-autonomous system*. An example may be a "fire and forget" missile on an aircraft, it is locked to a target identified by the pilot and then it attacks it without any further human involvement.[8] If the human is on the loop, it is a *supervised autonomous system*. The machine can act and decide on its own, however the human can observe the behaviours of AI and intervene if necessary. Examples of human-supervised autonomous systems are the US Aegis at sea and the Patriot on land – both designed to defend against short notice missile attacks – or the Israel's Iron Dome. Others are called automatic weapon defence systems, they are systems that respond nearly automatically when they detect incoming threats. Their main characteristic is that they are fixed. An example is the "close-in weapon system" or "Sea Whiz", which is used for point-defence of warship against incoming missiles, it can detect and immediately attack inbound missiles.[9] In

[6] Cumming, "Artificial Intelligence and the Future of Warfare", 5-7.
[7] Shkurti Özdemir, "Artificial Intelligence application in the military", 9.
[8] Schmitt, "Out of the Loop", 236
[9] Ivi., 235, 236.

the third case, the human is out of the loop and the machine is identified as a *fully autonomous system*. Here the human does not have any control over AI, acting and deciding by itself. The latter case has not yet been reached in the military field. The debates among experts about how much autonomy should be given to weapons using AI is particularly concerned about the so-called Lethal Autonomous Weapon Systems (LAWS).[10]

Indeed, there are different types of existing autonomous weapon systems. Examples are: air defence systems; active protection systems, which shield armoured vehicles by identifying and intercepting anti-tank missiles and rockets; guided munitions, that identify and engage targets that are not in sight of the attacking aircraft; robotic sentries, which have tasks of surveillance; loitering munitions, that can overfly an assigned area in search of targets and to bomb and destroy. The technological reality and prospect of autonomous weapon systems raise an ethical and legal issue: "Is it permissible to let a robotic system unleash destructive force and take attendant life-or-death decisions without any human intervention?".[11] States started to discuss the normative framework to govern developments, deployments, and uses of autonomous weapon systems. Diplomats have dialogued on this topic since 2014 at the United Nation in Geneve, within the institution of the Convention on Certain Conventional Weapons (CCW). Its main purpose is to restrict and possibly ban weapons that can cause unjustifiable or unnecessary suffering to combatant or affect civilians indiscriminately. Furthermore, the institution created a Group of Governmental Experts (GGE) on lethal autonomous weapon systems, which is the main annual forum where autonomy in weapon systems is debated internationally. An important report written in 2013 by Christof Heyns[12] explains the main ethical and legal concerns about autonomous weapon systems, that can be summarized in four points. First, the compliance with International Humanitarian Law (IHL) requires capabilities that only humans can have. Achieving situational awareness and formulating appropriate judgements in unstructured warfare scenarios, for instance. The second point is about the responsibility ascription problem: by removing human operators from the decision-making process, it would hinder responsibility ascriptions in case of errors. Third, robots' decisions concerning human life would be an affront

[10] Shkurti Özdemir, "Artificial Intelligence application in the military", 10.
[11] Daniele Amoroso, Guglielmo Tamburini, "Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues", in *Curr Robot Rep*, n. 1 (2020), 189. https://doi.org/10.1007/s43154-020-00024-3. (05/02/2021).
[12] In August 2010 Christof Heyns was appointed as United Nations Special Rapporteur on extrajudicial, summary or arbitrary executions.
https://www.ohchr.org/en/issues/executions/pages/christofheyns.aspx. (05/02/2021).

to human dignity. The fourth concerns the increasing risk for peace and international stability. Wars would be easier to wage reducing the number of soldiers involved, having unpredictable interaction between autonomous weapon systems and their harmful outcomes, and accelerating the rhythm of war beyond human reactive abilities. Because of these issues, a meaningful human control over autonomous weapon systems should be retained.[13]

Specific policy on the use of autonomous weapon systems were developed, one of them is the *boxed autonomy policy*. It assigns to humans the role of creatin an operational box with which constraining the autonomy of the system, constituted by fixed time period, geographic borders and a predefined target parameter. Another is the *denied autonomy policy*; it "rules out any autonomy whatsoever for weapon systems in the critical targeting function and therefore embodies a most restrictive interpretation of [meaningful human control]".[14] The latter is the *supervised autonomy policy*, which is in the middle between boxed and denied autonomy policies, because it requests humans on the loop. The former policy befits targeting situations, but it is criticised because it does not fit in dynamic targeting situations because they are not known in advance (unanticipated targets), or they are not localizable in advance (unplanned targets), so that they require changes during operations. Instead, the denied policy fulfils the normative which provides for human control as fail-safe actor[15], accountability attractor[16], and moral agency enactor[17]. However, it has been criticized because it sets an upscale threshold for machine's autonomy leading to abandon the use of certain weapons that have long been considered acceptable. An example is represented by all the systems classified as Sense and React to Military Objects (SARMO) weapons. They are air defensive systems that autonomously detect, track, and target incoming projectiles. They are used in highly predicable environment with scarce-risk civilian harm. Moreover, they are fixed and have constant human control and monitoring for rapid shutdown. Yet, these systems exceed the denied policy's threshold. Rather, the supervised autonomy policy occupies a middle ground between the denied policy and the boxed one, as it

[13] Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 189.
[14] Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 190.
[15] Contributing to prevent a malfunctioning of the weapon from resulting in excessive collateral damages; Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 189.
[16] "to secure the legal conditions for responsibility ascription in case a weapon follows a course of action that is in breach of international law"; Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 189.
[17] "by ensuring that decisions affecting the life, physical integrity, and property of people (including combatants) in- volved in armed conflicts are not taken by non-moral artificial agents"; Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 189.

requires humans to be on the loop. Although it may be used to defence installations and platforms from attacks – providing that they do not select humans as targets – supervised autonomy would not prevent faster and faster offensive autonomous weapon systems to be developed. Humans would supervise decisions taken at superhuman speed, while leaving the illusion that the human control is still crucial.[18] Analysing the diversity of autonomous weapon systems and their range of application, experts agree on the fact that a uniformed policy is not adequate to address such meaningful human control issue. Noel Sharkey[19] proposed an organization on levels about the autonomous weapon systems critical target selection and engagement functions.[20] Level 1 of autonomy is represented by a human that selects and engages with targets and initiates any attacks. Level 2 occurs when a program suggests alternative targets and a human chooses which one to attack. Then, level 3 is when a program selects targets and a human have to approve the selection before the attack. Level 4 explains the situation in which a program selects and engages targets, but it is supervised by a human that can override machine's choices and abort the attack. Finally, level 5 happens when a program selects targets and stats an attack according to mission's goal – defined at the planning stage – without any further human involvement. Connecting these levels with the policies analysed previously, level 5 basically corresponds to the boxed autonomy policy, where human control is exerted only at the planning stage by human commander. Rather, level 4 corresponds to the supervised autonomy policy. Humans have to be advised against the possibility of automation bias – the human propensity to overtrust machine decision-making process and outcomes – risks and increasing marginalization of human control. However, in certain operational conditions, it may be an acceptable level of human control. At level 3, human operators and weapon systems have the same control's privileges on critical targeting functions. Though, human deliberative role is limited to approve or reject machine's decisions. As for level 4, there is the possibility of automation bias, and it should not be adopted as a general policy. The last two levels correspond to shared control policies. Autonomy of the weapon system is totally absent, as in level 1, or it has just the role of adviser and decision support system for human deliberation, like level 2. Applying these

[18] Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 189.
[19] Noel Sharkey is a professor of AI and Robotics and of Public Engagement at the University of Sheffield. http://noelsharkey.com/. (05/02/2021).
[20] Noel Sharkey has proposed the organization on levels starting from the autonomous levels introduced in connection with surgical robots, unmanned commercial ships, and automated driving.

levels, it has to be aware of few autonomous weapon systems that have long been considered acceptable in warfare operations, as SARMO systems.[21]

The current state of autonomy in weapon systems

[21] Amoroso, "Autonomous Weapons Systems and Meaningful Human Control", 190, 191.

In a study of the Stockholm International Peace Research Institute (SIPRI) in 2017, scholars have created a dataset to obtain an overview on the level of autonomy in existing weapon systems. The dataset's report explains in details capabilities and levels of autonomy of the already-existing autonomous weapon systems. Systems are divided into five capability areas: mobility, targeting, intelligence, interoperability, and health management; and they are analysed answering to two questions:

1. What can autonomous systems do and not do autonomously?
2. What is the nature of the human-machine command-and-control relationship when the systems execute the relevant capability autonomously?

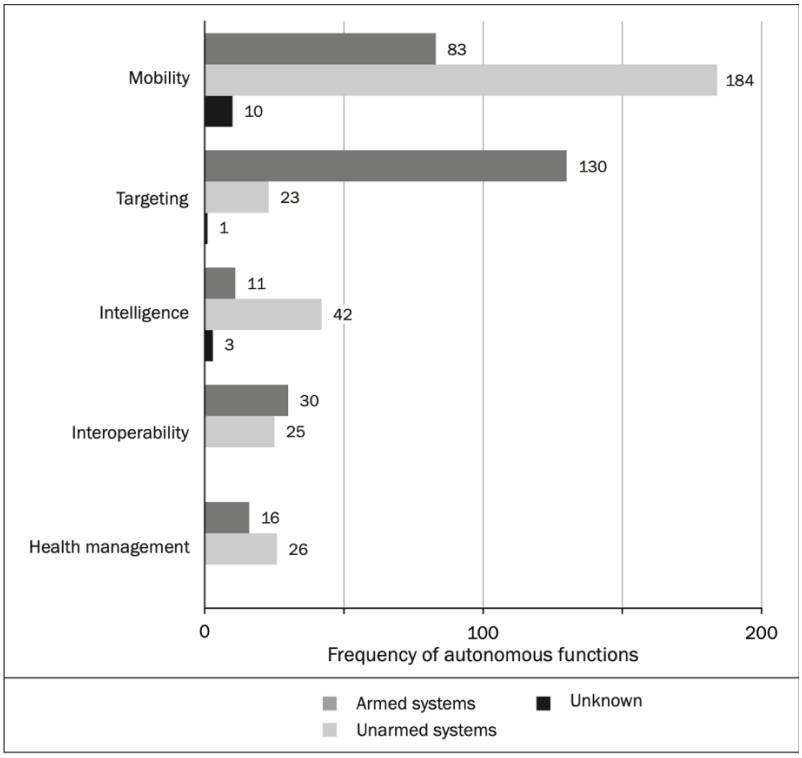


Figure 1 – Vincent Boulanin, Maaike Verbruggen, *Mapping the Development of Autonomy in Weapon Systems*, (Stockholm: SIPRI, 2017), 21

Autonomy for mobility

Mobility is the predominant area in military systems. Related autonomous functions change greatly in terms of capability, the most developed in exiting systems are homing/follow-me, autonomous navigation, and take-off and landing. With regard to the formers, homing is the system's capability to find and track its target. It is usually applicated in missiles technology. Instead, follow-me is the ability of an unmanned system to follow another system or a soldier. In both cases systems identify and track a target through a radar, acoustic or electromagnetic signal, or an electro-optical or infrared signature. The signal followed is pre-programmed and stored in

the system's memory; thus, existing systems have not the ability to pick up new signals once activated and deployed. As regards the autonomous navigation, it ensures that the system can determine its position and plan a route on its own. The systems that have this capability are not entirely autonomous because they rely on "waypoint navigation" – namely the system merely follows a series of geodetic coordinates set by a human operator. Few systems can plan a route by themselves but the general indicators, as speed, altitude, and mission objective, are set by humans. The autonomy of the systems is also conditioned by the domain in which it has to operate. The land domain, especially in a military context, exposes the system to greater complexity than air and sea. Indeed, existing ground systems with this capability rely on pre-mapping, therefore they can navigate autonomously in areas known in advance. This restriction – due to the state of art vision-based guidance technology that is not sophisticated enough – restricts the type of missions that systems can perform autonomously, such as perimeter surveillance and logistics. Moreover, these systems can operate only in non-adversarial condition, because they do not have sufficient perception or decision-making capabilities to address adversaries that might seek to defeat their guidance systems. Indeed, one of the key vulnerabilities of these systems is that the rely on Global Positioning System (GPS) guidance, making them vulnerable to jamming technologies and cyber-attacks. Concerning the latter abilities, it would be more appropriate to speak about automatic take-off and landing since these systems follow a very strict set of predefined rules, with the entire procedure operated by an algorithm. Anyway, machines have reached the point where they outperform humans in terms of precision and reliability.[22]

The relationship between human-machine command-and-control varies from one system to another, however, such systems are usually used to complement remote control. Indeed, autonomous navigation, homing and follow-me are used mostly to discharge humans from operating the system during phases of the mission where humans' capabilities are not essential. Rather, Autonomous take-off and landing capabilities are aimed at reducing the risk of accident. Nevertheless, there are three different types of systems that once they are launched, they can navigate in complete autonomy. The first category is aerial, land, and maritime systems that are deployed to conduct pre-programmed manoeuvres in known and semi-structured

[22] Vincent Boulanin, Maaike Verbruggen, *Mapping the Development of Autonomy in Weapon Systems,* (Stockholm: SIPRI, 2017), 21-23. https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf (05/02/2021)

environments. The second sees unmanned systems conducting long-term missions in an environment where communications are difficult. And the third category includes missile systems and unmanned combat systems striking targets in communication-denied environments.[23]

Autonomy for targeting

SIPRI found that the autonomy of targeting is used in at least 154 systems to support at the tactical level targeting process from identification, tracking, prioritization, and selection of targets to, in a few cases, target engagement. Automatic or automated target recognition software (ATR software) was invented in 1970s, and it relies on the principle of pattern recognition. This software is programmed to identify target types based on predefined target signatures. The target has to match these target signatures that are stored in software's identification target library. When multiple targets are identified by the software, it can prioritise among them based on strict predefined parameters that are specific to each operational situation. Automatic target recognition software has not deliberative autonomy. It can only identify and fire upon target types that are already determined by the human operator, and it has no capability to learn new target signatures once deployed. Therefore, there is an open debate over whether it is appropriate to discuss autonomy in relation to target recognition. In the majority of cases, the automatic target recognition software can only target large and well-defined military objects, such as tank, aircraft, submarines, and radar. The software uses simple criteria on the nature of the target to identify it. For example, tanks are recognized according to their shape and height, missiles are typically detected by their speed, radio-frequency emission, while submarines are identified based on their acoustic signature. Robotic sentry weapons[24] are the only type of system that uses the automatic target recognition software to detect human targets, but it cannot distinguish between soldier and civilians. Moreover, the systems using Automatic Target Recognition (ATR) software are highly sensitive to variations in the environment, so they cannot be used safely in all circumstances. ATR systems are unable to evaluate a situation to ensure that an attack complies with the

[23] Vincent Boulanin, Maaike Verbruggen, *Mapping the Development of Autonomy in Weapon Systems,* (Stockholm: SIPRI, 2017), 23, 24. https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf (05/02/2021)

[24] Robotic sentry weapons are gun turrets that can automatically detect, track and engage targets. They can be used as stationary weapons or be mounted on various vehicles; Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 44.

obligations of distinction (between combatant and civilians), *proportionality* (prohibits excessive civilian harm), and *precaution* (to avoid and minimize civilian harm).[25] They apply a rudimental principle of distinction ignoring everything that does not match the predefined target. Indeed, they are unable to understand if the target has surrendered or is *hors de combat* for some reason, neither they can detect if the target is surrounded by civilians and civilian objects, which would be an essential requirement to apply the principles of proportionality and precaution. There are two main problems that obstacle the developments of this technology. The first is the lack of training and test data. Indeed, target recognition algorithms have to be trained on a large set of data related to the mission scenario, in order to expose the algorithm to any variable that it will be expected to handle. Thus, the dataset has to include information about different possible background or the weather conditions. The challenge here is not only the training of the algorithm but to find the data needed about the target, particularly if it is human. The second main problem is about the concerns with regard to predictability that rise about the machine-learning techniques, such as deep learning. Actually, learning systems operate like "black boxes", the process, leading from input to output, is unknown or difficult to comprehend. This creates uncertainty about how the system could respond to an input different from the one used during the training phase. Accordingly, the use of machine learning for the development of automatic target recognition software has, so far, been restricted to experimental research.[26]

The weapon systems using ATR software are intended to operate as a decision aid in operational context where the presence of civilians and civilian objects is unlikely. Moreover, the human-machine command-and-control relationship is based on human in the loop in nearly one-third of the systems identified by SIPRI dataset. ATR software is mainly used when the target is beyond the visual range of human operators, or it is moving too fast for human capabilities. In these situations, humans retain the decision to engage the target. In other circumstances, weapon systems using ATR software can engage the target autonomously. Nevertheless, they are weapon systems that are intended to protect ships, ground installations, or vehicles against incoming projectiles. Usually, the human is on the loop, so the system is supervised, and it has different modes of engagement. Autonomy is used only in

[25] Rules and principles of international law in combat. "New SIPRI Reflection Film on limits on autonomy in weapon systems", in *SIPRI*, March 27, 2020. https://www.sipri.org/news/2020/new-sipri-reflection-film-limits-autonomy-weapons-systems. (05/02/2021)
[26] Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 24-26.

circumstance when the time would be to short for humans to be able to respond. Furthermore, the automated target recognition technology can also be found in unarmed military systems. This technology is typically used for intelligence, surveillance, and reconnaissance missions with the task of feeding target information to another weapon system, or to a command-and-control chain. An example can be the MQ-4C Triton, that can detect and classify targets using advanced image and radar returning recognition software. It can be pre-programmed to zoom on particular target types and relays images that can be of particularly interesting for human operators.[27]

Autonomy of intelligence

Military systems can use autonomy to collect and process different types of information that may be particularly relevant from the command-and-control perspective. The existing weapon systems use information processing taking the form of automated detection of simple objects or events matching specific predefined criteria. They can be employed for detection of explosive devices. This capability is usually found in robotic weapon systems designed for bomb ordnance disposal. They use different sensors depending on the type of the explosive that they are supposed to detect. These systems are usually managed by human operators remotely. However, few recent systems, as the Counter IED and Mine Suite (CIMS), can execute the entire process autonomously from detection to destruction. Another example of information processing task is the detection of perimeter intrusion – has in robotic platforms. It has the aim to secure known perimeters, such as military bases, borders, or warehouses. The detection process is relatively unsophisticated; indeed, the systems are programmed to detect movements or unauthorized presences using a set of sensors. An example is the Mobile Detection Assessment and Response System (MDARS), that can autonomously detect intrusions using forward-looking infrared, radar, light detection and ranging sensors, and radio-frequency identification. Another goal of such military system is detecting the location of gunfire or other weapon fire. In this scenario, the system is used to improve protection of human forces on the ground. The RedOWL optional sensor of the 510 Packbot, that is a ground robot, is an example of system used in this circumstance. It locates sniper and mortars, nevertheless, the system does not attack these targets, it simply communicates information about the direction and range to forces on the

[27] Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 26, 27.

ground. A further capability that military systems can have is the detection of objects of interest in intelligence, surveillance, and reconnaissance (ISR) missions. A limitation of most unmanned systems used for ISR missions is that they do not have on-board ability to analyse the intelligence information they collect. Human analysts off-board have to monitored and assessed the collected data. Hence, new-generation unmanned systems will include image data processing software that permits systems to autonomously process information of interest and convey that information to humans for disambiguation. An example of these new-generation unmanned systems is the ScanEagle, it is a small unmanned aircraft system (UAS) which is able to detect autonomously objects of interest on the sea surface. However, the system can only differentiate between water and non-water. Speaking about intelligence data generation, three specific functions can be identified. The first is mapping generation, the second is threat assessment, and the third is about big data analytics. The former is particularly common for underwater systems and is emerging in the last generation of reconnaissance aerial systems. It is the ability to autonomously generate details about the environment. An example is Shield AI, a tactical UAS that can generate three-dimensional maps and it requires no human piloting or GPS. The second function concerns defensive systems that are programmed to evaluate the level of risk according to predefined criteria. Israel's Iron Dome missile defence system can gauge where an incoming missile will detonate and suggest countermeasures. The latter function is the use of big data analytics for pattern recognition in intelligence data. This capability does not take place on-board weapon systems due to the high demand of computer power. Machine-learning algorithms allow military commands to find correlations in large and heterogeneous sets of intelligence data. An example is the algorithm used by the US to search the Global System for Mobile communication metadata of 55 million mobile phone users in Pakistan in order to track messages between al-Qaeda members.[28]

Analysing the human-machine command-and-control relationship, the majority of the function of these autonomous military systems are not safety critical; thus, they generally do not require direct supervision. However, detection of explosive devices and threat assessment in defensive systems are exceptions because they are related to the use of kinetic force.[29]

[28] Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 27-29.
[29] Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 29.

Autonomy for interoperability

Interoperability is the capacity of military equipment and troops to operate in conjunction with each other. There are autonomous military systems that have the ability to execute tasks or missions coopering with other systems (machine-machine teaming) or with combat troops (human-machine teaming). The two options of teaming have to be analysed separately.

Machine-machine teaming can be developed in different forms, among which the most basic is information sharing. Systems are connected and can communicate to share sensors or information, but each has its own goals. Instead, collaborative autonomy is a more complex model of interoperability, multiple systems are able to coordinate their action in order to achieve a common goal. The software architecture – that is employed in this kind of interoperability – has to be able in commanding and controlling the action of the "collective system" or the "system of systems" as a whole. The system of the systems can be formed by a "swarm" of identities or an heterogenous systems, for example a mix of UASs and unmanned surface systems. In the former case, software's architecture is designed to govern a collective behaviour in order to achieve effects that a single unit cannot reach by itself. In the latter case, software's architecture predetermines the specific role of each unite within the large group. The collaborative autonomy capability is a structure that is still under development. It has been tested in different scenarios. For example, the coordinated mobility is a capability that an increasing number of systems under development staring to use. As in the autonomous navigation, the main technical difficulty is the nature of the environment, particularly in land domain. Another example of interoperability is the development of coordinated ISR operations over a large geographical area. That could see involved small, low-cost UASs for ISR missions, for example a swarm of Perdix drones. A further case in which this technology can be applied is perimeter surveillance and protection, in anti-access/area-denial manoeuvres. But it can also be employed in distributed attacks. The development of a control architecture in which weapon systems could automatically distribute targets among themselves is being investigated. In this scenario, a higher-level UAS, which could act as the central authority, would identify a target and then hand it over to a lower-level UAS.[30]

[30] Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 30, 31.

The human-machine command-and-control relationship in machine-machine teaming is a nascent area of research. Indeed, system of systems can be supervised and controlled in two ways. Firstly, it can be supervised by *centralized control*, in which the system of systems is controlled by a human operator who sends commands to a system which then distributes them to the rest of the network. Secondly, it can be supervised by a *decentralized control*, in which the human operator gives instructions and commands to the system of systems as a whole. Both controls have pros and cons. In the specific case of swarms, experts prefer to use a decentralized control, because with this method the swarm has more resilience in case of individual unit loss.[31]

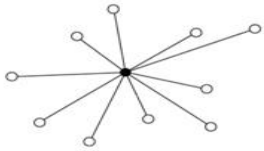
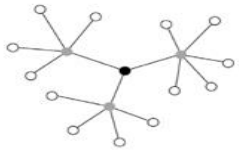
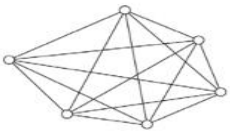
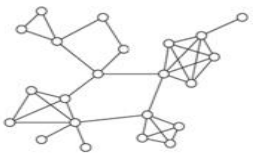
Command-and-control structure	Description	Pros	Cons
Centralized control (centralized) 	Individual elements communicate with a centralized planner that coordinates all tasks	Can find an optimal or 'good enough' solution quickly	Require high bandwidth to transmit data to centralized sources and send instructions back to the swarm Vulnerable to communication disruption
Hierarchical coordination (centralized) 	Individual elements are controlled by 'squad'-level agents that are, in turn, controlled by a higher-level controller		
Coordination by consensus (decentralized) 	All elements of the systems communicate with one another and use 'voting' or auction-based methods to converge to a solution	Can find solutions to complex problems Can work with low bandwidth between the different elements	Finding the optimal solution may take multiple iterations, and, hence, time
Emergent coordination (decentralized) 	Coordination arises by individual swarm elements reacting to one another, like an animal swarm	Can work with no direct communication between elements, hence immune to direct communication jamming	

Figure 2 – "Command-and-control structure for collective systems, including swarms". Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems*, 32.

In the human-machine teaming technological developments are still immature. AI limitations do not allow autonomous systems to have sufficient situational awareness and decision-making capacity to really work in peer with humans. Human-machine teaming remains an experimental capability. A technical obstacle to this teaming is

[31] Ivi., 32, 33.

the limitation of the existing human-machine communication that is still relying to visual interfaces. Instead, the use of voice command-and-control would give a real possibility of communicating between human and machine in operational and tactical situations.[32]

Autonomy for the health management of systems

The health management of systems is a less common application area of autonomy in weapon systems. Existing systems can self-recharging and self-refuelling (quite difficult for an UAS), detect and diagnose system's faults and failures. Self-maintenance and self-repair remain experimental capability. All the systems that include health management capabilities are remotely controlled or supervised by human operators. The use of autonomy for these abilities aims to relief human operators' tasks.[33]

Target detection and ethical issues

Analysing the autonomy in targeting it was shown that automatic target recognition software were developed in the 1970s. The most known and used robots for targeting detection are drones. Before analysing this capability, it has to be specified the

[32] Vincent Boulanin, *Mapping the Development of Autonomy in Weapon Systems,* 33, 34.
[33] Ivi., 35.

difference between drones and Unmanned Aerial Systems (UAVs). UAV is a category that includes drones and Unmanned Aerial Vehicles (UAVs). Drones are remotely autonomously guided aircraft, and UAV are drones characterized by their autonomous flight capability and the ability to operate over long distances with a secure live feed transmission. Their control can be classified in three main categories. The *Remote Pilot Control*, known as operator static automation, in which all decisions are made by a human operator. The *Remote Supervised Control*, or adaptive automation, in which the drone has the ability to launch and carry out a given mission process independently, allowing human intervention, if needed. The *Full Autonomous Control*, or system static automation, in which drones can take any decision for a successful mission completion, without human intervention. Unmanned aerial vehicles, or drone architecture usually consists in three different parts: Unmanned Aircraft (UmA), Ground Control Station (GCS), and Communication Data-Link (CDL). The first is the central drone's processing unit, the second is based on an On-Land Facility (OLF), providing human operators with the capabilities to monitor and control UAVs during the operation. The latter is composed by wireless links that control information flows between the UmA and the GCS. [34]

During the Cold War, aircrafts have been slowly substituted by drones for intelligence gathering, particularly after that Francis Gary Powers was shot down over the Soviet Union territory in 1960. The first prototype of reconnaissance drone was created by Ryan Aeronautical Company, a US government contractor. It was called Firebee target drone, that was developed in a reconnaissance plane known as Lightning Bug. Tagboard was the second developed by Lockheed. Both Tagboard and Lightning Bug usually flew pre-programmed and took photographs. They were used to identify and map enemy missile sites.

[34] Jean-Paul Yaacoub, Hassan Noura, Ola Salman, Ali Chehab, "Security analysis of drones systems: attacks, limitations, and recommendations" in *Internet of Things*, May 8, 2020, 3, 8. 10.1016/j.iot.2020.100218. (05/02/2021).

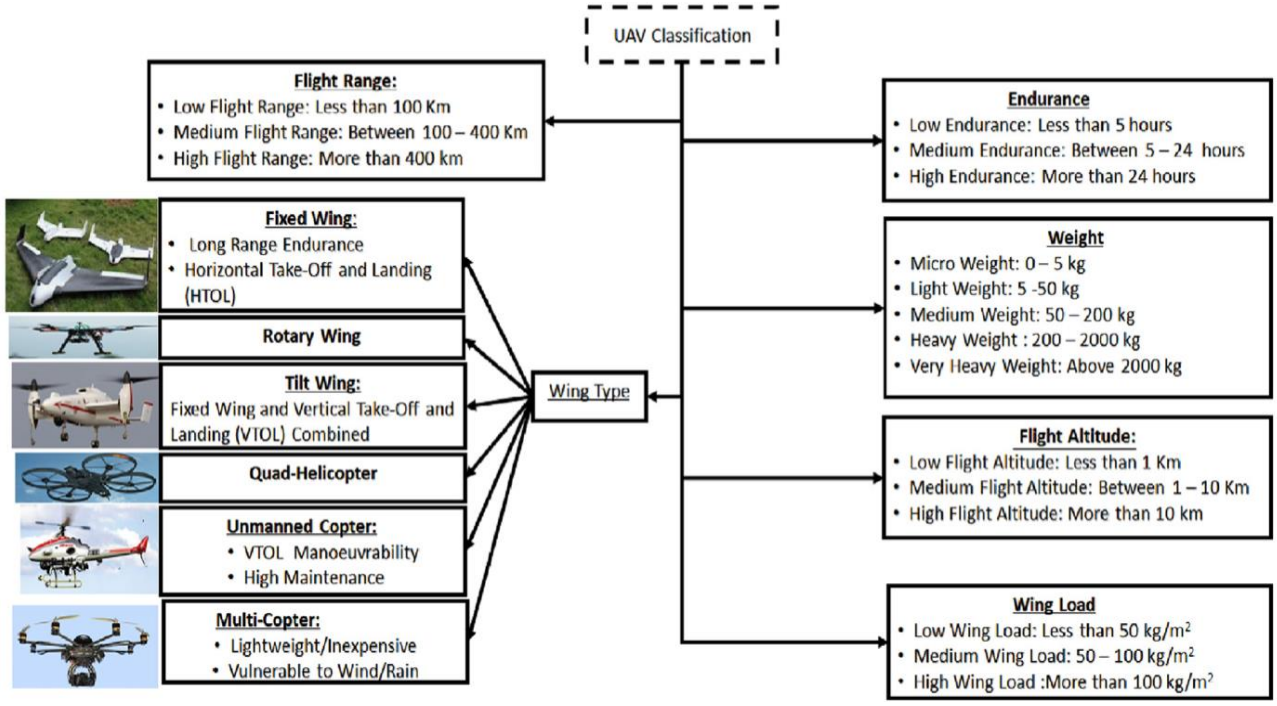


Figure 3 - UAV classification. Jean-Paul Yaacoub, "Security analysis of drones systems", 9.

During the Vietnam War, drones were used for approximate 3.000 missions and their mission expanded, they were used to take pictures of targets and dispersing propaganda materials and electronic listening devices. However, after the Vietnam War, the first real upgrade of drones took place during the Kosovo air war. Although the ISP capability of drone as well as its data transmission capability were increased, the role of air power and intelligence gathering began to shift dynamic targeting, and the necessity of transmission in real time became clear. So, in the mid-1990s the US Defence Advanced Research Projects Agency (DARPA) establish a new program on high-altitude unmanned aerial vehicles that brought about the Predator and the Global Hawk, two of the primary larger drone systems used today. These programs were put to the test during the US and NATO airstrikes over Kosovo in the later 1990s, and they included the Predator, the Pioneer and a German drone. Though the benefits that drones brought about in targeting, reconnaissance, and surveillance operation were clear, there were some challenges. Indeed, the time of targeting process was too long in same cases. At least two aircrafts, manned or unmanned, had to confirm the target, then the target's location had to be relayed to the bomber aircraft or to a mission control center. Thus, the next step was arming the Predator in 2001 in order to contract the killing-chain. Moreover, the drones started to deploy real-time video to anywhere. The last big step in the development of drone capabilities followed the 9/11 events, during military operations in Afghanistan and then Iraq. Drones were employed in both counterterrorism and counterinsurgency

operations thanks to their ability to monitor the movement of individuals for long periods of time and providing precision targeting capability. Moreover, in this period there was an effort to make drones more automated. "Since 9/11 the processes of surveillance and targeting started to collapse into each other, forming a single process of lethal surveillance".[35]

After the War on Terror, when armed UAVs were used for precision strikes, the moral and ethical worldwide discussion has associated armed UAVs with targeted killings. This is the reason why nowadays there is a strong aversion in arming a drone.[36] According to scholars, a targeted killing is "the use of internationally lethal violence against a prominent or culpable person or a small group of persons (the target) not in the physical custody of the agent using violence (the source)".[37] Targeted killing has always been a feature of human society, but, since the turn of the millennium, it has undergone a profound transformation involving three dimensions. First, the number of targeted killing missions has increased significantly, particularly against non-state, terrorist actors. Second, the technological revolution has transformed targeted killing tanks to the proliferation of surveillance and drone technology. Indeed, drones can monitor suspects for a long period of time and deliver deadly attacks with little risks for the operator. Third, states are slowly abandoning their policy secrecy on targeted killing.[38] The use of drones in targeted killing operation depends not only upon the technology available but also upon military commander decisions. If the commanders "see a legitimate and effective way of integrating autonomous weapons into the military-theoretical paradigm that they are applying to target killing operations".[39] According to the scholars, the value of autonomy of UAVs is greater where human supervision imposes a serious operational weight.[40]

Nowadays, fully autonomous combat drones are not a reality, but they are, anyway, becoming more and more autonomous. Indeed, drones employed in battlefield are remotely controlled or partially automated. Today's debate on ethical issues about

[35] Katherine Hall Kindervater, "The emergence of lethal surveillance", in *Security Dialogue*, Vol. 47, n. 3, 231, 232. https://www.jstor.org/stable/10.2307/26294130 (05/02/2021).
[36] Jack Watling, Nicholas Waters, "Achieving Lethal Effects by Small Unmanned Aerial Vehicles. Opportunities and Limitations", in *The RUSI journal*, 164:1 (2019), 40, 41. https://doi.org/10.1080/03071847.2019.1605017 (05/02/2021)
[37] Martin Senn, Jodok Troy, "The transformation of targeted killing and international order", in *Contemporary Security Policy*, 38:2 (2017), 186. https://doi.org/10.1080/13523260.2017.1336604 (05/02/2021)
[38] Ivi., 190.
[39] Michael Carl Haas, Sophie-Charlotte Fischer, "The evolution of targeted killing practices: Autonomous weapons, future conflict, and the international order", in *Contemporary Security Policy*, 38:2, 290. https://doi.org/10.1080/13523260.2017.1336407. (05/02/2021).
[40] Ivi., 297.

the autonomy of drones is addressing the question if the UAVs should behave more autonomously. This capability would add to the training, by machine learning algorithms, a program that would analyse the moral and legal legitimacy of a lethal attack on the target. New types of human-technology interactions in on-the-loop systems, where moral decision-making processes are distributed among humans and machine, are also addressed by the current debate on autonomy of UAVs. This is because humans have to rely more and more on the algorithms used in pattern recognition to identify suspect targets.[41]

There are different human-technology relations in drone warfare. One of them is the embodiment relations, that are crucial because the drone operator is separated from his weapon. The drone is a highly lethal weapon that allows the soldier to operate from a safe distance, thus the connection between weapon and soldier in battle is physically dissolved. However, on the other hand, there is a new quality of proximity due to the sophisticated surveillance technologies that allow the drone, and the operator, to be much closer to the target. The soldier experiences the immediate consequence of the attack more intensively. There is a new combination of physical distance and ocular proximity that can be interpreted in two opposite ways. On one hand, drone operators are able to act more ethically because they are not physically involved in the conflict and they do not suffer the stress of the battlefield. On the other hand, it was said that the proximity would privileges the view, and the implications are far more deadly. Another implication – that does not concern soldiers – is about the embodiment perceived also on the other side of the conflict. Citizens in battlefield regions have to live in the knowledge that they are constantly observed by invisible drones, thus, that a drone strike could happened anytime.[42]

Another relation among humans and machines is about the perception that operators have of the possible target. Usually, surveillance activities are a preventive operation, thus the target is followed for weeks and operators have to understand if the target is a terrorist or not. In order to do so, there is the risk of a certain "PlayStation mentality" that substitutes the real person in the world with a virtual target. Applying machine learning algorithms in an on-the-loop system, the situation becomes even more complex. Another criticized aspect of the pattern-of-life analysis is the "Kill lists". They are meta-data collected by drones and analysed by complex algorithms,

[41] Oliver Müller, "'An Eye Turned into a Weapon': a Philosophical Investigation of Remote Controlled, Automated, and Autonomous Drone Warfare", in *Philosophy & Technology*, December 15, 2020, 4, 5. https://doi.org/10.1007/s13347-020-00440-5. (05/02/2021).
[42] Ivi., 9, 10.

then interpretated by humans. Moreover, pattern recognition technologies are trained with machine learning algorithms in order to support the discrimination of targets, with regard to the discrimination of civilians and combatants in IHL. Though, aside from the fact that it is really difficult even to find a definition of "civilian" in a conflict area and it will be more and more difficult to identify enemy soldiers in the new asymmetrical wars, it is ethically highly problematic to delegate the decision of killing a person to an autonomous machine.[43] A legitimate lethal decision process has to "meet requirements that the human decision-maker involved in verifying legitimate targets and initiating lethal forces against them be allowed sufficient time to be deliberative, be suitably trained and well informed, and be held accountable and responsible".[44]

[43] Oliver Müller, "'An Eye Turned into a Weapon'", 14.
[44] Peter Asaro, "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making", in *International Review of the Red Cross*, Vol. 94, n. 886, (Summer 2012), 695. https://doi.org/10.1017/S1816383112000768 (05/02/2021).

Conclusion

On 8 April 2019, the High-Level Expert Group on AI presented the Ethics Guidelines for Trustworthy Artificial Intelligence. According to the Guidelines, in order to be trustworthy, AI should be lawful. Thus, AI has to respect all applicable laws and regulations, ethical principles and values, and to be robust from a technical perspective taking into account its social environment. Moreover, the Guidelines identifies seven key requirements that trustworthy AI should meet to be classified as such. The first regards the relation with humans that need to be allowed to make informed decisions, and it identifies three possible system relationships: human-in-the-loop, human-on-the-loop, and human-in command. The second is about technical robustness and safety: AI systems need to be resilient, accurate, reliable, and reproducible in order to minimize any unintended harm. Thirdly, AI systems have to ensure full respect for privacy and data protection, and legitimised access to data through adequate data governance mechanisms. The fourth addresses transparency issues, underlying the need for transparent data, system, and AI busines models. The fifth is related to the non-discrimination principle suggesting that unfair bias must be avoided. The sixth underline the importance of social and environmental well-being. Accordingly, it says that AI system should be sustainable and environmentally friendly and should take into account other living beings, considering their social and societal impact. The last is about accountability: mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes especially in critical applications.[45]

These are the latest laws that AI systems need to meet, at least in Europe. These are a halt on the development of fully autonomous weapon systems, particularly the first, the third and the seventh principles above. Few scholars agree about the position suggesting "to focus the development of future military technologies away from these so-called ethical systems and towards the development of systems that can actually improve the ethical conduct of humans in armed conflicts".[46] Other scholars suggest that "military organizations—network-centric ones, in particular—tend to have a strong preference for close operational supervision of frontline forces and that commanders will be reluctant to surrender control over sensitive operations."[47] It is

[45] "Ethics guidelines for trustworthy AI", in *European Commission*, (April 8, 2019). https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. (05/02/2021).
[46] Peter Asaro, "On banning autonomous weapon systems", 709.
[47] Michael Carl Haas, Sophie-Charlotte Fischer, "The evolution of targeted killing practices", 297.

unclear if autonomous weapon systems will be able to comply with IHL principles and, moreover, if it would be established the criminal individual responsibility for breaches of IHL through autonomous systems, given the opaqueness of algorithms.[48]

[48] Michael Carl Haas, Sophie-Charlotte Fischer, "The evolution of targeted killing practices", 300.

References

- Amoroso, Daniele. Tamburini, Guglielmo. "Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues". *Curr Robot Rep*, n. 1 187-194 (2020). Accessed: February 5, 2021. https://doi.org/10.1007/s43154-020-00024-3.
- Asaro, Peter. "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making". *International Review of the Red Cross*, Vol. 94, n. 886, 687-709 (Summer 2012). Accessed: February 5, 2021. https://doi.org/10.1017/S1816383112000768.
- Boulanin, Vincent. Verbruggen, Maaike. *Mapping the Development of Autonomy in Weapon Systems.* Stockholm, SIPRI. (2017). Accessed: February 5, 2021. https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf.
- Cumming, M. L.. "Artificial Intelligence and the Future of Warfare". *Chatham House* (2017). Accessed: February 5, 2021. https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings.pdf.
- "Ethics guidelines for trustworthy AI". *European Commission*, (April 8, 2019). Accessed: February 5, 2021. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
- Haas, Michael Carl. Fischer, Sophie-Charlotte. "The evolution of targeted killing practices: Autonomous weapons, future conflict, and the international order". *Contemporary Security Policy*, 38:2, 281-306 (2017). Accessed: February 5, 2021. https://doi.org/10.1080/13523260.2017.1336407.
- Müller, Oliver. "'An Eye Turned into a Weapon': a Philosophical Investigation of Remote Controlled, Automated, and Autonomous Drone Warfare". *Philosophy & Technology*, December 15, 2020. Accessed: February 5, 2021. https://doi.org/10.1007/s13347-020-00440-5.
- Kindervater, Katherine Hall. "The emergence of lethal surveillance". *Security Dialogue*, Vol. 47, n. 3 (2016), 223-238. Accessed: February 5, 2021. https://www.jstor.org/stable/10.2307/26294130 .
- "New SIPRI Reflection Film on limits on autonomy in weapon systems". *SIPRI*. March 27, 2020. Accessed: February 5, 2021.

https://www.sipri.org/news/2020/new-sipri-reflection-film-limits-autonomy-weapons-systems.

- Schmitt, Michael N.. Thurnher, Jeffrey S.. "”Out of the Loop”: Autonomous Weapon Systems and the Law of Armed Conflict”. *Harvard National Security Journal*, vol. 4 (2013), 231-281. Accessed: February 5, 2021. https://harvardnsj.org/wp-content/uploads/sites/13/2013/01/Vol-4-Schmitt-Thurnher.pdf.
- Shkurti Özdemir, Gloria. "Artificial Intelligence application in the military: the case of United State and China”. *SETA*, n. 51 (2019). Accessed: February 5, 2021. https://www.setav.org/en/analysis-artificial-intelligence-application-in-the-military-the-case-of-united-states-and-china/.
- Senn, Martin. Troy, Jodok. "The transformation of targeted killing and international order”. *Contemporary Security Policy*, 38:2, 175-211 (2017), Accessed: February 5, 2021. https://doi.org/10.1080/13523260.2017.1336604.
- Yaacoub, Jean-Paul. Noura, Hassan. Salman, Ola. Chehab, Ali. "Security analysis of drones systems: attacks, limitations, and recommendations” *Internet of Things*, May 8, 2020. Accessed: February 5, 2021. 10.1016/j.iot.2020.100218.
- Watling, Jack. Waters, Nicholas. "Achieving Lethal Effects by Small Unmanned Aerial Vehicles. Opportunities and Limitations”. *The RUSI journal*, 164:1, 40-51 (2019). Accessed: February 5, 2021. https://doi.org/10.1080/03071847.2019.1605017.