Digital War (2020) 1:173–177
https://doi.org/10.1057/s42984-020-00028-5

ORIGINAL ARTICLE

# Defining semi-autonomous, automated and autonomous weapon systems in order to understand their ethical challenges

**Jean-François Caron**[1,2]

Accepted: 10 November 2020 / Published online: 24 November 2020


**Abstract**
There is a lot of misunderstandings when it comes to what has been labelled as "autonomous killing robots". Indeed, the robotization of weapons is a very complex issue and requires a clear conceptualization of these various types of weapons that are currently being used in the military. This article offers a typology of these weapon systems by distinguishing between semi-autonomous, automated and autonomous weapons. This necessary distinction allows for a better understanding of the ethical challenges associated with these systems.

**Keywords** Autonomous weapon systems · Automated weapon systems · Semi-autonomous weapon system · War · Military technologies · Just war theory

If we are to believe some reports, the world of warfare is about to be profoundly changed. Not only are we to witness the deployment of super soldiers on the battlefields (Caron 2018), but autonomous killing robots will also replace human combatants. In light of the various Hollywood scenarios that have been made in the last decades, this prospect is for many of us very problematic. Indeed, how can we think objectively about this possibility when our minds are influenced by movies in which mankind is losing control over the machines, such as *War Games* or the *Terminator* franchise? Yet, ignoring that cultural legacy is essential if we are to assess the appropriateness of using these weapons systems and face the ethical questions connected with this new reality. The other challenge is to have a clear understanding of what we are talking about when we are talking about autonomous machines. There are a lot of confusion in this regard as a lot of people tend to assimilate technologies such as drones or the Israeli Iron Dome defense system in the same category as HAL in *2001: A Space Odyssey*, namely a computer able to make decisions on its own. This is a serious mistake that needs to be overcome. If the latter system that came out of Stanley Kubrick and Arthur C. Clark's minds can be labelled as an autonomous system, the formers fall within two different categories, namely of semi-autonomous and or automated weapon systems.

This therefore begs the question of how we can we differentiate between these types of weapons? It can be argued that one way of distinguishing between them is through the relationship humans have with the machines when they are performing tasks; in other words, whether there is a human in the loop as well as how their lethal capacities operate. In this perspective, some weapons' autonomy is solely preprogrammed, and their lethal capacities remain entirely the prerogative of a human operator. This is the case with many military technologies, such as drones that are able to fly autonomously to a certain location. However, they cannot fire their weapons without the direct intervention of a human being. We can think in this regard to the US Predator[1] and Reaper drones whose non-lethal autonomy remain largely akin to that of a standard plane with its auto-pilot function. However, when it comes to firing their Hellfire missiles, these weapons cannot act on their own. They are fired by a human agent according to certain rules of engagement.

Secondly, we find automated weapon systems that possess a destructive and/or lethal capacity. The best examples in this regard are the Israeli Iron Dome and the South Korean SGR-A1. Contrary to semi-autonomous weapons,

This text is an extract from Caron (2019).

✉ Jean-François Caron
jean-francois.caron@nu.edu.kz

[1] Nazarbayev University, Nur-Sultan, Kazakhstan

[2] University of Opole, Opole, Poland

[1] The Predator has been retired from active service by the US military in early 2018.

these systems are able to fire at specific targets without the direct intervention of a human operator. Indeed, both systems are programmed to either identify incoming rockets and other projectiles or enemy combatants and to intercept or fire at them.[2] We can also add to the list of examples the Sea Hunter, a prototype unmanned submarine tracking vessel developed by the Defense Advanced Research Projects Agency (DARPA) that will very soon join the US Naval fleet and has been described as "a highly autonomous unmanned ship that could revolutionize US maritime operations" and "a new vision of naval surface warfare" (Turner 2018). Even if the initial goal was to use this type of vessel for surveillance purposes, the US Navy tested the Sea Hunter in August 2017 with an offensive anti-submarine payload system which means that the likelihood is high that it might be able to locate, track, and engage enemy submarines in the near future. If the Navy ever decides to move forward with this vessel, it would become very similar to the Iron Dome. Similar to the Israeli defense system, the vessel could potentially replace whole fleets of destroyers that were previously dedicated to anti-submarine warfare thanks to a pre-programmed system that would only engage specific targets which would be detected because of their unique characteristics, such as Yasen or Akula class Russian submarines whose dimensions and features are different from US Los Angeles class submarines.

Even though these systems are often referred to as being "autonomous" (see for instance Sparrow 2007,[3] 63), this designation is misleading. Indeed, the notion of autonomy refers clearly to features that are not associated with the previously discussed weapons systems, since an autonomous agent is someone who is first and foremost able to pose a deliberate and independent action that results from his own will. This means that breathing is not sufficient in itself to define a living creature as an autonomous agent, since this action is involuntary and a natural result of the parasympathetic nervous system. The same logic would apply to the photosynthesis process of plants and other organisms. On the contrary, an autonomous action refers to an act that results from deliberate intent, which implies free will. Similarly, an individual who is under hypnosis or whose mental faculties are impaired cannot be considered an autonomous agent. This also implies that the intended action must result from a deliberative process that takes into account the difference between what is right and wrong. This faculty is at the core of how criminal responsibility is understood.[4]

In the case of military technologies, this understanding of autonomy would refer to their capacity to determine on their own and without any form of human interference when and against whom to use lethal force. This is clearly not the case with these aforementioned weapons systems since their lethal potential lies either with humans (as in the case of drones) or through a pre-programmed algorithm. In order to talk about autonomous weapons systems, these technologies would need to possess the capacity to exercise moral judgment in their killing process. However, it must be noted that these weapons do not exist at the current time and considering the inherent difficulties associated with their potential development, it is impossible to say if scientists will ever be able to create them. On the contrary, it is clear that there is an intention to transform this fantasy into a reality. This conclusion is supported by the rhetoric of many senior military officers and by the tremendous investments that states have allocated in recent years for the research and development of these weapons. Indeed, we cannot ignore the fact that Russian military commanders have openly said that "a fully robotized unit will be created [in the near future], capable of independently conducting military operations", while it has been stated by the US Department of Defense that the option of developing autonomous weapons able to determine on their own who should be targeted ought to be on the table (Scharre 2018, 6). Moreover, the Pentagon has recently announced that it will invest USD 18 billion in the research and development of such technologies.[5]

One of the main goals behind these investments is the capacity to develop autonomous platforms that are able to utilize artificial intelligence (AI) in order to operate on their own and to behave without a human in the loop. If it is ever realized, this innovation will open the door to the third revolution in warfare, after the invention of gunpowder and the nuclear bomb. Coined for the first time in 1956 by John McCarthy at the British Dartmouth Summer Conference, AI refers to "the science of making machines do things that would require intelligence if done by men" (Minsky 1968, V). In order for AI to be used to its full potential in the military domain, this advanced technology ought to be able to

[2] The autonomous nature of the SGR-A1 defence system has been a hotly debated topic in the last few years. Despite the fact that the spokesperson for Samsung Techwin—the company that developed the weapon—said in 2010 that the weapon "cannot automatically fire at detected foreign objects or figures", it was however revealed 3 years earlier that the technology could engage targets on its own without human intervention.

[3] Noel Sharkey is one of the few who adequately sees and engages with the problem of labelling weapons such as drones as autonomous (Sharkey 2010, 376).

[4] This is why "someone who does not know the difference between right and wrong is not a moral agent and not appropriately censured for her behaviours. This is, of course, why we do not punish people with severe cognitive disabilities like a psychotic condition that interferes with the ability to understand the moral character of her behaviour" (Einar Himma 2009, 23).

[5] It is estimated that global spending on military robotics will reach around USD 7.5 billion a year in 2018.

achieve three important features, namely: (1) to be able to analyze all possible outcomes and to suggest the best possible strategy and, if necessary, (2) to have intelligent robots coordinate a common action together, as well as (3) to display an analytical ability to show the same moral discernment as human beings. Currently, AI is only able to fulfill the first two tasks.

While it has taken more time than originally expected for scientists to develop computers intelligent enough to beat chess and Go players,[6] this was finally achieved in February 1996 when Deep Blue, an IBM supercomputer, was able to beat the world chess champion Gary Kasparov. As mentioned by Armin Krishnan, since then, "The days are now definitely gone when humans could seriously compete with computers in the chess domain" (Krishnan 2009, 47). We had to wait about twenty more years before a computer named AlphaGo was able to beat world champion Lee Sedol at the ancient Chinese game of Go, a game that requires far more intuition than chess. While Deep Blue relied on its capacity to evaluate millions of possible moves at the same time, AlphaGo relied for its part on its capacity for reinforcement learning and is therefore more akin to the decision-making process of humans. Indeed, AlphaGo's abilities were progressive and based on numerous attempts and errors. The machine was programmed to play countless games against itself, which helped it to learn from its mistakes and to devise alternative strategies. This is how it was eventually able to develop unprecedented moves that had never before been used by Go players.

Like in many other domains, AI has not been limited to games and has been integrated in different parts of military computer systems and robots. One of the best examples is certainly DARPA's Deep Green system that helps military commanders to have a better view of their strategies by generating their likely possible outcomes, thereby suggesting what might be the best course of action. In light of the accelerating stream of data and information that military commanders are now faced with, it is fairly easy to understand why this system has been developed, since human beings' capacity to process this information is not increasing. Of course, this system keeps a human in the loop and its aim "is not to replace the human military commander with a machine, but to enable [him] to master the enormous complexity of modern war"' (Krishnan 2009, 54). Alongside the use of computers as decision support tools (which is inherently non-problematic from the perspective of the morality of warfare), current technologies can also allow the armed forces to better coordinate their actions. Indeed, swarm systems are allowing machines to enjoy far more autonomy and have been developed with the aim of copying the swarm intelligence of ants and to use this in military technologies, especially for drones. There is indeed something fascinating about these insects. While they are vulnerable and unable to solve simple navigational puzzles when they are isolated from their peers, they show a strong collective intelligence by coordinating themselves in an effective manner without communicating with each other. This is why militaries have developed multi-agent systems that enable multiple robots to independently act in concert rather than hinder one another. This was done by the US military in 2003, when 120 small robots equipped with swarm intelligence flew in a coordinated manner. Since then, the Chinese military has also shown its ability to use this technology effectively. One can easily understand the effectiveness of this system through a hockey analogy. While it is easy for two defensemen to coordinate themselves when a forward from the other team is trying to enter their territory with the puck, this task would be impossible in a situation where there were five forwards with five different pucks against five defensemen. It would, however, be possible for a swarm of robot defensemen to very quickly devise a way of successfully defending their territory by collectively deciding on a course of action. The US Navy has shown the usefulness of this system during a test in 2014. A number of small swarm boats were deployed near a high-value ship. The human controller's task was simply to order the swarm boats to intercept a suspicious vessel by coordinating themselves autonomously. As stated by Paul Scharre:

> Bob Brizzolara, who directed the Navy's demonstration, called the swarming boas a 'game changer'. It's an often-overused term, but in this case, it's not hyperbole—robotic boat swarms are highly valuable to the Navy as a potential way to guard against threats to its ships. In October 2000, the USS Cole was attacked by al-Qaida terrorists using a small explosive-laden boat while in port in Aden, Yemen. The blast killed seventeen sailors and cut a massive gash in the ship's hull. Similar attacks continue to be a threat to US ships, not just from terrorists but also from Iran, which regularly uses small high-speed craft to harass US ships near the Straits of Hormuz. Robot boats could intercept suspicious vessels further away, putting eyes (and potentially weapons) on potentially hostile boats without putting sailors at risks (2018, 22).

These examples show that AI military technology has reached a point where machines can analyze situations, provide advice, and coordinate each other: features that have been labelled by Jacob Turner as a weak form of AI (2019, 6). In order to have a stronger form of AI, the technology would have to encompass many of the attributes of human

[6] It took almost 40 years before a computer was able to beat a chess Grand Master, while it was originally predicted that it would take 10 years (Krishnan 2009, 47).

intelligence and resemble the kind of robots portrayed in popular culture—namely, systems that would allow robots to identify targets and to decide whether to fire at them. This is likely the aspect of AI that scares most people and raises the question as to whether it is possible to have robots develop their own moral code and to make them behave in a way that is ethically responsible. Consider the following scenario: a fully autonomous drone is flying about a city and positively identifies a well-known terrorist who has been involved—and who is still involved—in terrorist acts against civilians. However, he is surrounded by young children and a group of elderly women. Should the robot fire a Hellfire missile at this high-value target? Such a decision comes with many fundamental moral questions and can either be solved through consequentialist (killing this man and innocent civilians in his vicinity will ultimately contribute to saving more lives) or deontological ethics (according to which it is immoral under any circumstances to justify the murder of civilians). The now famous example of the naked soldier evoked by Michael Walzer is another good example in this regard. He is asking us to imagine that "a soldier while on patrol or on sniper duty catches an enemy soldier unaware, holds him in his gunsight, easy to kill, and then must decide whether to shoot him or let the opportunity pass" (2006, 138–139). If killing this soldier is not problematic in the eyes of Walzer, he nonetheless agrees that for many of us, harming a soldier who is taking a bath, calmly smoking a cigarette, or doing his business behind a bush would be morally repulsive (and he gives many examples in this regard in his book). This hesitation is the quintessential representation of moral agency and the true meaning of an autonomous subject.[7] At the moment, the capacity to make such moral decision is not a feature of any weapon system.

As previously mentioned, AI is not sophisticated enough to allow machines to make the same moral judgments as human beings on a regular basis.[8] In fact, as argued by Kenneth Einar Himma, "it is clear that an artificial agent would have to be a remarkably sophisticated piece of technology to be a moral agent" (2009, 28) and that this might very well never be achieved. It is, however, what researchers are currently trying to accomplish through a pre-programmed mode. More precisely, they are trying to determine ethical patterns in how human beings make moral decisions in numerous circumstances. For instance, how would drivers react if a child on a bike were to swerve in front of them while the only option to avoid him was to swerve onto a sidewalk where a group of elderly women were taking a walk? Who would they choose to hit? When we would finally be able to determine how morality functions, computer engineers and technicians would then try to program these patterns into an AI. If this task is ever achieved by scientists, these intelligent robots will simply be a more advanced form of a pre-programmed decision-making process, which still does not qualify as autonomy. Since there are currently no robots that are able to engage targets independently of a human's will, there is thus no need at this point to entertain a conversation on these science fiction-esque weapons.

It is, of course, easy to have our attention diverted from the essence of the current debate by fears that are mainly supported by films such as *Slaughterbots,*[9] a movie that was shown as a side event hosted by the Campaign to Stop Killer Robots in November 2017 as a propaganda tool to convince people that autonomous weapons should be banned. Although controversial and conducive to debates, such films are an unfaithful representation of what these weapons are currently capable of. In this sense, it is not helpful to envisage that terminators and other technologies may end up turning against their creators and wipe out the entire human race.

The intelligence of weapons with a pre-programmed lethal capacity are not automatically more morally problematic because the life and death decision is no longer the sole prerogative of a human being. In fact, sticking solely to the notion of having or not having a human in the loop as a way of distinguishing the morality of these weapons is irrelevant in allowing us to determine the ethical permissibility of these weapons. Those who are honest will admit that while some of these weapon systems can be problematic, others have not led to the indiscriminate killing of countless civilians. This nuanced judgment is explained by the fact that other factors need to be considered—namely, their overall intelligence, the way they are used and programmed, as well as the identity of those who are targeted. It is from these various factors that these aforementioned technologies may be deemed to be or not to be ethically permissible. Let us take the example of the landmine that can be considered as a lethal weapon with pre-programmed autonomy, in the sense that it is designed to detonate on its own when pressure is exerted on it. This weapon, however, suffers from a major flaw: it has no freedom when it comes to determining if it should explode depending on the nature of those who are stepping on it. This incapacity to decide whether or not

[7] As Leveringhaus puts it correctly, "the act of programming negates any autonomy in a philosophical sense" (Leveringhaus 2016, 48).

[8] As noted by Vincent Conitzer, a Professor of Computer Science at Duke University who is working on allowing AI to make moral judgments, "Recently, there have been a number of steps towards such a system, and I think there have been a lot of surprising advances (…) but I think having something like a "true AI", one that's really as flexible, able to abstract, and do all these things that humans do so easily, I think we're still quite far away from that" (Creighton 2016).

[9] This short film depicts a future in which autonomous drones are going berserk and turn themselves against US Senators and university students.

to explode makes this type of autonomous weapon a rather indiscriminate and unintelligent one, which is why anti-personal landmines have been banned following the 1997 Ottawa Treaty. The German Falcon torpedo used during WWII is also another good example of a not-so-wise autonomous weapon. Equipped with an acoustic homing seeker, this type of torpedo did not travel in a straight line like traditional torpedoes. Using its acoustic sensors, it was able to detect ships and modify its trajectory accordingly. While it was a revolutionary weapon at the time that allowed for the more precise and deadly targeting of Allied merchant ships, it nonetheless faced a serious problem: two of the three U-boats equipped with this technology were sunk after their torpedoes detected the sound of the submarines' propellers and circled back on them. On the other hand, defense systems like the Israeli Iron Dome, the German Nächstbereichschutzsystem MANTIS, or the South Korean SGR-A1 can be considered as more clever autonomous systems because of their capacity to only fire on specific targets. Indeed, these systems are programmed to either identify incoming rockets and other projectiles or enemy combatants, and to intercept or fire at them without any human intervention. In the case of the Iron Dome, the Israeli military has displayed numerous batteries on strategical positions throughout the country that are constantly moved to fool the enemy and to adjust to new threats. When the radar system detects an incoming missile, the sophisticated algorithm determines in a few seconds the type of projectile that has been fired and if it is aimed at populated civilian areas or military infrastructure. If this is the case, interception missiles are launched. Since the deployment of this system in 2011, more than a thousand of Hezbollah and Hamas' rockets have been intercepted with an incredible success rate,[10] which has of course changed the lives of Israeli citizens living close to Lebanon or the Gaza Strip (Human Rights Council 2015, 151). In return, this system has not led to the destruction of commercial airliners or the death of innocent civilians.

At the end of the day, one of the main benefits of this conceptual clarifications is not only to defuse the fear that senseless computers can now decide to harm us based on their own free will. As it stands now, human beings are still either directly (in the case of drones) or indirectly involved (in the case of automated weapons in the way algorithms are developed) in the decision to kill in warfare. It also gives us a better idea of the specific ethical challenges associated with these respective weapon systems. One of most important being our capacity to determine the criminal responsibility of those responsible of a system malfunction or the unfortunate death of innocent civilians. If attributing this responsibility can be difficult in the case of an autonomous system, it is easier to determine in the two other systems since their respective reaction depends either on the decision of a human being (in the case of drones) or of individuals who have determined how the automated reaction ought to work.

[10] During Operation Pillar of Defense, the system successfully intercepted 84% of rockets and mortars fired against Israel, while the success rate reached 91% during the first part.

## References


Caron, Jean-François. 2018. *A Theory of the Super Soldier: The Morality of Capacity-Increasing Technologies in the Military*. Manchester: Manchester University Press.

Caron, Jean-François. 2019. *The War of the Machines: Contemporary Technologies and the Morality of Warfare*, 2019. London: Routledge.

Creighton, Jolene. 2016. The Evolution of AI: Can Morality be Programmed?. *Futurism*, 1 July. https://futurism.com/the-evolution-of-ai-can-morality-be-programmed/.

Einar Himma, Kenneth. 2009. Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent have to be a Moral Agent? *Ethics and Information Technology* 11 (1): 19–29.

Human Rights Council. 2015. Report of the detailed findings of the independent commission of inquiry established pursuant to Human Rights Council resolution S-21/1. 29th session.

Krishnan, Armin. 2009. *Killer Robots: Legality and Ethicality of Autonomous Weapons*. London: Ashgate.

Leveringhaus, Alex. 2016. *Ethics and Autonomous Weapons*. Oxford: Palgrave Macmillan.

Minsky, Marvin. 1968. *Semantic Information Processing*. Cambridge, MA: The MIT Press.

Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. New York, London: W.W. Norton & Company.

Sharkey, Noel. 2010. Saying No! To Lethal Autonomous Targeting. *Journal of Military Ethics* 9 (4): 369–383.

Sparrow, Robert. 2007. Killer Robots. *Journal of Applied Philosophy* 24 (1): 62–77.

Turner, Jacob. 2019. *Robot Rules. Regulating Artificial Intelligence*. London: Palgrave MacMillan.

Turner, Julian. 2018. Sea Hunter: Inside the US Navy's Autonomous Submarine Tracking Vessel. *Naval Technology*, 3 May. https://www.naval-technology.com/features/sea-hunter-inside-us-navys-autonomous-submarine-tracking-vessel/.


**Jean-François Caron** is an Associate Professor in the Department of Political Science and at the Institute of Political Science and Administration at the University of Opole.