

Article

Beyond Deadlock: Low Hanging Fruit and Strict yet Achievable Options in AWS Regulation

Maciej Zajac¹

¹ Institute of Philosophy and Sociology, Polish Academy Of Sciences; maciej.zajac@ifispan.edu.pl

Abstract: Efforts to ban Autonomous Weapon Systems were both unsuccessful and controversial. Simultaneously the need to address the detrimental aspects of AWS development and proliferation continues to grow in scope and urgency. The article presents several regulatory solutions capable of addressing the issue while simultaneously respecting the requirements of military necessity and so attracting a broad consensus. Two much stricter solutions – regional AWS bans and adoption of a no first use policy – are also presented as fallback strategies in case achieving AWS' compliance with the Laws of Armed Conflict proved elusive. Together, the solutions presented form an outline of a flexible regulatory strategy able to adjust to different technological outcomes and providing a sensible compromise to solve the current deadlock on the AWS issue.

Keywords: Autonomous Weapon Systems, Killer Robots, Military Ethics, Ethics of Technology, Laws of Armed Conflict, International Humanitarian Law, arms control, no first use policy, Ethics of War

Citation: Zajac, Maciej. 2022. Beyond Deadlock: Low Hanging Fruit and Strict yet Achievable Options in AWS Regulation. *Journal of Ethics and Emerging Technologies* 32: 2. <https://doi.org/10.55613/j eet.v32i2.116>

Received: 30/09/2022

Accepted: 30/11/2022

Published: 31/12/2022

Publisher's Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The issue of regulating Autonomous Weapon Systems (AWS) – weapon systems capable of selecting and/or engaging their targets based only on their own programming – has received comparatively little attention. This is surprising given that AWS are one of the most frequently discussed subjects in present-day military ethics and ethics of technology. However, debates have so far centered around the proposals for a broad and unconditional ban on AWS (Amoroso & Tamburrini 2017; Asaro 2012; Human Rights Watch 2016; Heyns 2017; Sauer 2020; Sparrow 2007, 2016). Such proposals have been rejected by other scholars (Anderson & Waxman 2013; Arkin 2010; Baker 2022; Dinstein & Dahl 2020; Homayounnejad 2019; Kershner 2013; Muller & Simpson 2014; Schmitt 2013), including this author, and were unable to attract necessary diplomatic support to become law. Whether one considers these efforts fundamentally misguided or noble yet regrettably unsuccessful, finding an alternative to blanket prohibition seems necessary unless one is willing to risk the prospect of AWS being completely unregulated. I take it to be uncontroversial that some regulation of AWS development, proliferation and use should take place, whether through an international treaty (or several such treaties), domestic law or internal military regulations (Zajac & Bober 2017).

In this article I am going to discuss a number of ideas on AWS regulation that should be introduced or restated in the debate as a part of a larger palette of options. Some of these solutions are largely uncontroversial and almost universally acceptable, and the

main problem consists of them not being seen, or regarded, in separation from much more controversial proposals. Others are almost as restrictive as a ban would be, but devoid of key difficulties. Taken together, these regulatory ideas should be considered both a starting point and a reference frame for more productive, more urgent and more substantial discussion of the AWS' future, while simultaneously demonstrating the breadth of solutions available to a regulationist policymaker, in contrast with her prohibitionist counterpart.

2. Low Hanging Fruit Regulatory Norms

In this section I will introduce the Reader to the low hanging fruit of AWS regulation – the politically easy, conceptually and technologically straightforward fixes to the most dire threats connected to AWS development, use and proliferation. These include: a ban on arming AWS with nuclear weapons and conventional weapons of comparable yield; a moratorium on automating positions higher up the chain of command; a moratorium on anti-personnel use of AWS; a moratorium on AWS miniaturization; a moratorium on in-combat learning and a ban on development, possession, trading and use of AWS by non-state actors. While the items on this list would not, either separately or jointly, suffice to avoid all the controversies connected with eventual AWS introduction, they would do much to address the most potent and most uncontroversially justified concerns. In this section I will explain which of the AWS-related concerns each of these items would accommodate or at least mollify, before proceeding to making my case for adopting these as soon as possible, instead of holding them hostage to the passage of much stricter, and much more controversial prohibitionist measures.

The ban on arming AWS with nuclear weapons and other weapons capable of producing similar effects is the most obvious and uncontroversial of the regulatory solutions one could possibly propose. Reasonable persons may argue whether the power to attack certain military targets should be granted to AWS of a certain kind; no one argues that AI agents should have the power to end human civilization as we know it. The AWS debate is a discussion on automating decision-making on the level of an individual infantryman or fighter pilot, not on the level of commander-in-chief of a nuclear superpower, questionable arguments for automating nuclear retaliation notwithstanding. To automate nuclear warfare would be risking the very survival of humanity for no discernible reason at all – literally none of the pro-AWS arguments apply in the case of nuclear weapons, if only because these are meant to never be used. Neither does anybody stand to benefit from such a move, while everybody would be immensely and immediately endangered by it. Consequently even the paradigmatically immoral state of North Korea should be willing to restrain itself in this manner. An immediate declaration of all nuclear powers to this effect, followed by a binding treaty, should be the very first item on the regulatory agenda, and one suspects that it is not so only because the scenario itself is considered to absurd and outlandish to be trying to preempt. Nevertheless, explicitly removing it as possibility would have value, if only by assuring the public opinion sound limits will be placed on the scope of the AWS revolution.

Bans or long-term moratoria on arming AWS with very high-yield conventional weapons or automating combat tasks currently overseen by mid- or high-ranking officers and their staffs would answer the very same concerns and follow the same logic. Placing a cap on the potency of weapons controlled by a given AWS lowers the uppermost limit of the damage it could do, and so the amount of risk taken by placing it in autonomous mode. A ban on automating positions higher up the chain of command would also limit the firepower placed at the machine's discretion, while at the same time limiting the area of machine's discretion to the lower level of tactical tasks, rather than, say, company or equivalent level of operations¹. We are talking about the levels of command and prerogatives few if any have an appetite or need for automating in the discernible future. Moreover, the tasks in question require high performance general intelligence, such as drafting local rules of engagement, that are likely to stay out of the reach of AI in near to mid-term. Absent a situation in which WMD-like use of indiscriminate AWS swarms would provide truly decisive military advantage, to be discussed in a subsequent section, there exists no incentive to automate these levels of command, and many reasons for not doing so, making a regulatory norm against it likely to gain universal acceptance.

A moratorium on targeting human personnel has less to do with limiting the amount of power delegated to AWS (they would still be, on some occasions, allowed to directly kill people) but with limiting the intellectual and moral difficulty of the tasks assigned to them. Limiting AWS either to attacking targets explicitly marked by human commanders, or to military objectives by nature (mostly weapon platforms such as tanks, warships and warplanes) would markedly lower the targeting difficulties faced by the machines. Even a uniformed human combatant may become non-targetable in the course of the battle for a number of reasons; most military platforms, meanwhile, either remain targetable even after their occupants surrender or are incapacitated or, like naval vessels, surrender in a machine-recognizable way. As most of the military advantage to be realized by using AWS would be realized by targeting equipment rather than enemy combatants as such, eschewing the latter would be a small price to pay for greatly simplifying the problem of LOAC-compliance and avoiding a public relations disaster that would follow the shooting of a surrendering or visibly incapacitated enemy.

Another moratorium to be proposed would prevent states from enabling their AWS to conduct in-combat learning, that is, changing their algorithms in the hope of improving them in light of their experience, either in an individual AWS or fleet-wide. While in-combat learning could yield huge rewards for the side implementing it, as the machines would steadily improve their combat performance; however, it would also mean that behaviors not tested for their permissibility under LOAC could be introduced. While the pressures to implement in-combat learning could be substantial, especially in prolonged conflicts, and detecting violations of such a moratorium would be difficult, with

¹ In many air forces even the lowest ranking pilots are officers, so automating the operations of an individual fighter would necessarily automate an officers position. This, however, would not be tantamount to automating a position higher up the chain of command.

detrimental implications for the prospects of its enforcement, LOAC compliance would most probably require in-combat learning to remain off-limits until significant progress in real-time AI explainability is not achieved.

The final two low hanging fruit regulatory solutions have to do with limiting the access of non-state actors to the novel and extremely potent class of weapons AWS' are shaping up to be. A ban or moratorium on their miniaturization below a certain size (to be set by persons more knowledgeable than this author) would prevent militaries from stockpiling weapons that could easily be transported covertly and used for unpredictable mass casualty attacks. Simply put, smuggling a tank is harder than smuggling an ATGM, which in turn is harder than smuggling a penknife. A pen-knife sized AWS, even if capable of killing a single person, could come to be, if mass produced, a weapon impossible to contain, enabling and perhaps even strategically favoring, out-of-the-blue first strikes against civilian targets and sensitive infrastructure. There is, and there should be, no place for such a weapon among civilized nations.

Non-state actors can be expected to acquire AWS on their own, without intentional or unintentional help from a state military or weapons industry – a point that eludes many prohibitionists. Constructing a crude, indiscriminate AWS – say, an autonomous delivery drone with a time-bomb strapped on – with off-the shelf component would not be that problematic for a terrorist organization, nor even for a lone wolf terrorist right now. As drones and robots will become more of a staple of modern life, this will grow yet easier. Governments need to react to this challenge, just as they need to react to advances in 3-D printing or biotechnology making producing other kinds of home-spun weapons easier. Given their potential to cause mass casualties against an unsuspecting, undefended crowd, any attempts to weaponize drones and robots, to own such systems, trade in them etc. should be treated as analogous actions are treated in the case of high explosives – as serious offenses in and of themselves, tantamount to terrorist activity. Countries should introduce legislation to this effect and work towards a level of uniformity of both sanction and enforcement in this respect. While preventing the proliferation of home-made AWS is not what most people have in mind when they talk about their banning or regulation, preventing governments only from using or abusing AWS while private citizens gain increasing access to such weapons is hardly the state of affairs to be aimed for.

Most items on this list are moratoria, rather than bans; this reflects the uncertainties regarding AWS and the underlying AI technologies, chiefly machine learning and its explainability problem. Given what we know in 2022 about the state of the art in AI and the history of its development, believing the tougher LOAC-compliance problems will never be solved is as much irresponsible speculation about the future of technology as is believing that they certainly will be solved; we simply do not know, and cannot know, if the solutions are possible, and if they are possible, if they will be stumbled upon. Banning some type of machine action forever because we cannot figure out how to compliantly conduct it now would foreclose potential benefits which would be realized if LOAC compliance is eventually achieved; yet permitting this actions before it is demonstrably

achieved is also not an option. Hence my preference for renewable, multi-year moratoria, which can be responsive to the changes in the technological state-of-the-art.

This list of bans and moratoria on AWS sub-types is substantially longer, and stricter, than a minimal compromise solution proposed three years ago by a group of prominent scholars featuring both AWS proponents and opponents (Arkin, Russell, Scharre et al. 2019). It is not as strict as the list of prohibitions recommended lately by the ICRC (2021), although it features recommendations that ICRC did not make, such as a moratorium on miniaturization and bans on use by non-state actors. What these three short-of-a-ban regulatory frameworks have in common is a desire to limit most controversial aspects of AWS use while leaving space for compliance oriented research-and-development and use in civilian scarce environments of high-end warfare where the pressure to make use of this militarily transformative technology will be the greatest.

The only reason a person caring about the humanitarian values could possibly have to oppose the immediate introduction of either some or all of the regulatory solutions outlined above would be a desire to promote even stricter regulation (or a ban) by binding them together with these commonsensical measures in the hope of advancing the former. The most needed and urgent measures would in effect be held hostage to ethically controversial and politically perilous parts of the anti-AWS agenda. The fifteen years of self-acknowledged failure by the prohibitionist camp show that this “no-plan-but-a-ban” agenda should not be considered politically savvy; what I would argue is that it also should be considered immoral to pursue it. The amount of suffering and damage caused by a completely unregulated AWS arms race would clearly be orders of magnitude greater than the amount of damage to be wrought by their restricted anti-platform use in high-end warfare. Gambling with the prospect of unregulated proliferation for the sake of realizing dubious and limited humanitarian gains, or protecting a controversial interpretation of human dignity is, I would argue, not just a tactical mistake, but a display of callous disregard for the actual success of arms control efforts. This is not to condemn those who wish to pursue stricter regulation than outlined above as pursuing an illegitimate goal; it is only to state that the passage of this politically palatable, ethically uncontroversial agenda cannot and should not be obstructed by making it conditional on the passage of stricter measures.

Indeed, stricter measures not only need not collide with this middle-of-the-road regulatory framework, they can be voluntarily adopted alongside on the basis of regional treaties, or serve as a back-up solution in case LOAC-compliance proves to be technologically impossible. I explore these options in the following sections.

3. Regional Bans

The ban advocates often tout the fact that so far as many as forty countries have endorsed it². Analyzing this list of endorsees may render several quite different conclusions. One

² These include: Algeria, Argentina, Austria, Bolivia, Brazil, Chile, China (on use only), Colombia, Costa Rica, Croatia, Cuba, Djibouti, Ecuador, Egypt, El Salvador, Ghana, Guatemala, the Holy See, Iraq, Jordan, Kazakhstan, Malta, Mexico,

may point out that only two countries on the list – Pakistan and China – are genuine military powers, with China’s endorsement being qualified and almost certainly insincere (Kania 2017), given that China was simultaneously developing drones with significant lethal autonomous capabilities (Trevithick 2019) and advertising them for export (Allen 2019, 6). And while a fair number of middle-rung military powers figure on the list, none of these are military technological leaders. A few of the endorsees (Costa Rica, Panama, the Vatican) do not even have standing armies, making their willingness to constrain themselves in this respect both backed by hard evidence and irrelevant. If one excludes China as a special case, one is justified to say that the list includes two kinds of states: those who do not expect to fight a regular war in the next two decades and those who expect to fight such a conflict against a technologically superior enemy. Countries whose decisions and example really matter – the military and technological leaders – are mostly absent from the list. This may lead one to conclude the list, far from being proof of the international mood being favorable to calls for the ban, is evidence of the opposite.

There is, however, another way of looking at the set of countries in question. If the ban endorsees are grouped by region, it is evident that in some major world regions prohibitionist views are much more popular, and even a majority position. Only four ban endorsing states are from Europe, with two being militarily insignificant micro-states. In contrast, as many as 13 endorsees come from Latin America – a decisive majority among the region’s 20 states including all of the region’s major military powers. While just 11 out of Africa’s 56 countries have declared support for the ban, these include militarily and regionally significant Algeria, Egypt, Morocco, Nigeria and South Africa.

Thus the analysis of the ban support’s geographic distribution shows potential for regional, continent-wide bans in Latin America and, to a lesser extent, Africa. For the remainder of this section I would like to show that such regional bans: have an established precedent; would fulfill at least some goals of the countries seeking the ban, as well as some other goals that are more universally shared; and would be game-theoretically stable and politically viable, or at least much more viable than a global ban.

The clear precedent for a continental/regional treaty effectively prohibiting the use of a militarily decisive weapon class is found in the African Nuclear-Weapon-Free-Zone Treaty, also known as the Treaty of Pelindaba³. The treaty, ratified as of August 2022 by 43 African states⁴, requires them “not to conduct research on, develop, manufacture, stockpile or otherwise acquire, possess or have control over any nuclear explosive device by any means anywhere” (Article 3 point 1), and prevents the stationing of any nuclear weapons on their soil. Having entered into force in 2009, the Pelindaba Treaty has since been abided by its parties and by all other African countries.

Morocco, Namibia, New Zealand, Nicaragua, Nigeria, Pakistan, Panama, Peru, Philippines, Sierra Leone, Spain, the State of Palestine, South Africa, Sri Lanka, Uganda, Venezuela, and Zimbabwe.
<https://www.hrw.org/news/2021/12/19/killer-robots-military-powers-stymie-ban>

³ <https://www.iaea.org/publications/documents/treaties/african-nuclear-weapon-free-zone-treaty-pelindaba-treaty>

⁴ <https://treaties.unoda.org/t/pelindaba>.

As evident from the fact that many of the parties to the Pelindaba Treaty are also parties to the worldwide Treaty on the Prohibition of Nuclear Weapons, the continent-wide ban on nuclear weapons does not fulfill all the goals of the African states seeking to ban them. This would also be the case of a regional AWS ban. A regional ban has, by its very nature, a limited scope; it can increase regional stability and prevent arms races between neighbors, but can do little to protect its parties from aggression by countries outside the weapon-x-free-zone and from spillover effects of arms development and use taking place elsewhere. It also does not fulfill the aim of shaping the humanitarian landscape outside the region – and many arms control and disarmament efforts are explicitly framed in terms of such universal aspirations.

However, as demonstrated by the example of the Treaty on the Prohibition of Nuclear Weapons, such aspirations frequently stay just that. In contrast, a regional treaty may not only be much more easily achievable but may accomplish a number of goals, including almost all of the goals that matter from the perspective of the more narrowly construed national interest and humanitarian needs of a region's citizens. Given that using AWS, unlike using nuclear weapons, would not directly affect persons living outside of a given warzone, the case for realizing these limited but tangible and nonetheless substantial goals is even more powerful.

A regional ban would also be conducive to the goals of countries from outside the region, just as the current African nuclear ban is. Even those nuclear-armed powers who have no intention of ridding themselves of their nuclear weapons acknowledge that universal nuclear proliferation would constitute a catastrophic outcome, both directly and indirectly by increasing the chances of nuclear weapons eventually falling into the hands of a malevolent non-state agent. Hence their support for the universal Treaty on the Non-Proliferation of Nuclear Weapons and the regional African arrangement. It may even be argued that it is the outside powers' interest in non-proliferation that set the conditions for the Treaty of Pelindaba's success. The outside powers have provided the diplomatic pressure and intelligence overwatch preventing African states from trying to acquire nuclear weapons, a feat of which even a broad coalition of African nations might not be capable themselves. Similarly nuclear blackmail or attack directed against an African state by a nuclear state has been successfully deterred by other nuclear powers interested in maintaining the nuclear taboo.

The incentives of the external powers align the same way in the case of a regional AWS ban, especially in the near to medium term. Even countries willing to aggressively pursue AWS development, deployment and use themselves are still most likely going to be interested in limiting their proliferation to countries deemed politically unstable and non-state agents. Multiple countries voluntarily eschewing AWS development would surely make these non-proliferation goals more likely to be realized (how likely and how costly meeting these goals would be all things considered is another matter). It is thus quite likely a regional ban would be welcome and even aided by outside powers, especially before AWS technology fully matures, and other regulatory frameworks aimed at non-proliferation goals solidify. While some powers may believe it in their interest to make

AWS part of their arms trade portfolio, it is doubtful this alone could make them act against the regional ban frameworks, and that these efforts at undermining regional frameworks would succeed. Supplying their own forces and those of geographically and politically closer allies would have priority for unscrupulous powers. This opposite incentive alone is not a sufficient reason to view regional bans as too fragile to be pursued.

That said, enforcing any AWS ban, even a local one, would be much more difficult than enforcing a ban on nuclear weapons. AWS will be much more easily acquired, concealed and possessed by a much higher number of outside actors (Altmann & Sauer 2017). Some of these actors might feel encouraged to meddle within the ban zone utilizing their AWS-based military superiority. Such actions would need to be deterred by a coalition of ban-abiding actors from within the zone, ban-friendly actors from outside, and most likely a combination of these. The use of AWS against a self-restricting actor would have to carry a similar stigma the use of nuclear weapons against a non-nuclear state. Developing these norms, defending them and preventing the collapse of a carefully crafted regional balance would be difficult, yet still much easier than sustaining a ban on a global scale. And if a regional ban framework were to fail after, say, twenty years, the benefits of these twenty years of arrested AWS development would still be substantial. Instead of being flooded by early, insufficiently LOAC-compliant AWS designs, the region would experience the AWS revolution in its more mature form.

All in all, whether one is worried about LOAC-incompliance, unsustainable arms races, runaway proliferation or unwilling to use AWS out of an absolutist moral conviction, a regional ban offers a less than perfect yet realistically attainable solution. The success of such a framework would be dependent on the actions of both the regional and extra-regional actors. However, both types of parties would generally have more to gain than to lose from abiding with the arrangement, and even temporary success of the framework might be of substantial value.

4. No First Use Policy⁵

No first use policies have been adopted, openly and tacitly, by a number of countries possessing weapons of mass destruction (WMD). No first use policy is a paradigmatic compromise between humanitarian concerns and military necessity; it is born of a realization that there are weapons too indiscriminate and/or inhumane to be used in warfare, but simultaneously powerful enough that to refuse to use them while the enemy does not do so is to eschew not only victory, but survival itself. This is certainly the case of nuclear weapons⁶. Their destructive potential has been successfully employed by global

⁵ I am thankful to professor Thomas Simpson for indicating this option to me.

⁶ It is also almost certainly not the case of chemical weapons. A war against an adversary employing chemical weapons on a mass scale could successfully be prosecuted by conventional means, provided friendly units were issued and used appropriate protective gear – as the Gulf War has demonstrated, a definite nuisance, but not a major problem. It can even be said that in such a case the user of chemical weapons would be at a disadvantage, investing scarce resources into a largely inefficient class of weapons. Biological weapons are also a non-obvious case, since their use threatens

and regional powers to deter each other from using them offensively, and to deter each other from using other kinds of weapons of mass destruction. In a world where nuclear weapons are a reality it makes moral and strategic sense to field them in order to prevent them from ever being used (Corbett 2021).

Why apply a policy designed for deterring WMD use to AWS? Some have claimed that AWS would necessarily be both deadly and indiscriminate enough to warrant the WMD label. This assertion is highly controversial and almost certainly wrong. Israeli counter-radar Harpy missile, which searches for a very specific targets and attacks them on their own, may plausibly be claimed to be an AWS, and it would be classified as such under the criteria adopted in this article. Still, no one claims that the Harpy is a weapon of mass destruction, and the ICRC does not consider it to be an indiscriminate AWS (2019, 10)⁷. It is certainly conceptually possible for an AWS to be precise and selective enough to attack only targets classified as military by nature with minimal collateral damage potential. Whether this conceptual possibility may become a technological one remains to be seen, but the Harpy seems to offer an early proof of the concept.

What is uncontroversial, however, is AWS' definite potential to be turned into weapons of mass destruction (Kallenborn 2020). Nobody could dispute that completely LOAC-incompliant AWS, deployed in large enough quantity, or with potent enough effectors, would be exactly that. One of the greatest challenges of the regulationist project is preventing the situation in which AWS of the WMD variety would see common use or, even worse, become a dominant weapon of the future battlefield⁸.

Why worry about the latter possibility? Being unrestrained by the limitations of human biology, AWS, if viable at all, would almost certainly greatly exceed top performance levels possible for human combatants. Never distracted, always acting at the upper limit of their capability, supernaturally strong and precise, considering themselves fully expendable and much more damage resistant, the robots would outperform human warfighters (or human drone operators) on the merits of any one of these advantages, let alone all these synergically combined. While humans would have some advantages over AWS, such as much superior general intelligence, the machines would not have to equal humans in all domains, just get tolerably proficient in each domain so that their own inherent advantages in some fields would outweigh their shortcomings in others. We already see them advancing to this level in the relatively uncluttered air and sea

their user as well, and laboratory leaks constitute an all too real possibility. The issue of whether keeping biological weapons as a deterrent against other biological weapon users would be prudent is rendered moot by the existence of nuclear weapons as a far more reliable alternative.

⁷ The ICRC lists enemy radar installations among military objectives by nature they explicitly state could be a valid AWS target, and the Harpy targets only these.

⁸ There is also a possibility of otherwise LOAC-compliant AWS being used to conduct mass casualty attacks against civilians, just like artillery or aircraft may be used today. However, the use of otherwise legal weapons in such a manner is already prohibited by LOAC.

environments, which coincidentally happen to be key theaters of contemporary warfare. Humans being made obsolete in some combat roles seems a very real possibility.

Were AWS to dominate in at least some combat environments, yet simultaneously remain LOAC-incompliant to a serious extent, state leaders would face essentially the same dilemma as at the dawn of the nuclear era. They would be unwilling to normalize AWS use for ethical and legal reasons, but at the same time they would not be able to afford rendering their countries defenseless against unscrupulous adversaries willing to use AWS despite the moral cost. To field AWS but pledge not to use them first would be a sensible solution in such a case. The only other alternative not involving either contemptuous disregard for law and ethics or effective surrender to less scrupulous adversaries would be deterring AWS use with nuclear weapons, an option burdened by even more serious downsides⁹.

If the conditions outlined above actually obtained, there seems to be no argument against the moral and strategic soundness of adopting a no-first-use policy towards AWS. Yet how likely is it that these conditions actually will obtain? I would argue that this is quite likely indeed. One reason is AWS' propensity to dominate at least some combat environments. It is also probably much easier to create an indiscriminate AWS – a killer robot – than to create a discriminate, LOAC-compliant one. An indiscriminate AWS can be as simple as a bomb with a timer strapped to a drone following a pre-programmed route – in fact, much more sophisticated AWS already exist and most probably have seen actual combat (UNSC Panel of Experts on Libya 2021, 17). LOAC-compliance, however, is another matter. Complying just with the principle of Distinction – discerning combatants from non-combatants – requires the capability to correctly perceive and analyze complex situations arising in combat, such as enemy combatants surrendering or non-combatants participating in hostilities. Compliance with other principles of *ius in bello*, such as the Principle of Proportionality, is even more difficult. While claims of LOAC-compliance being unachievable in principle are to be dismissed, it is true that full LOAC-compliance would at least require more sophisticated AI than will likely be available for some time, coupled with strict oversight by human officers. The longer the AWS would operate on their own, the more powerful the weapons at their disposal, and the greater their area of independent operations, the tougher compliance gets (ICRC 2019, 10).

Consequently a state of affairs in which LOAC-compliant AWS would still be unavailable, but indiscriminate, WMD-like AWS would already be able to overpower any conventional opposition is likely to transpire at some point. This state of affairs could be

⁹ Given their lasting environmental effects and the extremely painful deaths and injuries they cause to some of their victims, nuclear weapons would almost certainly be more indiscriminate and more inhumane than even the most indiscriminate AWS utilizing firearms or explosives as their armament. Using nuclear weapons against AWS to simply neutralize them within one's own territory would also be highly problematic, whereas even indiscriminate AWS could be used defensively against enemy AWS with much less collateral damage. Last but not least, any actual use of nuclear weapons could trigger a nuclear exchange, escalating AWS-based war into a nuclear one. Using AWS on AWS would not threaten that outcome.

temporary, if sufficient LOAC compliance was ultimately achievable, or permanent, if it proved lastingly elusive. In either scenario adopting a no first use policy coupled with a research effort aimed at achieving LOAC compliance as fast as possible would be ethically and strategically sound.

No first use policies may prove attractive to some actors even if the issue of LOAC compliance will be decisively solved at some point. Some governments may be persuaded by deontological arguments against using AWS even on legitimate military targets. This class of arguments, especially the sub-class of dignity-based objections, has the tendency to draw very different intuitions in experts and laymen alike (Baker 2022, 96-99); they also seem to be more popular in some countries, such as Germany, than in others, such as the US. It is thus reasonable to expect they might be adopted as core policy determinants by some, but not by all governments. Still, the AWS-eschewing countries would most probably like to preserve some defensive parity against potential AWS-wielding opponents, and so no-first-use pledge, perhaps coupled with a strictly defensive military posture and a non-interventionist foreign policy, may well prove the most appealing option.

No first use policy should be viewed as important tool of early stage AWS regulation – and as vital fallback position in case the more optimistic predictions about this technology’s potential fail. It would allow states not to actually use AWS while at the same time retaining military parity with their adversaries. Universal or almost universal adoption of no first use would actually result in a more stable version of a prohibitionist regime, with countries paying for developing and fielding weapons they never plan to use, but avoiding the strategic gamble on other nations compliance with the ban and the cost of monitoring or enforcing this compliance. Given the risks inherent in unilateral disarmament, the no first use framework simply offers a better deal, and one much more likely to receive support from key military powers. It also provides a degree of flexibility if previously elusive LOAC compliance is achieved, at which point no-first-use adopters would face a much smoother transition to AWS-wielding military, and no initial capability gap¹⁰. Knowing that no first use is an option consequently influences one’s approach to the ban-or-regulate debate, as it shows most benefits of the ban may be achieved by strict enough regulatory framework without the ban-related risks being incurred.

5. Conclusion.

For more than a decade ethical and legal debates about AWS have been dominated by controversial and unprecedented efforts to ban such a broad weapon class. These efforts were almost completely unsuccessful, in part because they have almost completely ignored the perspective of military necessity and the geopolitical context in which

¹⁰ Would not countries actively using AWS in real warfare hold an edge over countries merely fielding, but never actually using them? AWS nature takes an edge of this possible objection to no first use policy. AWS, after all, may be pitched against each other in real life combat scenarios in a way human troops cannot, allowing commanders to practice actual AWS-based combat on the proving ground.

decisions on AWS are being taken. Given the urgency of the issue and the catastrophic cost of inaction, it is high time to focus on a more nuanced, more politically palatable, and more easily enforceable regulatory approach, akin to that taken towards most other broad weapon classes. In this paper I discussed a number of regulatory options, dividing them into two categories.

The first category involved readily available regulatory solutions that would face little or no opposition from great military powers while addressing the biggest threats posed by unregulated proliferation of AWS. I have argued all humanitarians and arms control advocates have reason to support this regulatory package. I have also argued that trying to make one's support for this package conditional on the passage of much more restrictive measures would be tantamount to sabotaging urgently needed arms control agenda. The stakes are simply too high to indulge ideological purity at the expense of making tangible progress on the issue.

I have also discussed two solutions almost as restrictive as an actual AWS ban – regional bans and no first use policy. The existence of a regulatory toolbox containing such strict options weakens the case for an absolute ban, since choosing a ban over these would entail adopting very substantial risks in the name of much diminished and dubious gains. Discussing these options shows that the space of potential regulatory solutions is a rich spectrum, allowing persons greatly varying in their assessment of AWS negative and positive potential, the relative weight of military necessity and humanitarian concerns, and chances of requisite technological breakthroughs to be nevertheless relatively comfortable with some point on that spectrum.

It is entirely possible that some of the measures discussed in this paper will prove to be overly stringent and counterproductive from the perspective of humanitarian values. Were AWS to prove capable of compliance with LOAC, measures such as the no-first use policy or local bans on AWS would be as uncalled for and outright detrimental as such restrictions on using precision guided munitions would be today. This could also be the case of the moratorium on anti-personnel use. I do not want to preclude wider employment of AWS in the long term. My point is merely to show that such highly restrictive options are available if need be. For an AWS' enthusiast, they may not differ much from an outright prohibition; but for a person drawn to the prohibitionist case, but aware of its difficulties, this is their chief attraction. Just as prohibitionists are not justified in assuming their project will succeed, nor can regulationists simply assume they will succeed in making AWS LOAC-complaint. Regulatory frameworks of varying strictness need to be prepared, then used or left unused based on the regularly updated appraisal of the everchanging technological and geopolitical landscape. To prepare just one model is to set oneself up for failure.

Funding: This research was funded by National Science Centre, Poland grant number 2022/44/C/HS1/00051

Conflicts of Interest: The authors declares no conflict of interest.

References

- Allen (2019) – Allen, Gregory C. 2019. "Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security." Center for a New American Security, <https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNASUnderstanding-Chinas-AI-Strategy-Gregory-C.-Allen-FINAL-2.15.19.pdf>
- Altmann & Sauer (2017) – Altmann, Jurgen, and Frank Sauer. 2017. "Autonomous weapon systems and strategic stability." *Survival* 59.5: 117-142. <https://doi.org/10.1080/00396338.2017.1375263>
- Amoroso & Tamburrini (2017) – Amoroso, Daniele, and Guglielmo Tamburrini. 2017. "The ethical and legal case against autonomy in weapons systems." *Global Jurist* 18.1: 1-20. <https://doi.org/10.1515/gj-2017-0012>
- Anderson & Waxman (2013) – Anderson, Kenneth and Waxman, Matthew C. 2013. "Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can." Columbia Public Law Research Paper 13-351, April 13th 2013, <http://dx.doi.org/10.2139/ssrn.2250126>
- Arkin (2010) – Arkin, Ronald C. "The case for ethical autonomy in unmanned systems." *Journal of Military Ethics* 9.4: 332-341. <https://doi.org/10.1080/15027570.2010.536402>
- Arkin, Russell, Scharre et al. (2019) - Arkin, Ronald, Leslie Kaelbling, Stuart Russell, Dorsa Sadigh, Paul Scharre, Bart Selman and Toby Walsh. 2019. "Autonomous Weapon Systems, a Roadmapping Exercise", *Georgia Tech Robot Laboratory Online Publications*, September 9th, 2019, <https://www.cc.gatech.edu/ai/robot-lab/online-publications/AWS.pdf>.
- Asaro (2012) – Asaro, Peter. 2012. "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making." *International Review of the Red Cross* 94.886: 687-709. doi:10.1017/S1816383112000768
- Baker (2022) – Baker, Deane. 2022. *Should We Ban Killer Robots?* Cambridge, UK, Polity Press.
- Corbett (2020) – Corbett, Andrew. 2020. "Igitur qui desiderat pacem, praeparet bellum atomica." *Journal of Military Ethics* 19.4 : 331-347. <https://doi.org/10.1080/15027570.2021.1893461>
- Dinstein & Dahl (2020) – Dinstein, Yoram, and Arne Willy Dahl. 2020. *Oslo Manual on Select Topics of the Law of Armed Conflict: Rules and Commentary*. Springer Nature. https://library.oapen.org/bitstream/handle/20.500.12657/39564/2020_Book_OsloManualOnSelectTopicsOfThe.pdf.
- Human Rights Watch (2016) – Docherty, Bonnie, Joseph Crupi, Anna Khalfaoui, and Lan Mei. 2016. "Making the case: the dangers of killer robots and the need for a preemptive ban." Human Rights Watch Publication, 2016, https://www.hrw.org/sites/default/files/report_pdf/arms1216_web.pdf.
- Heyns (2017) – Heyns, Christof. 2017. "Autonomous weapons in armed conflict and the right to a dignified life: an African perspective." *South African journal on human rights* 33.1: 46-71. <https://hdl.handle.net/10520/EJC-cfba69d5e>
- Homayounnejad (2019) – Homayounnejad, Maziar. 2019. *Lethal Autonomous Weapon Systems Under the Law of Armed Conflict*. Doctoral Dissertation, King's College London, London, UK. https://kclpure.kcl.ac.uk/portal/files/110384075/2019_Homayounnejad_Maziar_0222601_thesis.pdf.
- ICRC (2021) – International Committee of the Red Cross. 2021. "ICRC position on Autonomous Weapon Systems and Background Paper", May 17th, 2021, <https://www.icrc.org/en/publication/455001-icrcposition-autonomous-weapon-systems>.
- Kallenborn (2020) – Kallenborn, Zachary. "Are Drone Swarms Weapons of Mass Destruction?" *The Counterproliferation Papers Future Warfare Series No. 60*, U.S. Air Force Center for Strategic Deterrence Studies, May 6th, 2020. <https://media.defense.gov/2020/Jun/29/2002331131/-1/-1/0/60DRONESWARMS-MONOGRAPH.PDF>.
- Kania (2017) – Kania, Elsa B. "Battlefield singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power." Center for a New American Security, November 2017, <https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/Battlefield-Singularity-November-2017.pdf>
- Kershner (2013) – Kershner, Stephen. "Autonomous weapons pose no moral problems." In: "Killing by remote control: The ethics of an unmanned military": 229-245.
- Muller & Simpson (2014) – Vincent C. Muller & Thomas W. Simpson, "Killer Robots: Regulate, Don't Ban." University of Oxford, Blavatnik School of Government Policy Memo. <https://philpapers.org/archive/MLLKRR.pdf>
- Sauer (2020) – Sauer, Frank. "Stepping back from the brink: Why multilateral regulation of autonomy in weapons systems is difficult, yet imperative and feasible." *International review of the red cross* 102.913 (2020): 235-259. <https://doi.org/10.1017/S1816383120000466>
- Schmitt (2013) – Schmitt, Michael. 2013. "Autonomous weapon systems and international humanitarian law: a reply to the critics." *Harvard National Security Journal* 4: 1-37. Official URL: <https://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/>
- Sparrow (2007) – Sparrow, Robert. 2007. "Killer robots." *Journal of applied philosophy* 24.1: 62-77. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>
- Sparrow (2016) – Sparrow, Robert. 2016. "Robots and respect: Assessing the case against autonomous weapon systems." *Ethics & international affairs* 30.1: 93-116. <https://doi.org/10.1017/S0892679415000647>
- Trevithick (2019) – Trevithick, Joseph. "China Is Hard At Work Developing Swarms Of Small Drones With Big Military Applications", *The Warzone*, June 30th 2019, <https://www.thedrive.com/the-war-zone/17698/chinas-is-hard-at-work-developing-swarms-of-small-drones-on-multiple-levels>

UNSC Panel of Experts on Libya (2021) – United Nations Security Council Panel of Experts on Libya. 2021. "Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011)." S/2021/229, March 8th, 2021, <https://undocs.org/S/2021/229>.

Zajac & Bober (2017) – Zajac, Maciej, and Wojciech J. Bober. 2017. "Autonomous Military Robots– critical topography of possible reactions to upcoming technological revolution." *Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Śląska* z. 110: 201-216. DOI: 10.29119/1641-3466.2017.110.20