To Ban or not to Ban. Analyzing the Banning Process of Autonomous Weapon Systems

Celien De Stercke
Ghent University, Department of Criminology, Criminal Law and Social Law, Ghent, Belgium
https://doi.org/10.38126/JSPG210102
Corresponding author: Celien.DeStercke@UGent.be
Keywords: autonomous weapons; killer robots; AWS; cyber warfare; policy analysis

Executive Summary: Over the last decade, autonomous weapon systems (AWS), also known as 'killer robots', have been the subject of widespread debate. These systems impose various ethical, legal, and societal concerns with arguments both in favor and opposed to the weaponry. Consequently, an international policy debate arose out of an urge to ban these systems. AWS are widely discussed at the Human Rights Council debate, the United Nations General Assembly First Committee on Disarmament and International Security, and at gatherings of the Convention of Conventional Weapons (CCW), in particular the Expert Meetings on Lethal Autonomous Weapons Systems (LAWS). Early skepticism towards the use of AWS brought a potential ban to the forefront of policymaking decisions with the support of a campaign to 'Stop Killer Robots' launched by the Human Rights Watch (HRW) in 2013. The movement is supported by Amnesty International, Pax Christi International, and the International Peace Bureau, among others. This campaign has catalyzed an international regulation process on the level of the United Nations (UN). Both a new protocol to the Convention on Conventional Weapons or a new international treaty have been considered. However, a lack of consensus stalls the process, and as such, leaves AWS in a regulatory grey zone.

I. Introduction
Autonomous weapon systems (AWS) are “weapon system[s] that, once activated can select and engage targets without further intervention by a human operator” (US Department of Defense, 2012, 13). Until recently, (lethal) autonomous weapon systems (LAWS), were under a primarily theoretical discussion to be regulated or even banned pre-emptively. In 2020, the theory became practice when the weaponry debuted in the Libyan war, as described in a report of the United Nations (2021; Hernandez 2021). The deployed STM Kargu-2 loitering munition produced in Turkey was explicitly described as a LAWS that was “programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true ‘fire, forget and find’ capability.” (UN 2021, 17). ‘Fire, forget and find’ is the ability “to select and attack targets on its own” (Hernandez 2021; Werkhäuser 2022, 2).

LAWS “can take the form of drones, land vehicles, or submarines” (Werkhäuser 2022, 2). Drones, like Kargu-2, are more technically labelled as unmanned aerial vehicles (UAVs), such as the military Predator and Reaper drones (Williams 2011; Scharre 2018). UAVs were first developed in the 1950s for surveillance purposes and made their appearance on the battlefield in the 1980s (Williams 2011). Although still in a developmental stage, they were weaponized around the turn of the century with the rising terrorism threat (Williams 2011). Their development, capabilities, and deployment have grown exponentially, including their autonomous characteristics (Williams 2011; Scharre 2018). The next stage of military drones can be marked by loitering munitions, also known as suicide or kamikaze drones (Rogoway 2021; Gao 2022). As one of the first of its kind, Israel's (IAI) Harpy and its successor the Harop are prominent examples, next to more recent systems such as the Switchblade and
Phoenix Ghost used in Ukraine (Rogoway 2021; Lopez 2022).

“The Harpy would be launched and enter a searching pattern, waiting for a radar to activate. If a radar activated, the Harpy would then home in on and destroy the radar using a blast fragmentation warhead in its body. The Harpy could loiter over the battlefield for up to six hours after launch.” (Gao 2022, 6)

These highly autonomous loitering munitions were already introduced in the 1990s, and in specific scenarios can operate fully autonomously (Scharre 2018). "Autonomy is simply the ability for a machine to perform a task or function on its own" (Scharre 2018, 27). However, autonomy as a concept is much more nuanced. Paul Scharre (2018), “a leading expert in next-generation warfare” (447), divides the concept into three dimensions: (1) “the type of task the machine is performing”, (2) “the relationship of the human to the machine when performing that task”, and (3) “the sophistication of the machine’s decision-making when performing that task” (27).

An example of the first dimension is landing: several drones do have the ability to land on their own. The second dimension can take various forms: a human can be pushing a button to initiate pursuit ('human in the loop'), supervise and manually override ('human on the loop'), or not intervene at all ('human out the loop')(Scharre 2018). To what extent a machine makes the decision, is encompassed by the third dimension. Hence, technology can be more or less autonomous depending on a spectrum of factors. The Harpy and Harop “can search for, decide to engage, and engage targets on their own and no human can intervene” (Scharre 2018, 46), as such, making them fully autonomous weapon systems under this conceptualization.

The latest game-changer in the field has been the LAWS Kargu-2. It operates with greater autonomy relative to its siblings on two main levels: (1) it uses machine learning algorithms (artificial intelligence) to select its targets (instead of primarily sensory-inputs and predefined programs) and (2) it is designed to attack humans (Nasu 2021). Notwithstanding the absence of an international consensus definition of (L)AWS (Sayler 2021), autonomous weapon systems with the capacity for lethal action are being used currently and further developed.

Prof. Noel Sharkey, by many considered the pioneer of the banning process of AWS, called for regulation as early as 2007 (Sharkey 2007; Campaign to Stop Killer Robots 2021). He warned against the development of fully autonomous robots making lethal decisions and stipulated the necessity for international regulation (Sharkey 2007; Campaign to Stop Killer Robots 2021). This precautionous standpoint has been followed by the International Committee of the Red Cross (ICRC) in 2011 (Expert meeting ICRC 2014) and became more formalized in 2013 since a campaign to ‘Stop Killer Robots’ has been launched by the Human Rights Watch (2020). Since then, a process to design an international regulatory framework has been initiated at the United Nations (HRW 2020). As a result of this movement, a new protocol to the CCW or a novel international treaty on AWS may be developed (HRW 2020).

The debate to regulate or even ban LAWS is still ongoing. This brings us to an important question: “Which arguments can be made to ban (or not to ban) autonomous weapon systems from across societal, ethical, and legal points of view?”.

i. To ban

The social argument against AWS is that its deployment can negatively influence global peace and stability by contributing to a new arms race (Tamburrini 2016; Scharre 2018).

AWS also pose potential security concerns. The technology industry cannot guarantee that these weapons would be invulnerable to hacking or spoofing, indicating that they are not infallible (United Nations Institute for Disarmament Research (UNIDIR) 2017). If an (L)AWS is hacked while being at its native military location, it can easily become a Trojan horse. Notwithstanding that “any computer system is, in principle, susceptible to hacking, greater complexity can make it harder to identify any vulnerabilities” (Scharre 2016, 14). Moreover, another known threat is called spoofing: a system that is “intentionally deceived into thinking that it sees objects – or military targets – that do not exist” (Reim 2020, 2), which could also encompass falsifying physical visual data (UNIDIR 2021). For
instance, by not wearing a uniform (and being seen as a civilian) or by looking like the adversary on the battleground (falsifying data cf. spoofing). This could result in the LAWS sensing “correctly” according to its programming and responding to it, though in reality interpreting the situation incorrectly. Furthermore, malfunctions and misclassifications because of changes in light, reflection, shadows, colors, and animals pose a risk as well. Although the security risks of AWS are considered by the technological and military field, the importance of security safeguards cannot be emphasized enough since “cyber fault management simply has higher stakes with autonomous systems” (UNIDIR 2017, 15). Moreover, note that some technologies are developed privately and for non-military use, though could be integrated into weapon systems at a later stage (e.g. artificial intelligence), which poses a risk for potential security gaps in a warfare context (UNIDIR 2017). Taken together, precaution regarding the security of the systems is needed.

Furthermore, the fundamental ethical question is whether we could delegate the decision over life and death to AWS. However, AWS as defined in this article does not necessarily mean solely ‘lethal’ but follows the notion of ‘use of force’, since this is more in line with current international humanitarian law (UNIDIR 2014). Some authors argue that delegating this right to use force against humans, to machines is inherently degrading to humans (Heyns 2016; Sharkey 2016). In addition, if the right to use force were delegated to a machine, then another issue arises: can the machine be held accountable for its acts? Can it be held morally responsible for its deeds? This ‘accountability issue’ has significant overlap with the legality of AWS.

From a legal point of view, a parallel can be made with human criminal liability: if an individual acts outside of his own control, then he cannot be held accountable for his deeds (Heyns 2016; Sartor and Omicini 2016). This would imply that machines wouldn’t be responsible for their deeds and even if they were, “there is clearly no point in putting a robot in jail” (Heyns 2016, 12). Furthermore, some experts argue that international law implicitly requires humans to make the decisions (Asaro 2013; Boulanin 2016; Heyns 2016), which has never been necessary to make explicit as the decision over (lethal) use of force was always made in human-to-human interactions (Heyns 2016). This latent argument supports the need for the notion of ‘human control’ to be compliant with international (humanitarian) laws.

An international consensus around the accountability issue as part of the larger AWS framework has already been reached, by retaining “a responsible chain of human command and control” over AWS (HRW 2020; UN 2019, Annex III, d; ICRC 2016). All High Contracting Parties to the CCW agreed to this stipulation in 2019, as encompassed by the principal guidelines from the Group of Governmental Experts on LAWS (see infra) (HRW 2020; UN 2019). What this notion of keeping ‘human control’ entails, is however unclear (HRW 2020; UN 2019; ICRC 2016). Various stakeholders prefer a strict sense of ‘meaningful human control’, indicating concretely an informed soldier pushing a button to launch an attack for example. Others prefer a lower threshold in retaining human control over the machines: it could for instance suffice that a human writes the AI program (Heyns 2016; UNIDIR 2014; HRW 2020; Davison 2018).

One of the most contested arguments about AWS is that their superior performance might allow them to “respect international humanitarian law or human rights law better than humans do” (Sharkey 2016; Heyns 2013; Tamburrini 2016; UNIDIR 2014, 6). For instance, it is unclear whether robots could comply with the discrimination and proportionality principle, both crucial elements in international humanitarian law (IHL) (Tamburrini 2016). Think for example of recognizing wounded soldiers in order to hold its fire, as attacking wounded soldiers is illegal under IHL (Nasu 2021; ICRC 2022). If AWS can’t differentiate them from combatants in action, their ability to comply with IHL is theoretically contested. The performance and discriminating abilities of the systems are therefore an important necessary condition for the systems to be IHL compliant. Arguably, this point is highly intertwined with the state of the art of the included technologies, such as artificial intelligence (AI). This argument could be addressed, however, by mandatory modeling and testing before producing the weapon systems. Though what if preliminary testing indicates that AWS could comply with international humanitarian law, but they fail to do so on the battlefield (Sharkey 2016)? Hence, Prof. Sharkey
(2016) underlines the necessity for precaution in developing violent robotic applications.

If AWS is able to comply with IHL, they would probably be more legally compliant than humans, without consideration of some practical concerns (Scharre 2018). Scharre, “leading expert in next-generation warfare” (Scharre 2018, 447), gives an example from his military past when a little girl was spotting the location of Scharre and his squad aiding Taliban fighters and in doing so, making herself an eligible target under the laws of war (Scharre 2018). However, they did not because it would have been morally wrong to do so. Scharre (2018) points out that even if robots could comply with laws, that “sometimes doing the right thing entails breaking the rules - what’s legal and what’s right aren’t always the same” (6).

### ii. Not to ban

AWS from a societal perspective, could save lives and make war more humane altogether (UNIDIR 2014; Heyns 2016; 2013; Scharre 2018; Etzioni and Etzioni 2017). Remotely operated AWS reduce the risk taken by military personnel and can minimize collateral damage due to greater weapon precision relative to human operators (Heyns 2016; Etzioni and Etzioni 2017). They also wouldn’t panic, lose attention, “get bored” “nor do they need food, water, or sleep” (Williams 2011, 5). They can operate “day and night” and could reach places people couldn’t (Williams 2011, 5). When the resources needed are compared to the return on investment, AWS seem more financially, efficiently, and effectively attractive than human soldiers (Heyns 2013; UNIDIR 2014; Scharre 2018; Etzioni and Etzioni 2017). Furthermore, in situations that challenge communications, the ability of the systems to operate fully autonomous could be for the better, compared to a semi-autonomous weapon that due to lack of communication wouldn’t be able to function properly (Scharre 2016). Hence, the case can be made that in some circumstances, they would be safer to use than their non-fully autonomous siblings (Scharre 2016). Notwithstanding possible limitations to their use, this argument implies not banning AWS.

Building upon the social argument against AWS, relating to a global arms race, the point can be reversed as well. If a global ban regarding AWS would be reached, it “will be ineffective in stopping their use by the states whose acquisition of such weaponry would be most dangerous” (Anderson and Waxman 2013, 15). As such, this would impose a significant disadvantage for other states to not develop AWS in international security dynamics, at the minimum as a form of deterrence (Horowitz 2019; Anderson and Waxman 2013).

As mentioned above, the technical security of intelligent weaponry is a reason for concern. However, this is not specific to autonomous weapon systems, but rather one against information technology in a wider sense (Scharre 2016). The imposed risks could be mitigated through minimum safeguards and technical requirements needing to be met before the production or deployment of AWS. With the support of the scientific, technical, and private domains (UNIDIR 2017), this seems a feasible counter measurement to apply. This implies a formalization of industrial requirements and as such argues in favor of a regulatory framework.

Referring to the IHL compliance aspect of AWS, a case was made for the systems required to be performant and discriminatory. This argument can be reverse-engineered against a potential ban. Pre-emptively banning the systems, including their development, could hamper the needed operational requirements being met (Anderson and Waxman 2013). Hence, banning their development would be “unjustified” (Anderson and Waxman 2013, 15).

At the crossroads of technology and law, the functional superiority in processing could imply that AWS would “be able to respect international humanitarian law or human rights law better than humans do” (UNIDIR 2014, 6). The continued research into and development of AWS, could “reduce misidentification of military targets, better detect or calculate possible collateral damage, or allow for using smaller quanta of force compared to human-decision making”, and probably even “more and more over time” (Anderson and Waxman 2013, 15).

This brings us to the accountability issue. AWS under the current IHL, could fall under command responsibility, a principle in customary international law (Sehrawat 2021). This could indicate that “individual(s) can be held criminally responsible for
their role as operators, commanding officers, programmers, engineers, technicians, or other relevant functions” (Sehrawat 2021; Stürchler and Siegrist 2017, 23). However, the applicability of this reasoning can be discussed as well (Stürchler and Siegrist 2017; HRW 2015; Sehrawat 2021). Besides, the individual criminal liability argument could be overly stipulated in the debate as individual liability is only one way to enforce IHL (Anderson and Waxman 2013; Sehrawat 2021). “[M]echanisms of state (or armed party) responsibility” (17) are also traditionally used to enforce compliance with IHL (Anderson and Waxman 2013).

Note that the applicability of IHL regarding AWS is not questioned by either side of this debate: IHL applies to AWS (HRW 2020; United Nations 2019). The interpretation regarding AWS under IHL on the contrary is, as it is not “specifically regulated” (Davison 2018, 7). A minority of country positions (cf. Russia) argue that an additional regulatory framework is redundant altogether, as IHL encompasses high autonomous weaponry already, and as such (L)AWS (HRW 2020).

On a final note, AWS impose various advantages in warfare as well foremost their potential to save the lives of civilians and military personnel, as substantiated above. In doing so, AWS could make war more humane altogether, adding an ethical dimension as well.

In summary, various arguments can be made both for and against AWS. There is no solid answer to the question of whether AWS should be banned or not, though it can be stated that precaution, retaining human control and some form of international regulation (whether that may be a ban or not) is highly necessary.

II. The regulation process: a ban?
These theoretical implications are practically mirrored in the international regulation process. Mainly situated on the United Nations level, two possible paths are explored: an additional protocol to the CCW or a new treaty could be established (HRW 2020). However, a lack of consensus results in a continuation of the status quo.

The United States, Israel, and Russia are the most notable countries not in favor of stringent regulation for AWS, such as a ban. The United States stipulate the humanitarian argument and therefore takes a neutral position in the debate: neither supporting a ban, nor explicitly in favor of the weaponry (HRW 2020). Israel takes a more extreme standpoint by openly rejecting any restrictions (HRW 2020), arguing that the systems could be able to respect IHL better than humans. Russia, on the other hand, acknowledges the risks such as peace destabilization and the potential of undermining international law (HRW 2020). Though, they argue that the current legislative framework suffices for the new developments as well (HRW 2020).

China takes a more precautious, though complex position. It explicitly agreed to a ban regarding the use of AWS, due to their uncertainty regarding IHL compliance and possible global peace destabilization (HRW 2020). In addition, China later added that they do not, however, support a strict regulation of the development or production of LAWS (HRW 2020).

The European Union and over 30 other states, such as Austria, Brazil, and Iraq, prefer more stringent regulation or even a ban on fully autonomous weapon systems (European Parliament 2018; 2014; Campaign to Stop Killer Robots 2021; HRW 2020).

i. A new treaty
The first Human Rights Council debate on lethal autonomous robotics was held in May 2013, with various countries and other stakeholders delivering their statements (Campaign to Stop Killer Robots 2021). Pakistan launched the first call for a ban during the convention (HRW 2020; Campaign to Stop Killer Robots 2021).

In the years following, gatherings continued without any progress on a new treaty. Both the UN General Assembly First Committee on Disarmament and International Security came together several times, as well as additional gatherings of the CCW on lethal autonomous weapons systems (Campaign to Stop Killer Robots 2021). For now, a new treaty regulating (L)AWS has been side-lined in favor of an additional protocol to the CCW.

ii. Additional protocol under the CCW
The starting signal for an additional protocol to the CCW was given in Geneva in 2013 (Campaign to Stop Killer Robots 2021). The event culminated in an
agreed mandate to start work in 2014 on regulating the emerging technology of lethal autonomous weapon systems (Campaign to Stop Killer Robots 2021). Various initiatives and gatherings of the multilateral expert meeting on lethal autonomous weapon systems (LAWS) have passed over the following years, without a real breakthrough.

The first concrete steps were taken at the Fifth Review Conference of the CCW in 2016 when a Group of Governmental Experts (GGE) was established to formalize the regulation process on lethal autonomous weapons (Campaign to Stop Killer Robots 2021). The GGE continued their mandate to explore options in the following years, despite multiple demands for a legally binding instrument by several countries such as Austria, Brazil, Chile, and Benin - on behalf of the African Group - (HRW 2020).

In 2019, the annual CCW meeting was held and concludes that the GGE mandate will be further expanded with their premised guidelines as a foundation (UN 2019). The 11 guidelines (UN 2019, Annex III, 10; UN Office for Disarmament Affairs (UNODA) n.d.) can be summarized in 7 points and conclude that:

1. international (humanitarian) law is applicable and should be taken into account at all times – (a, c, d, e, h);
2. human responsibility and accountability “must be retained” since this cannot be referred to machines – (b, d);
3. the extent of human-machine interactions needs to be considered over different factors, such as the “operational context, and the characteristics and capabilities of the weapon system as a whole” – (c);
4. risks need to be taken into account and measures to meet them, need to be made – (f, g)
5. that the systems should not be humanized in policies – (i);
6. the discussions should not hamper the positive development of autonomous technologies for peaceful uses and – (j);
7. that “the CCW offers an appropriate framework” to deal with the matter – (k).

Notwithstanding the relevance of these guidelines in the ongoing debate and regulation process, it took the GGE 4 years to draft them. As such, important years have passed as the technology continues to evolve, amid an unregulated AWS arms race. However, the guidelines put a concrete path forward to do so. The last provision could be interpreted as suggesting the drafting of an additional protocol to the CCW rather than a new treaty. The wording leaves some room for discussion though, as they state that “the CCW offers an appropriate framework” (UN 2019, Annex III, 10; UNODA n.d.). There was no discussion about the appropriateness of the CCW as a framework. Its main disadvantage is the fact that all parties must reach a consensus to amend the CCW, whereas a new treaty could form a starting point for future regulatory work (HRW 2020).

This lack of consensus is highly problematic in the current debate as it hampers progress toward regulation. Some countries are not necessarily against or in favor of regulating AWS, but in not being explicit, hampering any regulatory progress. Nonetheless, even hampering the regulation process has enormous consequences when compared to other international judicial initiatives since so-called “precursor weapons” are out of the scope: “everything that is already on the ground doesn’t count” (Topol 2016).

III. Conclusion and discussion
As discussed above, highly autonomous weapons do already exist (Scharre 2018; Topol 2016; Williams 2011), and they impose various societal, ethical, and legal benefits and concerns at the same time (see supra). Therefore, the weaponry became the subject of widespread international debate and even the subject of a banning call. A regulatory framework to encompass the systems into IHL has been on the political agenda for over a decade. The United Nations explored two possible paths to do so: in the form of an extra protocol to the CCW or design a new treaty. The lack of consensus on the debate is mirrored in the stagnating regulation process. Taken together with the pace of technological advancement in the field, hampering the regulatory process is enough to develop the weapons and evade the final regulation. Today, a status quo has been reached. A ‘coalition of the willing’ or industrial requirements, could perhaps initiate action in the field.
A so-called ‘coalition of the willing’ in a new treaty could form a baseline for further judicial work. In doing so, minimum rules could be established and a legal threshold could be enforced amongst the signatories. The point of reference could also deter unsigned states, as it would set the tone in the international community.

Furthermore, it could be useful to take other options into account as well, such as the International Organization for Standardization (ISO). ISO norms could be a stepping stone towards consensus by agreeing on industrial quality and minimum standards. Such norms are being prepared or do already exist for UAVs and military certifications (ISO 2014; 2021). The industrial safeguards could enforce high performance, certain security levels, and state requirements such as ‘human in or on the loop’ functionalities. This does not replace a legislative framework whatsoever, though could help the global status quo in moving forward.

References


Celien De Stercke is a Criminologist at the Institute for International Research on Criminal Policy (IRCP) and a member of the imec-DistriNet, a Research Group Computer Sciences at the KU Leuven. In 2022, she graduated from Ghent University with an M.Sc. in Criminological Sciences. She aims to pursue a PhD on the intersection of her technical and social background, being the phenomenon of cyber warfare. Celien researches the applicability of governance of security theories to cyberspace, in particular cyber mercenaries, in order to develop governance architecture model scenarios to counter cyber threats.

Acknowledgements
The author would like to thank the reviewers, Joseph Long and Ethan FitzGerald, for their valuable feedback on the manuscript.