# Lethal Autonomous Weapons Aff-Neg

## Topic Overview

Lethal Autonomous Weapons (LAWs, AWS, LARs) are a kind of military technology that utilizes artificial intelligence to be able to track, identify, and attack military targets. For the purposes of clarity it’s important to understand that while the most recent and common abbreviation of Lethal Autonomous Weapons is LAWs, it also goes by AWS (for Autonomous Weapons Systems), and LARs (for Lethal Autonomous Robots), this change is primarily for purpose of clarity.

Lethal Autonomous Weapons are only one kind of artificial intelligence that are used by the military, others include automated defense systems, tracking systems, etc. Advocates for banning LAWs focus predominately on the role that human agency plays in de-escalating situations, identifying civilians, and preventing unnecessary harm. Those who defend LAWs tend to be more optimistic on the progress that artificial intelligence will have in the future, which would provide autonomous weapons the ability to de-escalate situations more efficiently.

An important question for this topic is how effective banning weapons actually is. States often come out with different definitions of what they believe lethal autonomous weapons are, and fears about non-state actors developing them are a prevalent influence to their development. On the other hand, previous weapon systems and categories have been effectively banned, advocates point to the bans on chemical weapons, anti-personal landmines, and blinding laser weapons as key instances of bans working. Establishing the connection, or lack of connection, between these weapons and LAWs becomes important when debating the effectiveness of these bans.

There exists key philosophical differences that influence how advocates and opponents of LAW bans view the effectiveness, and need for prohibition. Those opposing bans tend to utilize a framework of ‘International Realism,’ which tends to analyze international relations based on the preferences of individual states – states often attempt to gain more power and influence which would cause them to see LAWs as politically important. Meanwhile advocates for LAW bans tend to place more emphasis on cooperation between states and international institutions (e.g. UN). While this cooperation isn’t entirely opposed by Realism, in the case of LAWs this cooperation would be seen as flawed because it’s in the states best interest to develop them because it could deter attacks from others who develop them, and give them a military edge against those who don’t.

## Further Reading

Umbrello, S., Torres, P., & De Bellis, A. F. (2020). The future of war: could lethal autonomous weapons make conflict more ethical?. AI & SOCIETY, 35(1), 273-282. Retrieved from: <https://www.researchgate.net/publication/330621132_The_Future_of_War_Could_Lethal_Autonomous_Weapons_Make_Conflict_More_Ethical>

Surber, R. (2018). Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats. ICT4Peace Foundation and the Zurich Hub for Ethics and Technology (ZHET) p, 1, 21. Retrieved from <https://ict4peace.org/wp-content/uploads/2018/02/2018_RSurber_AI-AT-LAWS-Peace-Time-Threats_final.pdf>

Wallach, W. (2017). Toward a ban on lethal autonomous weapons: surmounting the obstacles. Communications of the ACM, 60(5), 28-34. Retrieved from: <https://cacm.acm.org/magazines/2017/5/216318-toward-a-ban-on-lethal-autonomous-weapons/fulltext>

Roff, H. M., & Moyes, R. (2016, April). Meaningful human control, artificial intelligence and autonomous weapons. In Briefing Paper Prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, Geneva, Switzerland. Retrieved from: [www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf](http://www.article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf)

Warren, A., & Hillas, A. (2017). Lethal Autonomous Weapons Systems: Adapting to the Future Unmanned Warfare and Unaccountable Robots. Yale J. Int'l Aff., 12, 71. Retrieved from: yalejournal.org/wp-content/uploads/2017/08/2017a\_71\_hillas.pdf

# LAWs Aff

### Value-Criterion

#### For this debate I offer the value of Human Dignity

Schachter 83 Oscar Schachter was an American international law and diplomacy professor, and United Nations aide. "Human Dignity as a Normative Concept." Published by American Journal of International Law, Vol. 77(4), pp. 848-854. Published October 1983. Available here: (https://www.cambridge.org/core/services/aop-cambridge-core/content/view/F5C2D6F4C7A31D7DE2F6AD55670C24D4/S0002930000071074a.pdf/human\_dignity\_as\_a\_normative\_concept.pdf) - AP

The conception of respect for dignity suggested above can also be given more specific meaning by applying it to actions of psychological significance. Indeed, nothing is so clearly violative of the dignity of persons as treatment that demeans or humiliates them. This includes not only attacks on personal beliefs and ways of life but also attacks on the groups and communities with which individuals are affiliated. Official statements that vilify groups or hold them up to ridicule and contempt are an especially dangerous form of psychological aggression resulting in a lack of respect by others for such groups and, perhaps even more insidious, destroying or reducing the sense of selfrespect that is so important to the integrity of every human. We can also point to the widespread practice of using psychogenic drugs or other forms of psychological coercion to impose conformity and ideological obedience, These should clearly be seen as violations of the inherent dignity of the person. Put in positive terms, respect for the intrinsic worth of a person requires a recognition that the person is entitled to have his or her beliefs, attitudes, ideas and feelings. The use of coercion, physical or psychological, to change personal beliefs is as striking an affront to the dignity of the person as physical abuse or mental torture. Our emphasis on respect for individuals and their choices also implies proper regard for the responsibility of individuals. The idea that people are generally responsible for their conduct is a recognition of their distinct identity and their capacity to make choices. Exceptions may have to be made for those incapable of such choices (minors or the insane) or in some cases for those under severe necessity. But the general recognition of individual responsibility, whether expressed in matters of criminal justice or civic duties, is an aspect of the respect that each person merits as a person. It is also worth noting as a counterpart that restraint is called for in imputing responsibility to individuals for acts of others such as groups of which they are members. In general, collective responsibility is a denigration of the dignity of the individual, a denial of a person's capacity to choose and act on his or her responsibility. We do not by this last comment mean to separate individuals sharply from the collectivities of which they are a part. Indeed, we believe that the idea of 1983] EDITORIAL COMMENT 851 human dignity involves a complex notion of the individual. It includes recognition of a distinct personal identity, reflecting individual autonomy and responsibility. It also embraces a recognition that the individual self is a part of larger collectivities and that they, too, must be considered in the meaning of the inherent dignity of the person. We can readily see the practical import of this conception of personality by considering political orders that, on the one hand, arbitrarily override individual choice and, on the other, seek to dissolve group ties. There is also a "procedural" implication in that it indicates that every individual and each significant group should be recognized as having the capacity to assert claims to protect their essential dignity.

#### As a criterion I present the prevention of unnecessary suffering

Linklater 02 Andrew Linklater FAcSS is an international relations academic, and is the current Woodrow Wilson Professor of International Politics at Aberystwyth University. "The Problem of Harm in World Politics: Implications for the Sociology of States-Systems." Published by International Affairs, Vol 78(2), pp. 319-338. Published April 2002. Available here: (https://www-jstor-org.ezproxy.lib.utah.edu/stable/pdf/3095684.pdf) - AP

These cosmopolitan orientations are central to the sociological approach to harm in world politics which I want to outline here. What is most interesting from this point of view is how far different international systems have thought harm to individuals a moral problem for the world as a whole-a problem which all states, individually and collectively, should labour to solve-and have developed what might be called cosmopolitan harm conventions. These are moral conventions designed to protect individuals everywhere from unnecessary suffering, irrespective of their citizenship or nationality, class, gender, race and other distinguishing characteristics.6 In world politics, unnecessary suffering or superfluous injury-expressions used in the Hague Conventions-usually result from state-building, conquest and war, from the globalization of economic and social relations, and from pernicious racist, nationalist and related doctrines. The function of a sociology of harm conventions is to ask to what extent different states-systems drew on the idea of a universal community of humankind to create agreements that individuals should be protected from the harm such phenomena cause-moral conventions which reveal that human sympathies need not be confined to co- nationals or fellow citizens but can be expanded to include all members of the human race.

### Escalation

#### Human intervention is necessary to prevent escalating violence – obedient AI weapons only worsen the situation

Coeckelbergh et al. 18 Mark Coeckelbergh is a Belgian philosopher of technology. He is Professor of Philosophy of Media and Technology at the Department of Philosophy of the University of Vienna and former President of the Society for Philosophy and Technology. Janina Loh is a postdoctoral researcher for the Institute of Philosophy at the University of Vienna. Michael Funk is a professor of Philosophy at the University of Vienna, with research focusing on the Philosophy of Media and Technology. Johanna Seibt is a regular faculty at Aarhus University, Department of Philosophy and specializes in the areas of analytical ontology and metaphysics; most recently she works in robophilosophy. Marco Nørskov is an Assistant Professor at the Department of Philosophy and History of Ideas at Aarhus University, Denmark. "Envisioning Robots in Society – Power, Politics, and Public Space : Proceedings of Robophilosophy 2018 / TRANSOR 2018." Published by IOS Press in 2018. Available here: (<http://web.b.ebscohost.com.ezproxy.lib.utah.edu/ehost/ebookviewer/ebook?sid=8f3d98ec-1aae-4e83-a800-7e2b617e3535%40pdc-v-sessmgr03&vid=0&format=EB>) - AP

5. The Exclusive Focus on Dispositional Explanations The abstraction from the pragmatic features of killings in war also has a striking impact on the understanding of problems of noncompliance with Jus in Bello and International Humanitarian Law. A prominent argument for robotic combatants is that they “in principle” comply much better than human combatants, prone as they are to well-known but inconvenient and counterproductive emotional responses on the battlefield: fear, panic, stress, burn out, guilt, hatred, vengeance, bloodlust etc. We often see such dispositions and emotions presented as the sole explanation of war crimes and other inappropriate conduct. An obvious example is Arkin when he argues that we ought to reduce human inhumanity by replacing human combatants with robotic ones [10]. Robotic combatants can perform effectively in battlefield scenarios that would make human combatants panic and stress and they would abide by their pre-programmed rules of engagement no matter how many of their “comrades” they see dismembered next to them. There are at least two reasons to be deeply skeptical about Arkin’s solution. First, we need much more pragmatic context to get an adequate understanding of war crimes and other immoral behavior than Arkin’s narrow dispositional perspective. There is strong evidence from social psychology that atrocious behavior is correlated (and indeed caused) by situational and systemic factors to a much higher degree that the dispositional factors favoured by Arkin. Second, if war crimes are not typically due to decisions of the direct perpetrator but instigated by commanders or politicians it does not seem to be the best solution to have mindless and obedient robots to execute the orders. To see this, let us take a brief look at a concrete example of human malfunction in stressful and hostile environments: the case of the Abu Ghraib detainee abuse and torture. Philip Zimbardo has argued convincingly that essential parts of the explanation of these atrocities is to found in situational and systemic factors. In fact, there was no evidence that the direct perpetrators harbored any sadistic or psychopathic dispositions. But they were subject to a work environment of extreme overload, stress, and danger: “[...] twelve-hour night shifts (4:00 p.m. to 4:00 a.m.), seven days a week for forty days [...] unsanitary and filthy surroundings that smelled like a putrid sewer [...] frequent insurgency attacks. Five U.S. soldiers and twenty prisoners were killed, and many others were wounded by almost daily shelling [...]” [11]. Zimbardo argues, convincingly I think, that the abuse and torture was a result of burn out and total moral disorientation due to these situational factors. More than that, we have to understand these situations in light of the instructions and directives about prisoner treatment from commanders and political authorities. These instructions were at best vague and suggestive and at worst direct orders to “soften up” prisoners during and in between interrogations. No doubt, robots would “cope” much better with such conditions but you may wonder if that would really be an improvement. Human break down can be a very important indicator that something is more fundamentally wrong in the broader context in which the break down takes place. Arkin’s mistake is to take for granted that LAWS will be deployed only in an ideal situation of wise and just political leadership. It seems to me a real risk that a robotic “fix” will only reinforce an atmosphere and rhetoric of dehumanization and total enmity that dominates much of the “war on terror.” At least, the symbolic message seems clear: we do not even bother to put real human beings in harm’s way to deal with these people. 6. Conclusion A typical move by apologists for LAWS is to claim that the many problematic features of robotic warfare are merely contingent and something we can ignore in ideal theorizing. My objection is that this construal of ideal theory simply ignores too much to be relevant. Even on ideal terms, a theory should still be feasible. In the real world, we need some sort of legal framework for armed conflicts in order to be confident that we know what we are doing, morally speaking. It will not do that some philosopher king might be able to declare a war good and just. For certain actions to feasible (e.g. in reasonable compliance with the principles of distinction, liability, immunity, humane treatment etc.) we need an institutional framework. LAWS that observe the ideal morality of war are feasible only in so far an institutional framework supporting the norms of the ideal morality is in place. The question now is how feasible is that? The main reason why I think the answer is “not really” is that the basic strategy of LAWS is a response to a type of conflict in which essential parties to the conflict will have little if any motivation for compliance. I quoted Seneca above for the view that a sword never kills anybody. Seneca was a wise man and this was actually not a view he held himself. He ascribed it to “certain men”. Here is another quote, which is probably closer to his own view: “Arms observe no bounds; nor can the wrath of the sword, once drawn, be easily checked or stayed; war delights in blood.” (The Madness of Hercules, lines 403-05)

#### Machine Decision Making characteristic of LAWs escalates conflict further

Wong et al 20 Yuna Huh Wong is a policy researcher at the RAND Corporation. Her research interests include scenario development, futures methods, wargaming, problem-structuring methods, and applied social science. John M. Yurchak is a senior information scientist at the RAND Corporation who focuses on defense-related analysis. Robert W. Button is an adjunct senior researcher at the RAND Corporation. His research interests include artificial intelligence and simulation. Aaron Frank is a senior information scientist at the RAND Corporation. He specializes in the development of analytic tradecraft and decision-support tools for assessing complex national security issues. Burgess Laird is a senior international researcher at the RAND Corporation. His subject-matter areas of expertise are defense strategy and force planning, deterrence, and proliferation. Osonde A. Osoba is an information scientist at the RAND Corporation. Randall Steeb is a senior engineer at the RAND Corporation. Benjamin N. Harris is an adjunct defense analyst at the RAND Corporation and a student at the Massachusetts Institute of Technology, where he is pursuing a Ph.D. in political science. Sebastian Joon Bae is a defense analyst at the RAND Corporation. His research interests include wargaming, counterinsurgency, hybrid warfare, violent nonstate actors, emerging technologies, and the nature of future warfare. "Deterrence in the Age of Thinking Machines. Published by the Rand Corporation in 2020. Available here: (https://apps.dtic.mil/sti/pdfs/AD1090316.pdf) – AP

How Escalatory Dynamics May Change In this section, we explore some more general ideas prompted by the wargame that also have the potential to affect deterrent and escalatory dynamics. We hypothesize that the different mixes of humans and artificial agents in different roles can affect the escalatory dynamics between two sides in a crisis. We also examine how signaling and understanding, important elements to successful deterrence, could be adversely affected with the introduction of machine decisionmaking. Decisionmaking and Presence One insight from our wargame is that the differences in the ways two sides configure their human versus machine decisionmaking and their manned versus unmanned presence could affect escalatory dynamics during a crisis. In the wargame, confrontations occurred between unmanned U.S. forces with humans-in-the-loop decisionmaking and Chinese forces that were manned but had more humans-on-the-loop and humans-out-of-the-loop decisionmaking. These confrontations appeared to put the onus on U.S. forces to deescalate the situation and inspired Table 7.2. We hypothesize the further ways that mixes of human and machine could result in different escalatory dynamics. In the upper left of Table 7.2, we propose that when systems are manned and the decisionmaking is primarily done by humans, there is a lower escalatory dynamic. We argue that humans in the decisionmaking process have time to slow down how quickly things can escalate, but that the presence of humans means that there is a higher cost to miscalculating events, because human lives could be lost. This quadrant represents the most common situation today. 64 Deterrence in the Age of Thinking Machines In the lower left-hand quadrant, there are primarily unmanned systems with primarily human decisionmaking—this may be the least escalatory combination of all. Having humans in the loop again slows down the decision cycle compared with configurations that are more heavily driven by machine decisions, which may mean more time to consider deescalatory offramps during a crisis. Humans may also be better at understanding signaling. Additionally, having mostly unmanned systems lowers the risk to human life, as the consequences of miscalculating are destroyed systems but not loss of human lives. This is the quadrant that best represents the United States in the wargame in Chapter Five. In the upper right is a situation in which systems are manned but decisions are made mostly by machines (humans on the loop or humans out of the loop). We argue that this is the most escalatory situation of all. With more decisions happening at machine speeds, there is likely a greater risk of inadvertent escalation during a crisis. However, the presence of humans means that there is the higher risk to human life with miscalculation and escalation. This is where the notional, future Chinese forces were in the wargame. In the final, lower right-hand quadrant, we see the combination of unmanned systems and machine decisionmaking. This is perhaps what the public imagines futuristic war will be like one day. We argue that this has a higher escalatory dynamic because of the rapid machine decisionmaking, but the costs of miscalculation are lower because human lives are not at risk.

#### Miscalculations and escalation due to LAWs makes the likelihood of increasing violence and even nuclear warfighting more likely

Laird 20 Burgess Laird is a senior international defense researcher at the nonprofit, nonpartisan RAND Corporation. He is a contributing author of Deterrence in the Age of Thinking Machines. "The Risks of Autonomous Weapons Systems for Crisis Stability and Conflict Escalation in Future U.S.-Russia Confrontations." Published by the Rand Corporation on June 3, 2020. Available here: (https://www.rand.org/blog/2020/06/the-risks-of-autonomous-weapons-systems-for-crisis.html) - AP

While holding out the promise of significant operational advantages, AWS simultaneously could increase the potential for undermining crisis stability and fueling conflict escalation. First, a state facing an adversary with AWS capable of making decisions at machine speeds is likely to fear the threat of sudden and potent attack, a threat that would compress the amount of time for strategic decisionmaking. The posturing of AWS during a crisis would likely create fears that one's forces could suffer significant, if not decisive, strikes. These fears in turn could translate into pressures to strike first—to preempt—for fear of having to strike second from a greatly weakened position. Similarly, within conflict, the fear of losing at machine speeds would be likely to cause a state to escalate the intensity of the conflict possibly even to the level of nuclear use. Second, as the speed of military action in a conflict involving the use of AWS as well as hypersonic weapons and other advanced military capabilities begins to surpass the speed of political decisionmaking, leaders could lose the ability to manage the crisis and with it the ability to control escalation. With tactical and operational action taking place at speeds driven by machines, the time for exchanging signals and communications and for assessing diplomatic options and offramps will be significantly foreclosed. However, the advantages of operating inside the OODA loop of a state adversary like Iraq or Serbia is one thing, while operating inside the OODA loop of a nuclear-armed adversary is another. As the renowned scholar Alexander George emphasized (PDF), especially in contests between nuclear armed competitors, there is a fundamental tension between the operational effectiveness sought by military commanders and the requirements for political leaders to retain control of events before major escalation takes place. Third, and perhaps of greatest concern to policymakers should be the likelihood that, from the vantage point of Russia's leaders, in U.S. hands the operational advantages of AWS are likely to be understood as an increased U.S. capability for what Georgetown professor Caitlin Talmadge refers to as “conventional counterforce” operations. In brief, in crises and conflicts, Moscow is likely to see the United States as confronting it with an array of advanced conventional capabilities backstopped by an interconnected shield of theater and homeland missile defenses. Russia will perceive such capabilities as posing both a conventional war-winning threat and a conventional counterforce threat (PDF) poised to degrade the use of its strategic nuclear forces. The likelihood that Russia will see them this way is reinforced by the fact that it currently sees U.S. conventional precision capabilities precisely in this manner. As a qualitatively new capability that promises new operational advantages, the addition of AWS to U.S. conventional capabilities could further cement Moscow's view and in doing so increase the potential for crisis instability and escalation in confrontations with U.S. forces. In other words, the fielding of U.S. AWS could augment what Moscow already sees as a formidable U.S. ability to threaten a range of important targets including its command and control networks, air defenses, and early warning radars, all of which are unquestionably critical components of Russian conventional forces. In many cases, however, they also serve as critical components of Russia's nuclear force operations. As Talmadge argues, attacks on such targets, even if intended solely to weaken Russian conventional capabilities, will likely raise Russian fears that the U.S. conventional campaign is in fact a counterforce campaign aimed at neutering Russia's nuclear capabilities. Take for example, a hypothetical scenario set in the Baltics in the 2030 timeframe which finds NATO forces employing swarming AWS to suppress Russian air defense networks and key command and control nodes in Kaliningrad as part of a larger strategy of expelling a Russian invasion force. What to NATO is a logical part of a conventional campaign could well appear to Moscow as initial moves of a larger plan designed to degrade the integrated air defense and command and control networks upon which Russia's strategic nuclear arsenal relies. In turn, such fears could feed pressures for Moscow to escalate to nuclear use while it still has the ability to do so. Finally, even if the employment of AWS does not drive an increase in the speed and momentum of action that forecloses the time for exchanging signals, a future conflict in which AWS are ubiquitous will likely prove to be a poor venue both for signaling and interpreting signals. In such a conflict, instead of interpreting a downward modulation in an adversary's operations as a possible signal of restraint or perhaps as signaling a willingness to pause in an effort to open up space for diplomatic negotiations, AWS programmed to exploit every tactical opportunity might read the modulation as an opportunity to escalate offensive operations and thus gain tactical advantage. Such AWS could also misunderstand adversary attempts to signal resolve solely as adversary preparations for imminent attack. Of course, correctly interpreting signals sent in crisis and conflict is vexing enough when humans are making all the decisions, but in future confrontations in which decisionmaking has willingly or unwillingly been ceded to machines, the problem is likely only to be magnified. Concluding Thoughts Much attention has been paid to the operational advantages to be gained from the development of AWS. By contrast, much less attention has been paid to the risks AWS potentially raise. There are times in which the fundamental tensions between the search for military effectiveness and the requirements of ensuring that crises between major nuclear weapons states remain stable and escalation does not ensue are pronounced and too consequential to ignore. The development of AWS may well be increasing the likelihood that one day the United States and Russia could find themselves in just such a time. Now, while AWS are still in their early development stages, it is worth the time of policymakers to carefully consider whether the putative operational advantages from AWS are worth the potential risks of instability and escalation they may raise.

Cooley and Nexon 20 ALEXANDER COOLEY is Claire Tow Professor of Political Science at Barnard College and Director of Columbia University's Harriman Institute. DANIEL H. NEXON is an Associate Professor in the Department of Government and at the Edmund A. Walsh School of Foreign Service at Georgetown University. "How Hegemony Ends: The Unraveling of American Power." Foreign Affairs, vol. 99, no. 4, July-Aug. 2020, p. 143+. Available here: (https://go-gale-com.ezproxy.lib.utah.edu/ps/i.do?id=GALE%7CA626467013&v=2.1&u=marriottlibrary&it=r&p=AONE&sw=w) - AP

CONSERVING THE U.S. SYSTEM

Great-power contestation, the end of the West's monopoly on patronage, and the emergence of movements that oppose the liberal international system have all altered the global order over which Washington has presided since the end of the Cold War. In many respects, the COVID-19 pandemic seems to be further accelerating the erosion of U.S. hegemony. China has increased its influence in the World Health Organization and other global institutions in the wake of the Trump administration's attempts to defund and scapegoat the public health body. Beijing and Moscow are portraying themselves as providers of emergency goods and medical supplies, including to European countries such as Italy, Serbia, and Spain, and even to the United States. Illiberal governments worldwide are using the pandemic as cover for restricting media freedom and cracking down on political opposition and civil society. Although the United States still enjoys military supremacy, that dimension of U.S. dominance is especially ill suited to deal with this global crisis and its ripple effects.

Even if the core of the U.S. hegemonic system--which consists mostly of long-standing Asian and European allies and rests on norms and institutions developed during the Cold War--remains robust, and even if, as many champions of the liberal order suggest will happen, the United States and the European Union can leverage their combined economic and military might to their advantage, the fact is that Washington will have to get used to an increasingly contested and complex international order. There is no easy fix for this. No amount of military spending can reverse the processes driving the unraveling of U.S. hegemony. Even if Joe Biden, the presumptive Democratic nominee, knocks out Trump in the presidential election later this year, or if the Republican Party repudiates Trumpism, the disintegration will continue.

The key questions now concern how far the unraveling will spread. Will core allies decouple from the U.S. hegemonic system? How long, and to what extent, can the United States maintain financial and monetary dominance? The most favorable outcome will require a clear repudiation of Trumpism in the United States and a commitment to rebuild liberal democratic institutions in the core. At both the domestic and the international level, such efforts will necessitate alliances among center-right, center-left, and progressive political parties and networks.

What U.S. policymakers can do is plan for the world after global hegemony. If they help preserve the core of the American system, U.S. officials can ensure that the United States leads the strongest military and economic coalition in a world of multiple centers of power, rather than finding itself on the losing side of most contests over the shape of the new international order. To this end, the United States should reinvigorate the beleaguered and understaffed State Department, rebuilding and more effectively using its diplomatic resources. Smart statecraft will allow a great power to navigate a world defined by competing interests and shifting alliances.

The United States lacks both the will and the resources to consistently outbid China and other emerging powers for the allegiance of governments. It will be impossible to secure the commitment of some countries to U.S. visions of international order. Many of those governments have come to view the U.S.-led order as a threat to their autonomy, if not their survival. And some governments that still welcome a U.S.-led liberal order now contend with populist and other illiberal movements that oppose it.

Even at the peak of the unipolar moment, Washington did not always get its way. Now, for the U.S. political and economic model to retain considerable appeal, the United States has to first get its own house in order. China will face its own obstacles in producing an alternative system; Beijing may irk partners and clients with its pressure tactics and its opaque and often corrupt deals. A reinvigorated U.S. foreign policy apparatus should be able to exercise significant influence on international order even in the absence of global hegemony. But to succeed, Washington must recognize that the world no longer resembles the historically anomalous period of the 1990s and the first decade of this century. The unipolar moment has passed, and it isn't coming back.

### Arms Race

#### Autonomous weapons harms the already fragile stability between the US, Russia, and China causing arms race

Wong et al 20 Yuna Huh Wong is a policy researcher at the RAND Corporation. Her research interests include scenario development, futures methods, wargaming, problem-structuring methods, and applied social science. John M. Yurchak is a senior information scientist at the RAND Corporation who focuses on defense-related analysis. Robert W. Button is an adjunct senior researcher at the RAND Corporation. His research interests include artificial intelligence and simulation. Aaron Frank is a senior information scientist at the RAND Corporation. He specializes in the development of analytic tradecraft and decision-support tools for assessing complex national security issues. Burgess Laird is a senior international researcher at the RAND Corporation. His subject-matter areas of expertise are defense strategy and force planning, deterrence, and proliferation. Osonde A. Osoba is an information scientist at the RAND Corporation. Randall Steeb is a senior engineer at the RAND Corporation. Benjamin N. Harris is an adjunct defense analyst at the RAND Corporation and a student at the Massachusetts Institute of Technology, where he is pursuing a Ph.D. in political science. Sebastian Joon Bae is a defense analyst at the RAND Corporation. His research interests include wargaming, counterinsurgency, hybrid warfare, violent nonstate actors, emerging technologies, and the nature of future warfare. "Deterrence in the Age of Thinking Machines. Published by the Rand Corporation in 2020. Available here: (https://apps.dtic.mil/sti/pdfs/AD1090316.pdf) – AP

Autonomous systems may also affect the credibility of deterrent threats.1 States with autonomous systems might appear more credible when making deterrent threats than states without them.2 Nonetheless, as with other conventional weapons, opponents who do not possess autonomous systems will not simply accede to the deterrent or coercive threats of states that do have them. Instead, they will develop strategies, operational approaches, and capabilities designed to counter, avoid, or mitigate the advantages of autonomous systems. When confronting states that do possess autonomous systems of their own, using autonomous systems could come to be seen as low-risk and thus attractive means for mounting probing attacks against adversaries. This could result in “salami” tactics employed to slice away at the adversary’s interests without overtly crossing a threshold or red line that invites the opponent to strike back. Widespread AI and autonomous systems could also make escalation and crisis instability more likely by creating dynamics conducive to rapid and unintended escalation of crises and conflicts. This is because of how quickly decisions may be made and actions taken if more is being done at machine, rather than human, speeds. Inadvertent escalation could be a real concern. In protracted crises and conflicts between major states, such as the United States and China, there may be strong incentives for each side to use such autonomous capabilities early and extensively, both to gain coercive and military advantage and to attempt to prevent the other side from gaining advantage.3 This would raise the possibility of first-strike instability. AI and autonomous systems may also reduce strategic stability. Since 2014, the strategic relationships between the United States and Russia and between the United States and China have each grown far more strained. Countries are attempting to leverage AI and develop autonomous systems against this strategic context of strained relations. By lowering the costs or risks of using lethal force, autonomous systems could make the use of force easier and more likely and armed conflict more frequent.4 A case may be made that AI and autonomous systems are destabilizing because they are both transformative and disruptive. We can already see that systems such as UAVs, smart munitions, and loitering weapons have the potential to alter the speed, reach, endurance, cost, tactics, and burdens of fielded units. Additionally, AI and autonomous systems could lead to arms race instability. An arms race in autonomous systems between the United States and China appears imminent and will likely bring with it the instability associated with arms races. Finally, in a textbook case of the security dilemma, the proliferation of autonomous systems could ignite a serious search for countermeasures that exacerbate uncertainties and concerns that leave countries feeling less secure.

#### An AI Arms race would lead to systemic vulnerabilities that could create catastrophic damage – even if the US wins an AI Race the damage could be irreversible

Scharre 19 Paul Scharre is a Senior Fellow and Director of the Technology and National Security Program at the Center for a New American Security. "Killer Apps: The Real Dangers of an AI Arms Race." Published by Foreign Affairs Vol. 98(3). Published May-June 2019. Available here: (https://go-gale-com.ezproxy.lib.utah.edu/ps/retrieve.do?tabID=T003&resultListType=RESULT\_LIST&searchResultsType=SingleTab&hitCount=1&searchType=AdvancedSearchForm&currentPosition=1&docId=GALE%7CA585763278&docType=Essay&sort=RELEVANCE&contentSegment=ZONE-MOD1&prodId=AONE&pageNum=1&contentSet=GALE%7CA585763278&searchId=R1&userGroupName=marriottlibrary&inPS=true) - AP

The nation that leads in the development of artificial intelligence will, Russian President Vladimir Putin proclaimed in 2017, "become the ruler of the world." That view has become commonplace in global capitals. Already, more than a dozen governments have announced national AI initiatives. In 2017, China set a goal of becoming the global leader in AI by 2030. Earlier this year, the White House released the American AI Initiative, and the U.S. Department of Defense rolled out an AI strategy. But the emerging narrative of an "AI arms race" reflects a mistaken view of the risks from AI--and introduces significant new risks as a result. For each country, the real danger is not that it will fall behind its competitors in AI but that the perception of a race will prompt everyone to rush to deploy unsafe AI systems. In their desire to win, countries risk endangering themselves just as much as their opponents. AI promises to bring both enormous benefits, in everything from health care to transportation, and huge risks. But those risks aren't something out of science fiction; there's no need to fear a robot uprising. The real threat will come from humans. Right now, AI systems are powerful but unreliable. Many of them are vulnerable to sophisticated attacks or fail when used outside the environment in which they were trained. Governments want their systems to work properly, but competition brings pressure to cut corners. Even if other countries aren't on the brink of major AI breakthroughs, the perception that they're rushing ahead could push others to do the same. And if a government deployed an untested AI weapons system or relied on a faulty AI system to launch cyberattacks, the result could be disaster for everyone involved. Policymakers should learn from the history of computer networks and make security a leading factor in AI design from the beginning. They should also ratchet down the rhetoric about an AI arms race and look for opportunities to cooperate with other countries to reduce the risks from AI. A race to the bottom on AI safety is a race no one would win. THE AIS HAVE IT The most straightforward kind of AI system performs tasks by following a series of rules set in advance by humans. These "expert systems," as they are known, have been around for decades. They are now so ubiquitous that we hardly stop to think of the technology behind airplane autopilots or tax-preparation software as AI. But in the past few years, advances in data collection, computer processing power, and algorithm design have allowed researchers to make big progress with a more flexible AI method: machine learning. In machine learning, a programmer doesn't write the rules; the machine picks them up by analyzing the data it is given. Feed an algorithm thousands of labeled photos of objects, and it will learn to associate the patterns in the images with the names of the objects. The current AI boom began in 2012, when researchers made a breakthrough using a machine-learning technique called "deep learning," which relies on deep neural networks. Neural networks are an AI technique loosely inspired by biological neurons, the cells that communicate with other cells by sending and receiving electrical impulses. An artificial neural network starts out as a blank slate; it doesn't know anything. The system learns by adjusting the strength of the connections between neurons, strengthening certain pathways for right answers and weakening the connections for wrong answers. A deep neural network--the type responsible for deep learning--is a neural network with many layers of artificial neurons between the input and output layers. The extra layers allow for more variability in the strengths of different pathways and thus help the AI cope with a wider variety of circumstances. How exactly the system learns depends on which machine-learning algorithm and what kind of data the developers use. Many approaches use data that are already labeled (known as "supervised learning"), but machines can also learn from data that are not labeled ("unsupervised learning") or directly from the environment ("reinforcement learning"). Machines can also train on synthetic, computer-generated data. The autonomous car company Waymo has driven its cars for over ten million miles on public roads, but the company clocks ten million miles every day in computer simulations, allowing it to test its algorithms on billions of miles of synthetic data. Since the deep-learning breakthrough in 2012, researchers have created AI systems that can match or exceed the best human performance in recognizing faces, identifying objects, transcribing speech, and playing complex games, including the Chinese board game go and the real-time computer game StarCraft. Deep learning has started to outstrip older, rules-based AI systems, too. In 2018, a deep-learning algorithm beat the reigning chess computer program after spending just four hours playing millions of games against itself on a massive supercomputer without any human training data or hand-coded rules to guide its behavior. Researchers are now applying AI to a host of real-world problems, from diagnosing skin cancers to driving cars to improving energy efficiency. According to an estimate by the consulting firm McKinsey, almost half of all the tasks people are paid to perform in the United States could be automated with existing technology (although less than five percent of jobs could be eliminated entirely). AI tools are also becoming more widely available. Large organizations are the most likely to make major breakthroughs, thanks to their ability to amass large data sets and huge quantities of computing power. But many of the resulting AI tools are available online for anyone to use. Free programming courses teach people how to make their own AI systems, and trained neural networks are free to download. Accessibility will spur innovation, but putting powerful AI tools into the hands of anyone who wants them will also help those who set out to do harm. AUTOCRATIC INTELLIGENCE Harm from AI misuse isn't hypothetical; it's already here. Bots are regularly used to manipulate social media, amplifying some messages and suppressing others. Deepfakes, AI-generated fake videos, have been used in so-called revenge porn attacks, in which a person's face is digitally grafted onto the body of a pornographic actor. These examples are only the start. Political campaigns will use AI-powered data analytics to target individuals with political propaganda tailored just for them. Companies will use the same analytics to design manipulative advertising. Digital thieves will use AI tools to create more effective phishing attacks. Bots will be able to convincingly impersonate humans online and over the phone by cloning a person's voice with just a minute of audio. Any interaction that isn't in person will become suspect. Security specialists have shown that it's possible to hack into autonomous cars, disabling the steering and brakes. Just one person could conceivably hijack an entire fleet of vehicles with a few keystrokes, creating a traffic jam or launching a terrorist attack. AI's power as a tool of repression is even more frightening. Authoritarian governments could use deepfakes to discredit dissidents, facial recognition to enable round-the-clock mass surveillance, and predictive analytics to identify potential troublemakers. China has already started down the road toward digital authoritarianism. It has begun a massive repression campaign against the Muslim Uighur population in Xinjiang Province. Many of the tools the government is using there are low tech, but it has also begun to use data analytics, facial recognition systems, and predictive policing (the use of data to predict criminal activity). Vast networks of surveillance cameras are linked up to algorithms that can detect anomalous public behavior, from improperly parked vehicles to people running where they are not allowed. The Chinese company Yuntian Lifei Technology boasts that its intelligent video surveillance system has been deployed in nearly 80 Chinese cities and has identified some 6,000 incidents related to "social governance." Some of the ways in which Chinese authorities now use AI seem trivial, such as tracking how much toilet paper people use in public restrooms. Their proposed future uses are more sinister, such as monitoring patterns of electricity use for signs of suspicious activity. China is not just building a techno-dystopian surveillance state at home; it has also begun exporting its technology. In 2018, Zimbabwe signed a deal with the Chinese company CloudWalk Technology to create a national database of faces and install facial recognition surveillance systems at airports, railway stations, and bus stops. There's more than money at stake in the deal. Zimbabwe has agreed to let Cloud-Walk send data on millions of faces back to China, helping the company improve its facial recognition systems for people with dark skin. China also plans to sell surveillance technology in Malaysia, Mongolia, and Singapore. China is exporting its authoritarian laws and policies, too. According to Freedom House, China has held training sessions with government officials and members of the media from over 30 countries on methods to monitor and control public opinion. Three countries--Tanzania, Uganda, and Vietnam--passed restrictive media and cybersecurity laws soon after engaging with China. WHAT AI WILL DO Whichever country takes the lead on AI will use it to gain economic and military advantages over its competitors. By 2030, AI is projected to add between $13 trillion and $15 trillion to the global economy. AI could also accelerate the rate of scientific discovery. In 2019, an artificial neural network significantly outperformed existing approaches in synthetic protein folding, a key task in biological research. AI is also set to revolutionize warfare. It will likely prove most useful in improving soldiers' situational awareness on the battlefield and commanders' ability to make decisions and communicate orders. AI systems can process more information than humans, and they can do it more quickly, making them valuable tools for assessing chaotic battles in real time. On the battlefield itself, machines can move faster and with greater precision and coordination than people. In the recent AI-versus-human StarCraft match, the AI system, AlphaStar, displayed superhuman abilities in rapidly processing large amounts of information, coordinating its units, and moving them quickly and precisely. In the real world, these advantages will allow AI systems to manage swarms of robots far more effectively than humans could by controlling them manually. Humans will retain their advantages in higher-level strategy, but AI will dominate on the ground. Washington's rush to develop AI is driven by a fear of falling behind China, which is already a global powerhouse in AI. The Chinese technology giants Alibaba, Baidu, and Tencent rank right alongside Amazon, Google, and Microsoft as leading AI companies. Five of the ten AI startups with the most funding last year were Chinese. Ten years ago, China's goal of becoming the global leader in AI by 2030 would have seemed fanciful; today, it's a real possibility. Equally alarming for U.S. policymakers is the sharp divide between Washington and Silicon Valley over the military use of AI. Employees at Google and Microsoft have objected to their companies' contracts with the Pentagon, leading Google to discontinue work on a project using AI to analyze video footage. China's authoritarian regime doesn't permit this kind of open dissent. Its model of "military-civil fusion" means that Chinese technology innovations will translate more easily into military gains. Even if the United States keeps the lead in AI, it could lose its military advantage. The logical response to the threat of another country winning the AI race is to double down on one's own investments in AI. The problem is that AI technology poses risks not just to those who lose the race but also to those who win it. THE ONLY WINNING MOVE IS NOT TO PLAY Today's AI technologies are powerful but unreliable. Rules-based systems cannot deal with circumstances their programmers did not anticipate. Learning systems are limited by the data on which they were trained. AI failures have already led to tragedy. Advanced autopilot features in cars, although they perform well in some circumstances, have driven cars without warning into trucks, concrete barriers, and parked cars. In the wrong situation, AI systems go from supersmart to superdumb in an instant. When an enemy is trying to manipulate and hack an AI system, the risks are even greater. Even when they don't break down completely, learning systems sometimes learn to achieve their goals in the wrong way. In a research paper last year, a group of 52 AI researchers recounted dozens of times when AI systems showed surprising behavior. An algorithm learning to walk in a simulated environment discovered it could move fastest by repeatedly falling over. A Tetris-playing bot learned to pause the game before the last brick fell, so that it would never lose. One program deleted the files containing the answers against which it was being evaluated, causing it to be awarded a perfect score. As the researchers wrote, "It is often functionally simpler for evolution to exploit loopholes in the quantitative measure than it is to achieve the actual desired outcome." Surprise seems to be a standard feature of learning systems. Machine-learning systems are only ever as good as their training data. If the data don't represent the system's operating environment well, the system can fail in the real world. In 2018, for example, researchers at the MIT Media Lab showed that three leading facial recognition systems were far worse at classifying dark-skinned faces than they were at classifying light-skinned ones. When they fail, machine-learning systems are also often frustratingly opaque. For rules-based systems, researchers can always explain the machine's behavior, even if they can't always predict it. For deep-learning systems, however, researchers are often unable to understand why a machine did what it did. Ali Rahimi, an AI researcher at Google, has argued that much like medieval alchemists, who discovered modern glassmaking techniques but did not understand the chemistry or physics behind their breakthroughs, modern machine-learning engineers can achieve powerful results but lack the underlying science to explain them. Every failing of an AI system also presents a vulnerability that can be exploited. In some cases, attackers can poison the training data. In 2016, Microsoft created a chatbot called Tay and gave it a Twitter account. Other users began tweeting offensive messages at it, and within 24 hours, Tay had begun parroting their racist and anti-Semitic language. In that case, the source of the bad data was obvious. But not all data-poisoning attacks are so visible. Some can be buried within the training data in a way that is undetectable to humans but still manipulates the machine. Even if the creators of a deep-learning system protect its data sources, the system can still be tricked using what are known as "adversarial examples," in which an attacker feeds the system an input that is carefully tailored to get the machine to make a mistake. A neural network classifying satellite images might be tricked into identifying a subtly altered picture of a hospital as a military airfield or vice versa. The change in the image can be so small that the picture looks normal to a human but still fools the AI. Adversarial examples can even be placed in physical objects. In one case, researchers created a plastic turtle with subtle swirls embedded in the shell that made an object identification system think it was a rifle. In another, researchers placed a handful of small white and black squares on a stop sign, causing a neural network to classify it as a 45-mile-per-hour speed-limit sign. To make matters worse, attackers can develop these kinds of deceptive images and objects without access to the training data or the underlying algorithm of the system they are trying to defeat, and researchers have struggled to find effective defenses against the threat. Unlike with cybersecurity vulnerabilities, which can often be patched once they are uncovered, there is no known way of fully inoculating algorithms against these attacks.

#### Banning LAWs now is key – stopping development before they’re completed prevents other actors from redoubling their efforts

Piper 19 Kelsey Piper is a Staff Writer for Vox's new vertical with a focus on the global poor, animal welfare, and risks affecting a stable future for our world. "Death by algorithm: the age of killer robots is closer than you think." Published by Vox on June 21, 2019. Available here: (https://www.vox.com/2019/6/21/18691459/killer-robots-lethal-autonomous-weapons-ai-war) - AP

For one thing, if LAWS development continues, eventually the weapons might be extremely inexpensive. Already today, drones can be purchased or built by hobbyists fairly cheaply, and prices are likely to keep falling as the technology improves. And if the US used drones on the battlefield, many of them would no doubt be captured or scavenged. “If you create a cheap, easily proliferated weapon of mass destruction, it will be used against Western countries,” Russell told me. Lethal autonomous weapons also seem like they’d be disproportionately useful for ethnic cleansing and genocide; “drones that can be programmed to target a certain kind of person,” Ariel Conn, communications director at the Future of Life Institute, told me, are one of the most straightforward applications of the technology. Then there are the implications for broader AI development. Right now, US machine learning and AI is the best in the world, which means that the US military is loath to promise that it will not exploit that advantage on the battlefield. “The US military thinks it’s going to maintain a technical advantage over its opponents,” Walsh told me. That line of reasoning, experts warn, opens us up to some of the scariest possible scenarios for AI. Many researchers believe that advanced artificial intelligence systems have enormous potential for catastrophic failures — going wrong in ways that humanity cannot correct once we’ve developed them, and (if we screw up badly enough) potentially wiping us out. In order to avoid that, AI development needs to be open, collaborative, and careful. Researchers should not be conducting critical AI research in secret, where no one can point out their errors. If AI research is collaborative and shared, we are more likely to notice and correct serious problems with advanced AI designs. And most crucially, advanced AI researchers should not be in a hurry. “We’re trying to prevent an AI race,” Conn told me. “No one wants a race, but just because no one wants it doesn’t mean it won’t happen. And one of the things that could trigger that is a race focused on weapons.” If the US leans too much on its AI advantage for warfare, other countries will certainly redouble their own military AI efforts. And that would create the conditions under which AI mistakes are most likely and most deadly. What people are trying to do about it In combating killer robots, researchers point with optimism to a ban on another technology that was rather successful: the prohibition on the use of biological weapons. That ban was enacted in 1972, amid advances in bioweaponry research and growing awareness of the risks of biowarfare. Several factors made the ban on biological weapons largely successful. First, state actors didn’t have that much to gain by using the tools. Much of the case for biological weapons was that they were unusually cheap weapons of mass destruction — and access to cheap weapons of mass destruction is mostly bad for states. Opponents of LAWS have tried to make the case that killer robots are similar. “My view is that it doesn’t matter what my fundamental moral position is, because that’s not going to convince a government of anything,” Russell told me. Instead, he has focused on the case that “we struggled for 70-odd years to contain nuclear weapons and prevent them from falling in the wrong hands. In large quantities, [LAWS] would be as lethal, much cheaper, much easier to proliferate” — and that’s not in our national security interests. The Campaign to Stop Killer Robots works to persuade policymakers that lethal autonomous weapons should be banned internationally. Courtesy of Campaign to Stop Killer Robots But the UN has been slow to agree even to a debate over a lethal autonomous weapons treaty. There are two major factors at play: First, the UN’s process for international treaties is generally a slow and deliberative one, while rapid technological changes are altering the strategic situation with regard to lethal autonomous weapons faster than that process is set up to handle. Second, and probably more importantly, the treaty has some strong opposition. The US (along with Israel, South Korea, the United Kingdom, and Australia) has thus far opposed efforts to secure a UN treaty opposing lethal autonomous weapons. The US’s stated reason is that since in some cases there could be humanitarian benefits to LAWS, a ban now before those benefits have been explored would be “premature.” (Current Defense Department policy is that there will be appropriate human oversight of AI systems.) Opponents nonetheless argue that it’s better for a treaty to be put in place as soon as possible. “It’s going to be virtually impossible to keep [LAWS] to narrow use cases in the military,” Javorsky argues. “That’s going to spread to use by non-state actors.” And often it’s easier to ban things before anyone has them already and wants to keep the tools they’re already using. So advocates have worked for the past several years to bring up LAWS for debate in the UN, where the details of a treaty can be hammered out.

### Civilian Casualities

#### Lethal Autonomous weapons are prone to error and intentional misuse that puts civilians at risk

Javorsky, Tegmark, and Hefland 19 Emilia Javorsky MD, MPH. Physician-Scientist at Arctic Fox, Co-Founded Sundaily, Former Researcher Harvard Medical School. Max Erik Tegmark is a Swedish-American physicist, cosmologist and machine learning researcher. ra Helfand, MD is co-chair of PSR's Nuclear Weapons Abolition Committee and also serves as co-president of PSR's global federation. "Lethal Autonomous Weapons." Published by BMJ on March 25, 2019. Available here: (http://oadoi.org/10.1136%2Fbmj.l1171) - AP

It’s not too late to stop this new and potentially catastrophic force Advances in artificial intelligence are creating the potential to develop fully autonomous lethal weapons.1 These weapons would remove all human control over the use of deadly force. The medical community has a long history of advocacy against the development of lethal weapons, and the World and American Medical Associations both advocate total bans on nuclear, chemical, and biological weapons.2 But while some nations and non-governmental organisations have called for a legally binding ban on these new weapons,34 the medical community has been conspicuously absent from this discourse. Third revolution in warfare Several countries are conducting research to develop lethal autonomous weapons. Many commentators have argued that the development of lethal autonomous weapon systems for military use would represent a third revolution in warfare, after the invention of gunpowder and nuclear weapons.5 Although semi-autonomous weapons, such as unmanned drones, are in widespread use, they require human oversight, control, and decision making to ensure, at least in theory, that targets are ethically and legally legitimate. In contrast, lethal autonomous weapon systems are defined as: “any system capable of targeting and initiating the use of potentially lethal force without direct human supervision and direct human involvement in lethal decision making.”6 In other words, they represent the complete automation of lethal harm. Once developed, such weapons could be produced rapidly, cheaply, and at scale.7 Furthermore, lethality will only increase with use as the machine’s learning algorithms gain access to more data. Without human decision making capability, autonomous weapons have great potential to target civilians in error or malfunction in other ways with no clarity around responsibility and justifiability. These weapons could quickly become ubiquitous on black markets and readily accessible to groups acting outside international laws. Professional voice The practical, legal, and ethical ramifications of dehumanising lethal warfare, combined with a high risk of both unintentional and intentional misuse, have amplified calls for an international ban on lethal autonomous weapons and a requirement for meaningful human control of all weapons systems.35 Healthcare professionals must engage in this conversation.

#### LAWs inability to grasp necessary complex situational awareness makes identifying civilians nearly impossible

Rosert and Sauer 19 Elvira Rosert is Junior Professor for International Relations at Universität Hamburg and at the Institute for Peace Research and Security Policy. Frank Sauer is a Senior Researcher at Bundeswehr University Munich. His work covers nuclear issues, terrorism, cyber‐security as well as emerging military technologies. "Prohibiting Autonomous Weapons: Put Human Dignity First." Published by the University of Durham and John Wiley & Sons. Published July 5, 2019. Available here: (https://onlinelibrary.wiley.com/doi/full/10.1111/1758-5899.12691) - AP

Lethal autonomous weapons systems: a threat to human dignity Numerous arguments motivate the current call for an international, legally binding ban on so‐called lethal autonomous weapons systems (LAWS).1 Strategic concerns include proliferation, arms races and escalation risks (Altmann and Sauer, 2017; Rickli, 2018). Military concerns include the incompatibility of LAWS with a traditional chain of command or the potential for operational failures cascading at machine speed (Bode and Huelss, 2018; Scharre, 2016). Ethical concerns include the fear that LAWS might further increase the dehumanization and abstractness of war (and thus its propensity), as well as its cruelty if warfare is delegated to machines incapable of empathy or of navigating in dilemmatic situations (Krishnan, 2009; Sauer and Schörnig, 2012; Sparrow, 2015; Sparrow et al., 2019; Wagner, 2014). Legal concerns include difficulties of attribution, accountability gaps, and limits to the fulfillment of obligatory precautionary measures (Brehm, 2017; Chengeta, 2017; Docherty, 2015). But the most prominent concern, focalizing some elements of the concerns just mentioned, is the danger these weapons pose to civilians. This argument's legal underpinning is the principle of distinction – undoubtedly one of the central principles of International Humanitarian Law (IHL), if not the central principle (Dill, 2015). As multifaceted and complex as the debate on military applications of autonomy is now, what has been articulated at its very beginning (Altmann and Gubrud, 2004; Sharkey, 2007) and consistently since then is that LAWS would violate IHL due to their inability to distinguish between combatants and civilians. This image of LAWS as a threat to civilians is echoed routinely and placed first by all major ban supporters (we substantiate this claim in the following section). That LAWS would be incapable of making this crucial distinction – and thus have to be considered indiscriminate – is assumed because ‘civilian‐ness’ is an under‐defined, complex and heavily context‐dependent concept that is not translatable into software (regardless of whether the software is based on rules or on machine learning). Recognizing and applying this concept on the battlefield not only requires value‐based judgments but also a degree of situational awareness as well as an understanding of social context that current and foreseeable computing technology does not possess. We unequivocally share this view as well as these concerns. And yet, in this article, we propose to de‐emphasize the indiscriminateness frame in favor of a deeper ethical assertion, namely that the use of LAWS would infringe on human dignity. The minimum requirement for upholding human dignity, even in conflicts, is that life and death decisions on the battlefield should always and in principle be made by humans (Asaro, 2012; Gubrud, 2012). Not the risk of (potential) civilian harm, but rather retaining meaningful human control to preserve human dignity should be at the core of the message against LAWS.2 Our proposal rests on normative considerations and strategic communication choices. In the remainder of this article, we elaborate on two basic lines of our argument, namely the IHL principle of distinction and the concept of human dignity, provide insights into how and why they have been mobilized in the global debate on LAWS, and discuss the benefits and challenges of putting our proposal into practice. LAWS and the principle of distinction Modern IHL identifies three different categories of persons: combatants, non‐combatants, and civilians. Those members of the armed forces who directly participate in hostilities count as combatants; those members who do not directly participate (e.g. military clergy) count as non‐combatants; and persons who do not belong to the armed forces count as civilians (Aldrich, 2000; Ipsen, 2008). These distinctions bring into being one major principle for the conduct of hostilities: Only members of the armed forces constitute legitimate targets, whereas civilians must never be deliberately made a target of attack (Best, 1991). With regard to the use of certain means and methods of combat, the prohibition of indiscriminate attacks implies a prohibition of indiscriminate weapons. Weapons may be deemed indiscriminate if they cannot be targeted at specific and discrete military objects, if they produce effects which cannot be confined to military objects during or after the use of the weapon, or if they are typically not targeted at specific objects despite being capable of precise targeting in principle (Baxter, 1973; Blix, 1974). This general principle has surfaced in several weapon prohibitions. First, indiscriminateness is a constitutive feature of the entire category of weapons of mass destruction (WMD). As of recently, each of these weapons – biological, chemical, and nuclear weapons – have been explicitly prohibited by a separate treaty.3 Second, several conventional weapons have been restricted or prohibited due to their indiscriminate effects, the treaties prohibiting anti‐personnel (AP) landmines (1997) and cluster munitions (2008) being the two most recent and most prominent examples. The processes resulting in these two prohibitions function as procedural and substantial precedents for the ongoing norm‐setting efforts on LAWS. In procedural terms, all three processes share their formal institutional origins in the United Nations (UN) Convention on Certain Conventional Weapons (CCW), and all were championed by NGO coalitions. In the cases of AP landmines and cluster munitions, the CCW's failure to reach an agreement provoked eventually successful processes conducted by like‐minded states outside the UN framework. The issue of LAWS initially gained traction within the UN framework in the Human Rights Council (HRC); it then moved to the CCW, where it has been debated since 2014, first in informal talks, and, since 2016, in a group of governmental experts (GGE), which used to spend 2 weeks’ time on the issue but has reduced the allotted time to 7 days in 2019. Yet, due to the lack of progress and the more or less open resistance to any regulation attempt by some major states, leaving the CCW is yet again being discussed. What is of more interest to us, though, is the substantial impact of previous ban campaigns on the framing of LAWS. The campaign against AP landmines succeeded in achieving the first complete ban on a conventional weapon by coining the image of AP landmines as ‘indiscriminate, delayed‐action weapons that cannot distinguish between a soldier and an innocent civilian’ (Price, 1998, p. 628). Some years later, the ban on cluster bombs was grafted onto this existing stigma by drawing an analogy between landmines and unexploded submunitions killing civilians long after the end of conflicts (Petrova, 2016; Rosert, 2019). As mentioned at the outset of this article, the legal argument against LAWS is more complex and also involves issues such as accountability and precautions in attack. Nevertheless, the frame of ‘indiscriminateness’, which has worked out well twice in the past, has been salient since the earliest warnings against LAWS and remains a focal point of the ongoing pro‐ban discourse, especially in communication from the international Campaign to Stop Killer Robots. Shortly after its formation in 2009, the International Committee for Robot Arms Control (ICRAC)4 announced in the first sentence of its foundational ‘Berlin Statement’ that such weapons systems ‘pose [pressing dangers] to peace and international security and to civilians’.5 When Human Rights Watch (HRW) embarked on the issue in 2011, the question most interesting to them was whether LAWS were ‘inherently indiscriminate’; when Article36 – an NGO advocating humanitarian disarmament, with civilian protection at its core – became another champion of a ban, the link between LAWS and civilian harm was further strengthened (Carpenter, 2014). The then UN Special Rapporteur on extrajudicial, summary, or arbitrary executions, Christof Heyns, emphasized the ‘specific importance’ of the ‘rules of distinction and proportionality’, and pointed out that the ability of LAWS to ‘operate according to these rules’ will likely be impeded (Heyns, 2013, pp. 12–13). Launched in fall 2012, the Campaign to Stop Killer Robots coordinated by HRW also placed special emphasis on the protection of civilians from the very beginning: ‘The rules of distinction, proportionality, and military necessity are especially important tools for protecting civilians from the effects of war, and fully autonomous weapons would not be able to abide by those rules. […] The requirement of distinction is arguably the bedrock principle of international humanitarian law’ (HRW, 2012, pp. 3, 24). While listing various risks raised by LAWS, the focus on civilians is still the most prominent element in the campaign's framing of the issue today. LAWS are diagnosed with a lack ‘of the human judgment necessary to evaluate the proportionality of an attack [and] distinguish civilian from combatant’, and are considered particularly prone to ‘tragic mistakes’ that would ‘shift the burden of conflict even further on to civilians’ (CSKR 2019a). The aim of sparing civilians from the effects of armed conflict is commendable, and we wholeheartedly support it. In relation to the specific case of LAWS, however, this legacy focus on IHL and civilian harm risks obscuring the much deeper ethical problem of delegating the decision to kill to machines. The LAWS problematique thus goes far beyond the question of whether a machine will be able to comply with the principle of distinction or not. We elaborate on this argument in the following section.

#### Civilian deaths are not an incidental feature of LAWS – they’re goal of assymetric warfare uniquely places civilians at risk

Coeckelbergh et al. 18 Mark Coeckelbergh is a Belgian philosopher of technology. He is Professor of Philosophy of Media and Technology at the Department of Philosophy of the University of Vienna and former President of the Society for Philosophy and Technology. Janina Loh is a postdoctoral researcher for the Institute of Philosophy at the University of Vienna. Michael Funk is a professor of Philosophy at the University of Vienna, with research focusing on the Philosophy of Media and Technology. Johanna Seibt is a regular faculty at Aarhus University, Department of Philosophy and specializes in the areas of analytical ontology and metaphysics; most recently she works in robophilosophy. Marco Nørskov is an Assistant Professor at the Department of Philosophy and History of Ideas at Aarhus University, Denmark. "Envisioning Robots in Society – Power, Politics, and Public Space : Proceedings of Robophilosophy 2018 / TRANSOR 2018." Published by IOS Press in 2018. Available here: (<http://web.b.ebscohost.com.ezproxy.lib.utah.edu/ehost/ebookviewer/ebook?sid=8f3d98ec-1aae-4e83-a800-7e2b617e3535%40pdc-v-sessmgr03&vid=0&format=EB>) - AP

3.1. Aggression Based on the experiences from the US and Israeli drone programs and what we learn from projected LAWS it seems obvious that they are designed for aggressive rather than defensive purposes. Their combination of surveillance and “surgical” lethal functions signifies their purpose. They are designed for targeted killings of persons who have been under surveillance for an extended period. This widens the “window of opportunity” for engaging in preventive attacks significantly. Intelligence agencies and governments typically portray the terrorists targeted in drone attacks as posing an ongoing, unavoidable threat due to their mere existence. Therefore, killing them whenever there is an opportunity to do so is justified as an act of self- and other-defence. Especially dubious are the widespread, so-called “signature strikes” based on “patterns of life analyses”. A US Senior Administration Official recently described the choice of targets as involving “[a] variety of signatures, from the information and intelligence that in some ways is unique to the US government, for example [...] to the extent an individual’s activities are analogous to those traditionally performed by a military” (Guardian 1 July 2016, emphasis mine). Now, second-guessing the militant or otherwise violent nature of a target hardly constitutes proof of guilt beyond reasonable doubt. Neither does it seem to meet the principle of discrimination. However, the future criteria for such signatory first strikes could be even more oblique, possibly generated by artificially intelligent systems in ways unfathomable for mere humans. 3.2. Asymmetry LAWS, e.g. autonomous military drones with their own systems for target selection and rules of engagement are designed for conflicts of extreme asymmetry, since they are ill equipped to deal with enemies with symmetrical anti-aircraft defence capabilities. Extreme asymmetry, however, can be highly counterproductive not least because it gives the impotent party a strong incentive to change battlefields and direct its attack against the stronger party’s civilians, thus enhancing the asymmetry in two respects. It enhances the asymmetry in the means applied by the parties and it potentially enhances the asymmetry in power due to the predictable reaction to such terrorist attacks: that we ought to uproot evil with all means necessary. The result is that we enter into an entirely new territory in terms of international law and the ethics of armed conflict. There is a point where a conflict of extreme asymmetry can be understood neither in terms of war nor in terms of self-defence against criminals but rather as some kind of annihilation program where our enemies, in the graphic formulation by Uwe Steinhoff, are “treated like cockroaches on the receiving side of pest control” [4]. 3.3. “Combatant Immunity” LAWS are obviously designed to radically increase the safety of “our” (supposedly just) combatants, ideally by removing them from the battlefield altogether. According to Strawser, this is not only morally justified but also morally required by his self-declared “principle of unnecessary risk”: “It is wrong to command someone to take an unnecessary potentially lethal risk in an effort to carry out a just action for some good; any potentially lethal risk incurred must be justified by some strong countervailing reason” [1]. However, the formulation of this principle is so qualified that it is of limited relevance in the real world. The real world is overpopulated with strong countervailing pragmatic reasons! One could consider, for example, that people in remote areas of Pakistan racked by terrorists and warlords may want to be “protected” by people who make a serious attempt to communicate and to live up to their noble intentions. People who make an effort to negotiate the terms of their strategy with those confronted by military robots with all the social and psychological repercussions of that. Otherwise, one should not be terribly surprised if insurgent recruitment tends to increase after each operation. Again, my point is that this is not a contingent, practical consequence but due to the asymmetry, distance, and isolation inherent in robotic war strategies. 3.4. Stealth Warfare Is such official secrecy a necessary feature of LAWS or merely a contingent feature of current drone strategies? The answer once again depends on the extent to which we take pragmatic considerations seriously. LAWS seem to be intimately linked to a security paradigm that has dominated strong military powers for quite some time, in which official secrecy, lack of transparency and intense political spin (if not propaganda) is indeed definitional. I see no sign of any fundamental change of this in the near future. Drone killings are carried out in a double isolation from public scrutiny. First, the drone strategy carries with it the dogma of giving “no safe haven” to terrorists. Consequently, drone strikes are typically carried out in remote areas and their radical unpredictability in time and place (“anytime, anywhere”) is once again definitional. Second, drone killings are in large measure ordered and carried out by intelligence agencies that are by definition not keen on a high level of transparency. There is thus a discrepancy or paradox involved in the advanced surveillance and documentation capacity of the drones and the secrecy and impenetrability of the operations. On this background, Waldron is right to remind us “how reluctant we should be to deploy principles authorizing homicide in an environment from which we know legal process will be largely banished” [5]. If it is so ethical and effective, where is the information?

## Extensions

### Escalation

#### LAWs make escalation more likely by misinterpreting efforts de-escalation as opportunities to attack

Wong et al 20 Yuna Huh Wong is a policy researcher at the RAND Corporation. Her research interests include scenario development, futures methods, wargaming, problem-structuring methods, and applied social science. John M. Yurchak is a senior information scientist at the RAND Corporation who focuses on defense-related analysis. Robert W. Button is an adjunct senior researcher at the RAND Corporation. His research interests include artificial intelligence and simulation. Aaron Frank is a senior information scientist at the RAND Corporation. He specializes in the development of analytic tradecraft and decision-support tools for assessing complex national security issues. Burgess Laird is a senior international researcher at the RAND Corporation. His subject-matter areas of expertise are defense strategy and force planning, deterrence, and proliferation. Osonde A. Osoba is an information scientist at the RAND Corporation. Randall Steeb is a senior engineer at the RAND Corporation. Benjamin N. Harris is an adjunct defense analyst at the RAND Corporation and a student at the Massachusetts Institute of Technology, where he is pursuing a Ph.D. in political science. Sebastian Joon Bae is a defense analyst at the RAND Corporation. His research interests include wargaming, counterinsurgency, hybrid warfare, violent nonstate actors, emerging technologies, and the nature of future warfare. "Deterrence in the Age of Thinking Machines. Published by the Rand Corporation in 2020. Available here: (https://apps.dtic.mil/sti/pdfs/AD1090316.pdf) – AP

Signaling and Nonhuman Decisionmaking What happens to signaling when not only humans but also machines are involved in sending and receiving signals? In a world with only humans doing the signaling, they do try to show resolve and communicate a deterrent threat, but they also seek to avoid further escalation and to deescalate conflicts. How will machines interpret such signals? Theory of mind, demonstrated from an early age, allows humans to understand that other humans may hold intentions and beliefs about a situation that are different from what they themselves hold to be true. It is this natural ability in most humans that allows them to make some predictions about the behavior of others.5 There is the chance that statistical machine learning could predict certain behaviors from signals, 5 Brittany N. Thompson, “Theory of Mind: Understanding Others in a Social World,” Psychology Today, blog post, July 3, 2017. 66 Deterrence in the Age of Thinking Machines but a very important question in this regard is what data have been used to train the models that make it to the battlefield. We acknowledge the numerous historical cases in which humans have misinterpreted signals from other humans. However, we argue that machines, on the whole, are still worse at understanding intended human signals than are humans, particularly because there is often a complex context that the machine will not understand. We also argue that machines lack theory of mind in novel situations with humans. Table 7.3 lists some of our hypotheses about how machines that are programmed to take advantage of changes in the tactical and operational picture might react to different human signals. Key here is the idea that machines that are set up to rapidly act on advantages they see developing on the battlefield may miss deescalatory signals. In other words, signals developed over decades between humans to deter or deescalate a conflict could have the opposite effect and rapidly escalate a situation if machines are not programmed, or taught, to take deterrence and deescalation into consideration. AI that is set to be aggressive may be at greater danger of misreading the intent behind such signals. We see in Table 7.3 that autonomous systems, programmed to take advantage of tactical and operational advantages as soon as they can identify them, might create inadvertent escalation in situations where the adversary could be trying to prevent further conflict and escalation. We are not arguing against implementing systems that can quickly identify opportunities on the battlefield. It is, however, advisable to ask how to review the situation for adversary signals that machines may miss. Level of Understanding Understanding of an adversary’s will, resolve, and intent are central to deterrence. Figure 7.1 is a simplified diagram of how deterrence has traditionally worked: humans signaling to, interpreting, understanding, and anticipating other humans. (We use blue to denote friendly forces and red to denote adversarial ones.) Put in simple terms, traditional deterrence primarily required humans understanding other humans. In Figure 7.2, we add the types of understanding that are required once machines are involved. Not only must humans understand adversary humans as in Figure 7.1, the following must also occur: • Humans understand their own machines. • Humans understand adversary machines. • Machines understand their humans. • Machines understand adversary humans. • Machines understand other machines. Misunderstanding along any of these dimensions introduces possibilities for misinterpretation, misperception, and miscalculation. Humans understanding their own machines and their range of potential behaviors is not a trivial undertaking. We already have historical examples of systems such as the Phalanx antimissile system firing on U.S. ships and aircraft in ways not anticipated by their human operators and killing U.S. servicemen.6 An even more difficult problem is humans trying to understand adversary machines, particularly machine learning systems. The first obvious problem is that humans do not have ready access to adversary algorithms or understanding of how an adversary system is programmed. For learning systems, even if the algorithm is known, it may be impossible to know the data on which the system trained. Even if the algorithms and data are somehow known, how the adversary intends to use the system and under what circumstances may still be unknown. That is, adversary human-machine collaboration may be a mystery. Rather, humans on one side of the equation may be left trying to infer intent and potential behavior from partial observations. Understanding between humans and machines is a two-way street. It is necessary for machines to accurately understand the intent of their own humans, adversary human behavior and intent, and adversary machine behavior in order to avoid misunderstanding and miscalculation. Will machines accurately understand adversary humans and machines if the adversary behaves differently during conflict, when most of the data on the adversary were collected during peacetime? If machines understand the future primarily through correlation, will they appropriately correlate unexpected adversary behaviors to the “right” things? Figure 7.2 becomes even more complicated when allied and coalition partners and their machines enter the picture. Interoperability with learning systems will pose challenges. And this does not even begin to address a future with a large number of autonomous civilian machines also operating throughout the environment. 6 Paul J. Springer, Outsourcing War to Machines: The Military Robotics Revolution, Praeger Security International, 2018.

#### Inadvertent engagement with autonomous systems is inevitable – but the complexity of LAWs makes this more likely

Wong et al 20 Yuna Huh Wong is a policy researcher at the RAND Corporation. Her research interests include scenario development, futures methods, wargaming, problem-structuring methods, and applied social science. John M. Yurchak is a senior information scientist at the RAND Corporation who focuses on defense-related analysis. Robert W. Button is an adjunct senior researcher at the RAND Corporation. His research interests include artificial intelligence and simulation. Aaron Frank is a senior information scientist at the RAND Corporation. He specializes in the development of analytic tradecraft and decision-support tools for assessing complex national security issues. Burgess Laird is a senior international researcher at the RAND Corporation. His subject-matter areas of expertise are defense strategy and force planning, deterrence, and proliferation. Osonde A. Osoba is an information scientist at the RAND Corporation. Randall Steeb is a senior engineer at the RAND Corporation. Benjamin N. Harris is an adjunct defense analyst at the RAND Corporation and a student at the Massachusetts Institute of Technology, where he is pursuing a Ph.D. in political science. Sebastian Joon Bae is a defense analyst at the RAND Corporation. His research interests include wargaming, counterinsurgency, hybrid warfare, violent nonstate actors, emerging technologies, and the nature of future warfare. "Deterrence in the Age of Thinking Machines. Published by the Rand Corporation in 2020. Available here: (https://apps.dtic.mil/sti/pdfs/AD1090316.pdf) – AP

Past Inadvertent Engagements by Autonomous Systems Military autonomous systems are not new, and neither is inadvertent engagement by such systems. Examples include landmines, torpedoes, close-in weapon systems such as Phalanx,1 and area defense systems such as Aegis. In use since the U.S. Civil War,2 landmines are unable to distinguish among friendly forces, adversary forces, and civilians.3 At least two German U-boats are believed to have been sunk by their own acoustically homing torpedoes during World War II.4 There were also many cases of “circular runs” by American torpedoes in World War II in which torpedoes circled back toward the submarines that launched them. The USS Tang and the USS Tullibee were sunk by their own torpedoes.5 The threat of a circular run by a torpedo persists today; it is mitigated by procedures and the capability to guide torpedoes after launch.6 Phalanx has also experienced several mishaps. In 1989, the USS El Paso used Phalanx to destroy a target drone. The drone fell into the sea, but the Phalanx reengaged it as it fell and struck the bridge of the nearby USS Iwo Jima, killing one and injuring another. During the 1991 Gulf War, the USS Missouri launched chaff to confuse an incoming Iraqi missile. The Phalanx system on the nearby USS Jarrett shot at the Missouri’s chaff and hit the ship four times.7 In 1996, a U.S. A-6E Intruder aircraft towing a radar target during gunnery exercises was shot down when a Phalanx aboard the Japanese destroyer Yūgiri locked onto the A-6E instead of the target. A post-accident investigation concluded that the Yūgiri’s gunnery officer gave the order to fire too early.8 Aegis has been involved in an especially high-profile case of inadvertent engagement. In 1988, the Aegis cruiser USS Vincennes mistook an Iranian civilian airliner for an Iranian fighter and shot it down. It fired two surface-to-air missiles at the airliner and killed all 290 crew and passengers aboard.9 There had been hostilities prior to the incident between U.S. and Iranian forces, including the USS Samuel B. Roberts striking a mine and Iranian forces firing on U.S. helicopters. New U.S. rules of engagement also authorized positive protection measures before coming under fire.10 After Vincennes inadvertently crossed into Iranian waters, Revolutionary Guard gunboats fired on Vincennes’s helicopter. The Vincennes crew also erroneously concluded that the airliner was descending toward the Vincennes when it was in fact climbing.11 These and other factors led to the Vincennes firing on the airliner. A review of the incident by the Chairman of the Joint Chiefs of Staff concluded that while errors had been made, the captain and crew had acted reasonably. The review also found that the Aegis system had performed as designed—particularly, that it was “never advertised as being capable of identifying the type of aircraft being tracked. That decision is still a matter for human judgment.” However, one recommendation was to improve the Aegis display systems in order to better identify important data.12 Table 8.1 summarizes these mistaken engagements with autonomous systems. We note the type of system, the nature of the incident, and reasons behind the mishap. Common reasons for mistaken engagements include target misidentification, an inability on the part of the system to account for friendly forces, and human error. Implications of More-Advanced Autonomous Systems We expect future autonomous systems to be more capable in a number of ways. This could include increased pattern recognition from statistical machine learning to improve target recognition and reduce risks during target selection. It could also involve improved sensing to shorten the decision cycles by which autonomous systems move from searching for and acquiring targets, to engaging them, to deciding to disengage. The proliferation of these more capable systems will likely increase the frequency of their use. How could inadvertent engagements such as those we discussed in Table 8.1 change with more-widespread and more-advanced systems?13 On the one hand, better AI could reduce mistaken engagements through improved target identification, addressing the current problem of discriminating between targets and nontargets. On the other hand, we have noted cases where human error contributed to mishaps. Human error interacting with even more-complex systems could very well contribute to future mistaken engagements.14 Lastly, Table 8.1 largely covers autonomous systems in naval environments with limited civilian presence. Even as AI could improve differentiating targets from nontargets, having more autonomous systems on the ground and in populated areas may come with significant challenges in accounting for friendly forces and noncombatants. We present some potential advantages and disadvantages of future systems in Table 8.2. What are the implications of more-advanced and more-widespread autonomous systems for deterrence and escalation? As more-complex systems and more-complicated human-machine interactions develop, there is clearly the possibility of technical accidents and failures. There is the possibility that one side may interpret accidental engagements by autonomous systems as deliberately escalatory or even preemptive in nature. This is particularly true because it is extremely difficult to surface the full range of behaviors that autonomous systems are capable of during testing. On the other hand, timely notification about accidents and inadvertent engagements, perhaps communicated through means or channels worked out in advance, could help avoid misinterpretation and escalation.

### Bans Work/Good

#### Domestic bans on lethal autonomous weapons establish international cooperation and allow for peaceful re-allocation of resources

Sauer 16 Frank Sauer is a senior research fellow and lecturer at Bundeswehr University in Munich. He is the author of Atomic Anxiety: Deterrence, Taboo and the Non-Use of U.S. Nuclear Weapons (2015) and a member of the International Committee for Robot Arms Control. "Stopping ‘Killer Robots’: Why Now Is the Time to Ban Autonomous Weapons Systems." Published by the Arms Control Association in October 2016. Available here: (https://www.armscontrol.org/act/2016-09/features/stopping-‘killer-robots’-why-now-time-ban-autonomous-weapons-systems) - AP

Implementing autonomy, which mainly comes down to software, in systems drawn from a vibrant global ecosystem of unmanned vehicles in various shapes and sizes is a technical challenge, but doable for state and nonstate actors, particularly because so much of the hardware and software is dual use. In short, autonomous weapons systems are extremely prone to proliferation. An unchecked autonomous weapons arms race and the diffusion of autonomous killing capabilities to extremist groups would clearly be detrimental to international peace, stability, and security. This underlines the importance of the current opportunity for putting a comprehensive, verifiable ban in place. The hurdles are high, but at this point, a ban is clearly the most prudent and thus desirable outcome. After all, as long as no one possesses them, a verifiable ban is the optimal solution. It stops the currently commencing arms race in its tracks, and everyone reaps the benefits. A prime goal of arms control would be fulfilled by facilitating the diversion of resources from military applications toward research and development for peaceful purposes—in the fields of AI and robotics no less, two key future technologies. This situation presents a fascinating and instructive case for arms control in the 21st century. The outcome of the current arms control effort regarding autonomous weapons systems can still range from an optimal preventive solution to a full-blown arms race. Although this process holds important lessons, for instance regarding the valuable input that epistemic communities and civil society can provide, it also raises vexing questions, particularly if and how arms control will find better ways for tackling issues from a qualitative rather than quantitative angle. The autonomous weapons systems example points to a future in which dual-use reigns supreme and numbers are of less importance than capabilities, with the weapons systems to be regulated, potentially disposable, 3D-printed units with their intelligence distributed in swarms. Consequently, more thinking is needed about how arms control can target specific practices rather than technologies or quantifiable military hardware. ‘We Will Not Delegate Lethal Authority...’ “We will not delegate lethal authority for a machine to make a decision. The only time we’ll delegate [such] authority [to a machine] is in things that go faster than human reaction time, like cyber or electronic warfare…. We might be going up against a competitor who is more willing to delegate authority to machines than we are and, as that competition unfolds, we’ll have to make decisions on how we can best compete. It’s not something that we have fully figured out, but we spend a lot of time thinking about it.” —U.S. Deputy Defense Secretary Robert Work, during a Washington Post forum “Securing Tomorrow,” March 30, 2016 Lastly, some policy recommendations are in order. The United States “will not delegate lethal authority for a machine to make a decision,” U.S. Deputy Secretary of Defense Robert Work said in March. Yet, he added that such self-restraint may be unsustainable if an authoritarian rival acts differently. “It’s not something that we have fully figured out, but we spend a lot of time thinking about it,” Work said.14 The delegation of lethal authority to weapons systems will not inexorably happen if CCW states-parties muster the political will not to let it happen. States can use the upcoming CCW review conference in December to go above and beyond the recommendation from the 2016 meeting on lethal autonomous weapons systems and agree to establish an open-ended group of governmental experts with a strong mandate to prepare the basis for new international law, preferably via a ban. Further, a prohibition on autonomous weapons systems should be pursued at the domestic level. Most countries actively engaged in research and development on such systems have not yet formulated policies or military doctrines. Member states of the European Union especially should be called to action. Even if the CCW process were to fizzle out, like-minded states could cooperate and, in conjunction with the Campaign to Stop Killer Robots, continue pursuing a ban through other means. The currently nascent social taboo against machines autonomously making kill decisions meets all the requirements for spawning a “humanitarian security regime.”15 Autonomous weapons systems would not be the first instance when an issue takes an indirect path through comparably softer social international norms and stigmatization to a codified arms control agreement. In other words, even if technology were to overtake the current process, arms control remains as possible as it is sensible.

#### Efforts to ban LAWs are already widespread – bans have only failed due to US and Russia opposition

Human Rigths Watch 19 Human Rights Watch is an international non-governmental organization, headquartered in New York City, that conducts research and advocacy on human rights. "'Killer Robots:' Ban Treaty Is the Only Credible Solution." Published by Human Rights Watch on September 26, 2019. Available here: (https://www.hrw.org/node/334152/printable/print) - AP

(New York) – France, Germany, and other nations that are committed to a rules-based international order should begin negotiations on a new international treaty to ban preemptively lethal autonomous weapons systems, also known as fully autonomous weapons or killer robots. On September 26, 2019, foreign ministers from France, Germany, and dozens of other countries endorsed a declaration at the United Nations addressing lethal autonomous weapons systems. “This declaration is yet another step down the path leading to the inevitable treaty that’s needed to prevent a grim future of killing by machine,” said Mary Wareham, arms advocacy director at Human Rights Watch and coordinator of the Campaign to Stop Killer Robots. “If these political leaders are really serious about tackling the killer robots threat, then they should open negotiations on a treaty to ban them and require meaningful human control over weapons systems and the use of force.” The foreign ministers participating in the “Alliance for Multilateralism” initiative that France and Germany spearheaded share the common goal of promoting a “rules-based international order” and have committed to address killer robots along with climate change and four other “politically relevant” issues. The political declaration endorsed during the annual opening of the UN General Assembly in New York marks the first time such a high-level group has acknowledged the killer robots threat. The killer robots declaration shows that efforts to tackle this urgent challenge are swiftly ascending the multilateral agenda, Human Rights Watch said. Since 2014, more than 90 countries have met eight times at the Convention on Conventional Weapons (CCW) to discuss concerns raised by killer robots. Most of the participating nations wish to negotiate a new treaty with prohibitions and restrictions in order to retain meaningful human control over the use of force. Yet, a small number of military powers – most notably Russia and the United States – have blocked progress toward that objective. As a result, while the talks were formalized in 2016, they still have not produced a credible outcome. At the last CCW meeting in August 2019, Russia and the United States again opposed proposals to negotiate a new treaty on killer robots, calling such a move “premature.” Human Rights Watch and the Campaign to Stop Killer Robots urge states party to the convention to agree in November to begin negotiations next year on a new treaty that requires meaningful human control over the use of force, which would effectively prohibit fully autonomous weapons. Only a new international law can effectively address the multiple ethical, moral, legal, accountability, security, and technological concerns raised by killer robots, Human Rights Watch said. A total of 29 countries have explicitly called for a ban on killer robots: Algeria, Argentina, Austria, Bolivia, Brazil, Chile, China (on use only), Colombia, Costa Rica, Cuba, Djibouti, Ecuador, El Salvador, Egypt, Ghana, Guatemala, the Holy See, Iraq, Jordan, Mexico, Morocco, Nicaragua, Pakistan, Panama, Peru, the State of Palestine, Uganda, Venezuela, and Zimbabwe. The new political declaration on killer robots is unambitious as it falls far short of the new international ban treaty sought by so many. It is ambiguous as it endorses a goal discussed at the Convention on Conventional Weapons of “developing a normative framework,” but there is little agreement among countries about what that means in practice. Some countries view such a framework as guidelines that would not amend existing international law, while others regard it as a new international treaty to prohibit or restrict lethal autonomous weapons systems. The Campaign to Stop Killer Robots, which began in 2013, is a coalition of 118 nongovernmental organizations in 59 countries that is working to preemptively ban fully autonomous weapons and require meaningful human control over the use of force. “It’s obvious that a new treaty to prevent killer robots is desperately needed to ensure a successful rules-based international order,” Wareham said. “Pressure to regulate will intensify the longer it takes nations to commit to negotiate the killer robots treaty.”

#### Efforts to ban LAWs can succeed – viewing them compared to nuclear weapons proves

Coeckelbergh 18 Mark Coeckelbergh is a Belgian philosopher of technology. He is Professor of Philosophy of Media and Technology at the Department of Philosophy of the University of Vienna and former President of the Society for Philosophy and Technology. Janina Loh is a postdoctoral researcher for the Institute of Philosophy at the University of Vienna. Michael Funk is a professor of Philosophy at the University of Vienna, with research focusing on the Philosophy of Media and Technology. Johanna Seibt is a regular faculty at Aarhus University, Department of Philosophy and specializes in the areas of analytical ontology and metaphysics; most recently she works in robophilosophy. Marco Nørskov is an Assistant Professor at the Department of Philosophy and History of Ideas at Aarhus University, Denmark. "Envisioning Robots in Society – Power, Politics, and Public Space : Proceedings of Robophilosophy 2018 / TRANSOR 2018." Published by IOS Press in 2018. Available here: (<http://web.b.ebscohost.com.ezproxy.lib.utah.edu/ehost/ebookviewer/ebook?sid=8f3d98ec-1aae-4e83-a800-7e2b617e3535%40pdc-v-sessmgr03&vid=0&format=EB>) - AP

1.2.Military AI as a ‘Hard Case’ for Global Governance; Nuclear Weapons as Case Study Given such ethical, legal and strategic risks, we may desire to avert or control the militarization of AI, and to develop governance arrangements that contain either ‘horizontal’ proliferation (i.e. more parties pursuing or deploying AI weapons), or ‘vertical’ proliferation (i.e. actors developing and deploying more advanced, potentially more destabilizing, ethically problematic or error-prone AI weapons). One challenge faced is that military AI appears to offer such strong unilateral strategic advantages to principals that develop and deploy them. In this, military AI is arguably a ‘hard’ case for AI governance approaches in general. In meeting this hard challenge, we might look to past military technologies—which enables us to learn from a long—if at times checkered—history of nonproliferation and arms control. Such a history can serve as a rich seam of lessons about the opportunities (and the pitfalls) in stopping, controlling, or containing a weaponized technology, potentially informing or expanding the dialogue on policies for military AI. Nuclear weapons offer one such fruitful historical case. Indeed, the comparison between early nuclear science and AI as strategically disruptive technologies has been drawn previously [20]. This is because while nuclear weapons and military AI are different technologies on an object level, in a strategic context they share key features: (i) they offer a strong and ‘asymmetric’ strategic advantage; (ii) they involve dual-use components, technologies, and applications, which mean that blanket global bans of the core technology are not politically palatable, enforceable, or even desirable; (iii) both involve an (initially) high threshold in technology and tacit scientific knowledge. We therefore next offer two brief explorations of governance lessons from nuclear history. 2.Two Lessons from Nuclear Weapons History for AI Governance 2.1.Global Norms and Domestic Politics Shape the Causes and Cures of Arms Races There is a widespread perception that self-interested states cannot be permanently barred from pursuing strategically important technology which they suspect their rivals might develop—that accordingly, military AI arms races are inevitable or even already underway, and the horizontal, global proliferation of fully autonomous ‘killer robots’ a matter of time. Such pessimism echoes historical fears from the nuclear era, that “proliferation begets proliferation” [21, p. 18]. Indeed, policymakers in the early Cold War perceived the possession of nuclear weapons—the ultimate deterrent—as desirable or necessary, and therefore anticipated a wildfire spread of these weapons. In 1963, on the basis of a memo by then-US Secretary of Defense Robert McNamara, President John F. Kennedy publicly and famously envisioned “the possibility in the 1970s of [...] a world in which 15 or 20 or 25 nations may have these weapons” [22]. Yet remarkably, given such pessimism, ‘horizontal’ nuclear proliferation since the 1960’s has proven less the ‘wildfire’ and more a ‘glacial spread.’ By some estimates, in the past seven decades up to 56 states have at one or another time possessed the (theoretical) capability to develop a nuclear weapons program [23, p. 30]. Even though many of these states—up to 39 by some estimates [24]—chose to engage in ‘nuclear weapons activity,’ the majority eventually voluntarily terminated these programmes, uncompleted [25, p. 273]. ‘Only’ ten states have actually managed to develop these weapons, and after South Africa dismantled its small nuclear arsenal, nine nuclear weapons states presently remain. How can this be explained? The literature on state decision-making behavior identifies a range of models, many of which focus on the respective roles of (1) security interests; (2) domestic politics, and (3) norms [26]. Under the security model, states pursue nuclear weapons in reaction to perceived security threats—either to offset a rival’s conventional military supremacy, or to match an adversary’s (feared) nuclear program. Under this ‘realist’ reading, nonproliferation policy can only slow down, but not eliminate, the spread of nuclear weapons. While intuitive and parsimonious, there are some problems with the security model: for instance, ‘national security’ often serves as a default post-hoc rationalization for decision-makers seeking to justify complex, contested choices by their administrations. Moreover, as noted by Sagan, “an all too common intellectual strategy in the literature is to observe a nuclear weapons decision and then work backwards, attempting to find the national security threat that ‘must’ have caused the decision” [26, p. 63]. Other scholarship has therefore turned to the role of domestic politics—to the diverse sets of actors who have parochial bureaucratic or political interests in the pursuing or foregoing of nuclear weapons. These actors include policy elites; nuclear research- or energy industry establishments; competing branches of the military; politicians in states where parties or the public favor nuclear weapons development. Such actors can form coalitions to lobby for proliferation. This happened in India, where the 1964 nuclear test by rival China did not produce a crash weapons program, and instead set off a protracted, decade-long bureaucratic battle between parties in the Indian elite and nuclear energy establishments. This struggle was only decided in 1974, when Prime Minister Indira Gandhi, facing a recession and crisis of domestic support, authorized India’s ‘Peaceful Nuclear Explosion,’ possibly to distract or rally public opinion [26, pp. 65–69]. Another example is found in the South African nuclear program, which saw first light not as a military project, but as an initiative, by the Atomic Energy Board, to develop these devices for mining uses [26, pp. 69–70]. Conversely, domestic politics can also turn against proliferation: after pursuing nuclear programs throughout the 1970s-1980s, regional rivals Brazil and Argentine eventually abandoned their nuclear ambitions—the result of new liberalizing domestic regimes supported by actors (e.g. banks, monetary agencies) who favored open access to global markets and opposed ‘wasteful’ defense programs [27]. Finally, a third model emphasizes the role of (domestic and global) norms on states’ desire to pursue nuclear weapons. In some cases, the perception of nuclear weapons’ symbolic value may have driven proliferation. In the French case, the experiences in the First Indochina War and the 1958 Algerian Crisis seem to have contributed to President de Gaulle’s strong desire to obtain the atomic bomb as a symbol of restored French great power status [26, pp. 78–79]. More often, however, norms—implicitly the ‘nuclear taboo,’ and explicitly the norms encoded by global international legal instruments including the Nuclear Non-Proliferation Treaty (NPT) and the Comprehensive Nuclear Test-Ban Treaty (CTBT)—have served as a major factor in constraining nuclear proliferation. Such global legal instruments provide shared normative frameworks, and thereby promote non-proliferation norms or interests at the domestic-political level, tipping the balance of power towards domestic coalitions seeking non-proliferation. Moreover, these global international regimes—defined by Krasner as “sets of implicit or explicit principles, norms, rules and decision-making procedures around which actors’ expectations converge in a given area of international relations” [28, p. 2]—also serve as ‘Schelling points’ around which global society can converge to coordinate collective sanctions, or jointly promise economic or political rewards [29]. Intriguingly, while public norms seem able to strengthen the hands of (non)proliferation coalitions, they do not seem to reliably shift state policymaking where such coalitions do not already exist in some strength: in 1994 Ukraine chose to join the NPT and renounce its nuclear arsenal in spite of Ukrainian public support for retaining the weapons [26, p. 80]. Conversely, in 1999 the US Senate rejected the CTBT in the face of widespread US public support. While it is hard to distill causal chains, and while elements of all three models may appear across almost all cases, a preliminary, incomplete survey (Tables 1-2) of cases of nuclear proliferation (Table 1) and nuclear non-proliferation (Table 2), illustrate a diverse array of motivational factors involved [23, 26, 27, 30-32]. This overview suggests that even when they involve strategically appealing technologies, proliferation is far from a foregone conclusion, and arms races can be slowed, channeled, or averted entirely. It also introduces some considerations for military AI arms races. In the first place, it suggests that security concerns are conducive but not decisive to arms races, and that ‘first-mover’ major powers may share an interest in supporting global legal regimes aimed at the non-proliferation (if not disarmament) of certain forms of military AI which might otherwise empower conventionally weaker (non-state) rivals. In the second place, the domestic-politics model suggests that strengthening the hand of domestic coalitions pursuing the non-proliferation (or the responsible development) of AI weapons is one possible pathway towards shifting state decision-making away from pursuing more problematic categories of military AI, even in the ostensible face of clear national security interests. Conversely, some inhibitive factors—such as excessive program cost—seem less applicable for military AI systems, which are both comparatively less expensive than nuclear weapons, and which also offers to ‘pay for themselves’ through possible civilian spinoff applications. Finally, while policy-makers may pursue the development of AI in general because of its ‘symbolic’ value as a marker of global scientific leadership, it is less clear if (or to who) the development of military AI, let alone ‘killer robots’, will confer similar global prestige. Additionally, it appears unlikely that states will face an ‘AI weapons taboo’ similar to the ‘nuclear taboo;’ nuclear weapons were not used in anger since Hiroshima, and have a publically visible and viscerally horrifying use mode that creates a clear ‘nuclear Rubicon’ not to be crossed. Conversely, the daily application of some AI in wars is in some sense already a fact—that red line been crossed. Moreover, these uses are more diverse, and only some visceral applications (e.g. ‘killer robots’) may generate public opprobrium, whereas less kinetic ones (e.g. surveillance; tracking missile submarines) may not. Moreover, while public norms or activism against military AI might strengthen domestic political coalitions already opposed to these weapons, they alone are not always able to sway policymakers in the first place. A key route lies in shaping policymakers’ norms or their domestic political landscapes. This relies on the (top-down) influence exerted by global legal instruments and regimes, and on the (bottom-up) institutionalization of norms by ‘epistemic communities.’

## AT:

### AT: Miscalculatioon Inevitable

#### Even if miscalculation is inevitable – the speed of LAWs Processing is likely to cause far more damage and spark flash wars

Scharre 18 Paul Scharre is Director for Technology and National Security at the Center for a New American Security (CNAS). "A Million Mistakes a Second." Published by Forein Policy on September 12, 2018. Available here: (https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/) - AP

Despite humans’ advantages in decision-making, an arms race in speed may slowly push humans out of the OODA loop. Militaries are unlikely to knowingly field weapons they cannot control, but war is a hazardous environment and requires balancing competing risks. Faced with the choice of falling behind an adversary or deploying a new and not yet fully tested weapon, militaries are likely to do what they must to keep pace with their enemies. As mentioned above, automated stock trading provides a useful window into the perils of this dynamic. In 2010, the Dow Jones Industrial Average lost nearly 10 percent of its value in just minutes. The cause? A sudden shift in market prices driven in part by automated trading, or what’s come to be known as a flash crash. In the last decade, financial markets have started to suffer such crashes, or at least miniature versions of them, on a regular basis. The circuit breakers installed by regulators to pull a stock offline can’t prevent incidents from occurring, but they can stop flash crashes from spiraling out of control. Circuit breakers are still regularly tripped, though, and on Aug. 24, 2015, more than 1,200 of them went off across multiple exchanges after China suddenly devalued the yuan. In competitive environments such as stock markets and battlefields, unexpected interactions between algorithms are natural. The causes of the 2010 flash crash are still disputed. In all likelihood, there were a range of causes, including an automated sell algorithm interacting with extreme market volatility, exacerbated by high-frequency trading and deliberate spoofing of trading algorithms. To prevent the military equivalent of such crises, in which autonomous weapons become trapped in a cascade of escalating engagements, countries will have to balance advantages in speed with the risk of accidents. Yet growing competition will make that balancing act ever more difficult. In 2016, Robert Work, then-U.S. deputy defense secretary, colorfully summed up the problem this way: “If our competitors go to Terminators, and it turns out the Terminators are able to make decisions faster, even if they’re bad, how would we respond?” Again, stock markets show how important it is that countries answer this question in the right way. In 2012, an algorithm-based trading accident nearly bankrupted the high-frequency trading firm Knight Capital Group. A glitch in a routine software update caused the firm’s computers to start executing a lightning-fast series of erroneous trades, worth $2.6 million a second. By the time the company reined in its runaway algorithm, its machines had executed 4 million trades with a net loss of $460 million—more than the company’s entire assets. To give a sense of scale: In 1994, it took more than two years of deception for the rogue trader Nick Leeson to bankrupt Barings Bank. In what came to be known as the Knightmare on Wall Street, a machine managed to inflict the same damage in 45 minutes. In that case, of course, although a company was destroyed, no lives were lost. A runaway autonomous weapon would be far more dangerous. Real-world accidents with existing highly automated weapons point to these dangers. During the initial invasion of Iraq in 2003, the U.S. Army’s Patriot air defense system accidentally shot down two friendly aircraft, killing three allied service members. During the initial invasion of Iraq in 2003, the U.S. Army’s Patriot air defense system accidentally shot down two friendly aircraft, killing three allied service members. The first fratricide was due to a confluence of factors: a known flaw that caused the radar to mischaracterize a descending plane as a missile, outdated equipment, and human error. The second blue-on-blue incident was due to a situation that had never arisen before. In the hectic march to Baghdad, Patriot operators deployed their radars in a nonstandard configuration likely resulting in electromagnetic interference between the radars that caused a “ghost track”—a signal on the radars of a missile that wasn’t there. The missile battery was in automatic mode and fired on the ghost track, and no one overruled it. A U.S. Navy F-18 fighter jet just happened to be in the wrong place at the wrong time. Both incidents were flukes caused by unique circumstances—but also statistically inevitable ones. Coalition aircraft flew 41,000 sorties in the initial phases of the Iraq War, and with more than 60 allied Patriot batteries in the area, there were millions of possible interactions, seriously raising the risk for even low-probability accidents. Richard Danzig, a former U.S. secretary of the Navy, has argued that bureaucracies actually systematically underestimate the risk of accidents posed by their own weapons. It’s also a problem that it’s nearly impossible to fully test a system’s actual performance outside of war. In the Iraq invasion, these accidents had tragic consequences but did not alter the course of the war. Accidents with fully autonomous weapons where humans cannot intervene could have much worse results, causing large-scale fratricide, civilian casualties, or even unintended attacks on adversaries.

### AT: ‘Just War’

#### Efforts to understand warfighting under just-war theory are idealized and misrepresent how violence occurs in the real world

Coeckelbergh et al. 18 Mark Coeckelbergh is a Belgian philosopher of technology. He is Professor of Philosophy of Media and Technology at the Department of Philosophy of the University of Vienna and former President of the Society for Philosophy and Technology. Janina Loh is a postdoctoral researcher for the Institute of Philosophy at the University of Vienna. Michael Funk is a professor of Philosophy at the University of Vienna, with research focusing on the Philosophy of Media and Technology. Johanna Seibt is a regular faculty at Aarhus University, Department of Philosophy and specializes in the areas of analytical ontology and metaphysics; most recently she works in robophilosophy. Marco Nørskov is an Assistant Professor at the Department of Philosophy and History of Ideas at Aarhus University, Denmark. "Envisioning Robots in Society – Power, Politics, and Public Space : Proceedings of Robophilosophy 2018 / TRANSOR 2018." Published by IOS Press in 2018. Available here: (<http://web.b.ebscohost.com.ezproxy.lib.utah.edu/ehost/ebookviewer/ebook?sid=8f3d98ec-1aae-4e83-a800-7e2b617e3535%40pdc-v-sessmgr03&vid=0&format=EB>) - AP

4. Abstraction of Ideal Theory People who defend military robots in principle thus do so in abstraction from the political realities of their realistic use in practice. The most obvious objection to such a move is to question its relevance. What good is it that LAWS would work ideally in a world with fundamentally different characteristics than our world? That may be too rash, however. People who defend LAWS by this method could insist that there is a valuable lesson to learn from an ideal moral theory of robotic war. Strawser and many other participants in the ethical debate on LAWS do favour a kind of ideal theorizing about war in general: the so-called “revisionist” school in the ethics of war (arguably led by Jeff McMahan). According to revisionists, the conventional view of the morality of war is false. First, the moral status of combatants depend entirely on which side they are. Unjust aggressors have no permission to fight just because their victims justifiably fight back. Only just combatants have a valid permission to fight. Secondly, non-combatants are not necessarily immune since the ideal criterion for liability is moral responsibility for unjust threats. Civilians may well bear such responsibilities. Ideally, then, only just combatants use their weapons and they target them at all liable persons, be they combatants or noncombatants [6]. Now, the inherent features of LAWS mentioned before (aggressive tactics, extreme asymmetry, combatant immunity and stealth warfare) is welcome news in an ideal situation in which such advantages favour the just side. Historically, however, the main purpose of the just war principles was always pragmatic: how do we handle a situation in which both just and unjust people take up arms? Revisionists often present their account as an account of the “deep morality” of war. It is thus an account of the ethics of war that overlooks contingencies, such as epistemic issues, as well as those to do with noncompliance and other unintended consequences. In short, it is an account of the necessary features of the morality of war [7]. Among the most significant contingent features overlooked by ideal theory are these: 1)It is typically very unclear which party (if any) is just, 2)Individual liability can be extremely difficult to determine, 3)Contrary to what weapons producers advertise, all weapons tend to kill not only those intended, 4)Permissions to target noncombatants will be misused. Given such contingencies in a non-ideal world, practicing such principles would probably be disastrous. First, since all parties to a conflict tend to consider themselves in the right, it will only enforce enmity to believe that the other party does not even have moral permission to participate in fighting. Second, making exceptions to the prohibition against targeting civilians would likely constitute the top of a slippery slope towards total war. Some revisionists are well aware of this problematic. For instance, McMahan thinks that the revised principle of permission to fight should not be enacted until we have some kind of institutional framework in place to decide in an impartial way which parties to a conflict are really just or unjust [8]. Nevertheless, we could still imagine a situation in which we had somehow remedied the problematic contingencies. In that situation, the ideal principles would not only be true but also workable. What we now need to consider is whether the ideal morality is feasible. The answer depends on the strength of what Holly Lawford-Smith has termed feasibility constraints[9]. Ideal moral theory takes into account only hard feasibility constraints. Such constraints only rule out principles that are logically, conceptually, metaphysically, or nomologically impossible.3 Ideal theorists often work from the assumption that as long as you do not ignore hard feasibility constraints, your principles may still be feasible. Non-ideal moral theory, on the other hand, insists on taking soft feasibility constraints into account as well, i.e. the economic, institutional and cultural circumstances that make the ideal moral principles infeasible in the real world. Exactly how soft such constraints are is a question of how realistic a development from non-ideal circumstances to (more) ideal circumstances is. Suppose we know from historical experience, from insights into the dynamics of armed conflicts, from experts in social psychology etc. that blatantly unjust combatants almost invariably consider themselves to be in the right. And suppose we are aware that any significant change would depend on fundamental reforms in military institutions and military training, several decades of intense education programs, and profound reorientation of cultural mores. In that case, it seems to me that we face more than a soft constraint on an ideal norm of moral inequality of combatants. I propose to call constraints of such nature a high-density constraint.4High-density constraints may strictly speaking be “merely contingent” but if they are contingent on a complex web of interrelated and deeply rooted circumstances, it seems to me simply irresponsible to overlook them in our moral evaluation of a given military practice or strategy. I think this happens a lot in the debates over LAWS. Another set of high-density constraints enters the picture the moment we start to consider what happens if we move, not towards an ideal situation e.g. in which only just people take up arms but the opposite, i.e. an even more non-ideal situation of wide-spread proliferation of LAWS. Even if we agree that the people we target with LAWS are all unjust and liable to be killed, we may be well advised to consider the possibility that they(i.e. the bad guys) may also be in possession of them. This possibility does not seem to be on the radar of proponents of LAWS like Strawser and Arkin. The obvious explanation once again is their exclusive focus on ideal reasoning.

### AT: Arms Race Rhetoric Bad

#### Governments and Media already frame military AI as an arms race – focusing on research around this is key to prevent its most harmful effects

Cave and ÓhÉigeartaigh 18 Stephen Cave, Leverhulme Centre for the Future of Intelligence, University of Cambridge. Seán S ÓhÉigeartaigh, Centre for the Study of Existential Risk, University of Cambridge. "An AI Race for Strategic Advantage: Rhetoric and Risks." Published In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’18). February 2-3, 2018. Available here: (https://dl.acm.org/doi/pdf/10.1145/3278721.3278780) - AP

Choices for the Research Community Given the dangers noted above of talking about AI development as a competitive race, one could argue that it would be better for the community of researchers considering AI and its impacts to avoid this terminology altogether. In this way, the language of the race for dominance would be seen as (in Nick Bostrom’s terms) an information hazard -- perhaps either an idea hazard (a general idea whose dissemination could increase risk) or an attention hazard (if we consider that the dangerous idea already exists, but is as yet largely unnoticed) (Bostrom 2011). But if we believe that the idea is already being disseminated by influential actors, such as states, including major powers, then the information hazard argument is weakened. It might still apply to particularly influential researchers -- those whose thoughts on the future of AI can become headline news around the world. But even in their case, and particularly in the case of lesser-known figures, there is a strong countervailing argument: that if the potentially dangerous idea of an AI race is already gaining currency, then researchers could make a positive impact by publicly drawing attention to these dangers, as well as by pursuing research dedicated to mitigating the risks of such a framing. Of course, many leading researchers are already speaking out against an AI arms race in the sense of a race to develop autonomous weapons -- see for example the Future of Life Institute’s open letter on this, signed by over three thousand AI researchers and over fifteen thousand others (Future of Life Institute 2015b). We believe this community could also usefully direct its attention to speaking out against an AI race in this other sense of a competitive rush to develop powerful general-purpose AI as fast as possible. Both media and governments are currently giving considerable attention to AI, yet are still exploring ways of framing it and its impacts. We believe that there is therefore an opportunity now for researchers to influence this framing1. For example, few countries have published formal strategies or legislation on AI, but a number have commissioned reviews that have sought expert opinion, eg, the UK Government-commissioned report on AI (Hall and Prescenti, 2017) and the White House report (NSTC, 2016). of the principles on AI agreed at the 2017 Asilomar conference offers a precedent on which to build -- it states: Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards (Future of Life Institute 2017).

### AT: US Hegemony

#### Hegemonic stability theory is flawed – the influence of the hegemon causes global ebbs and flows based on their own domestic trends ensuring crisis occurs

Gaiya 19 Abel B.S. Gaiya is a Commonwealth Shared Scholar and MSc Development Economics candidate at SOAS, University of London. "Tensions in Hegemonic Stability and Global Structural Transformation." Published by Developing Economics: A Critical Perspective on Development Economics." Published March 12, 2019. Available here: (https://developingeconomics.org/2019/03/12/tensions-in-hegemonic-stability-and-global-structural-transformation/) - AP

The Role of Power in International Trade Within the field of international relations, the Hegemonic Stability Theory (Gilpin, 1972; Kindleberger, 1986) has gained wide acceptance among many realists. In simple terms, it says that in an international system characterized by anarchy (which means that, unlike nations, the world lacks a single sovereign to mediate international politics), a hegemon is necessary to instil stability and order in the global system. The hegemon provides the “international public goods” necessary for global stability. Within modern capitalism, Great Britain in the 19th century very imperfectly served as one such hegemon; and the United States currently serves as the hegemon, underwriting the historically unprecedented post-war international liberal order (Kagan, 2018). Two major conditions for hegemony are industrial and/or economic (and technological) dominance, and military dominance to be able to balance other powers and maintain global stability. The hegemon also works with other major economic powers in the delivery of international public goods and the oversight of the global community. Its ability to maintain international peace is necessary for prosperity to ensue. However, the international responsibilities of hegemony (e.g. underwriting security alliances) also include responsibilities for global economic governance (i.e. underwriting economic alliances). It was the U.S., along with Britain to a subordinate degree that led the creation of the post-war new institutions of global economic governance – the Bretton Woods system. And it was England in the 19th century that first adopted the gold standard; and was a primary actor in the late 19th century Berlin Conference. The hegemon maintains unilateral power to influence the affairs of weaker nations around the world, as well as to do so through multilateral institutions. In other words, the hegemon and its cohorts can shape the global rules of the game in ways that “subordinate” nations cannot. This means that the hegemon and its cohorts are very capable of externalizing the (transitional) costs of their domestic economic tensions, unilaterally, bilaterally and/or multilaterally. When faced with the transitional turbulence associated with global structural change, the hegemon and its cohorts, being the economic leaders and hence those actually experiencing the turbulence and pressures, are able to use their channels of power to change the international rules of the game in order to limit the pace of global structural change and thereby stabilize their polities. The hegemon is particularly important as the prime market to absorb exports from the industrializing countries because it typically possesses the largest economy, and yet it has the most power to externalize the costs of adjustment to such absorption. Robert Brenner (2006) has written copiously about how the post-war (re)development of Germany and Japan, and later on the industrialization of the Asian Tigers, intensified international competition and instigated a crisis of overproduction and a crisis of profitability in the reigning hegemon, the U.S. Other factors were certainly at play (such as the oil shocks and the inflationary wage, fiscal and monetary pressures); but within the background of a crisis of profitability, the attempts made by the U.S. to deal with its crisis affected the rest of the world. For instance, the Volcker shock of the 1980s precipitated the debt crisis in the developing world. Silver and Arrighi (2003) extend this argument to the case of the Long Depression of 1873-1896 which saw England face domestic economic turbulence due to the intensified international competition resulting from the industrial rise of Germany and the U.S. Bhagwati and Irwin (1987) further show the protectionist rhetorics employed by 19th century England and 1980s U.S. when faced with the new competitors who employed protectionist policies as part of their industrialization measures. As a result of these pressures, and with the contribution of others, both hegemons in their different periods of reign engaged in significant externalization of the adjustment costs. In 19th century England there was the engagement of “protectionism, mercantilism, and territorial expansion overseas” (Silver and Arrighi, 2003:334). Colonial economic policy sought to coercively enforce the Ricardian comparative advantage logic by preventing colonies from developing manufacturing capabilities, explicitly limiting their production to raw materials as well as making the colonial power the primary export destination; and actively avoiding developmental capital transfers to the colonies. In post-war U.S. there was the neoliberalization of the economic rules of the game to coercively and consensually enforce the same logic, but with different tools (such as structural adjustment lending imposing the elements of the Washington Consensus, and eventually the creation of the WTO to further shrink development policy space). Confronting the Political and Economic Implications for Development What this all means is that with the structure of international relations and politics, and its intersection with international economic governance and development economics, it seems that there can, actually be no smooth trend of broad-based global development (Gaiya, 2018). With an accommodating global development space (more so if the developing country or region has geostrategic value to the leading power), some economies will successfully structurally transform (strongly aided by export growth and external development finance flows), but this would trigger a narrowing of the global development space under which many other economies will struggle to develop. This is not an indictment of the West; as it implies that no matter who the leader is (whether from East or West), these tendencies will be present. Perhaps it may also be a systematic feature of global capitalism for the opening up of the global development space to come at the heels of another series of crises – just as the post-war broadened development space came at the close of two world wars and a great depression which enabled leaders to fashion international institutions that allowed countries significant policy space. The implication is that, in creating a new global economic order, economics cannot be separated from politics and international relations. And the domestic affairs of the hegemon can no longer be treated as inconsequential to the global order. For instance, to improve the ease of the leader’s structural change process (and thereby reducing the frictions involved in global structural change), it may be necessary for the leader to maintain institutions that foster social protection, labour mobility (occupational and geographical) and social investment (for high-quality post-industrial servicification to occur) within its borders. These should no longer be seen as social-democratic luxuries, but as burdens of hegemony necessary for quicker global development. The Global South must also work to better understand these interrelations with global development. The implications are not trivial. For instance, emerging economies may need to be mandated to make plans for internal integration so as to avoid the global imbalances which put persistent pressure on the leaders. What this all reminds us is that Southern structural transformation is inextricably tied to Northern structural change; and therefore global structural change is more interactive than we commonly think, and requires much more international cooperation, checks and balances.

# LAWs Neg

### Value-Criterion

#### For this debate I offer the value of security

Rothschild 95 Emma Rothschild is director of the Centre for History and Economics at King's College. "What is Security?" Published by Daedalus, Vol. 124(3), pp. 53-98. Published Summer 1995. Available here: (https://www-jstor-org.ezproxy.lib.utah.edu/stable/pdf/20027310.pdf?refreqid=excelsior%3A02230441d546cbd2c47bf432fdc2bad5) - AP

The idea of security has been at the heart of European political thought since the crises of the seventeenth century. It is also an idea whose political significance, like the senses of the word "secu rity," has changed continually over time. The permissive or plural istic understanding of security, as an objective of individuals and groups as well as of states? the understanding that has been claimed in the 1990s by the proponents of extended security?was charac teristic, in general, of the period from the mid-seventeenth century to the French Revolution. The principally military sense of the word "security," in which security is an objective of states, to be achieved by diplomatic or military policies, was by contrast an innovation, in much of Europe, of the epoch of the Revolutionary and Napoleonic Wars. But security was seen throughout the pe riod as a condition both of individuals and of states. Its most consistent sense?and the sense that is most suggestive for modern international politics?was indeed of a condition, or an objective, that constituted a relationship between individuals and states or societies. "My definition of the State," Leibniz wrote in 1705, "or of what the Latins call Respublica is: that it is a great society of which the object is common security (ela seuret? commune')."24 For Montesquieu, security was a term in the definition of the state, and also in the definition of freedom: "political freedom consists in security, or at least in the opinion which one has of one's security."25 Security, here, is an objective of individuals. It is some thing in whose interest individuals are prepared to give up other goods. It is a good that depends on individual sentiments?the opinion one has of one's security?and that in turn makes possible other sentiments, including the disposition of individuals to take risks, or to plan for the future. The understanding of security as an individual good, which persisted throughout the liberal thought of the eighteenth century, reflected earlier political ideas. The Latin noun "securitas" re ferred, in its primary classical use, to a condition of individuals, of a particularly inner sort. It denoted composure, tranquillity of spirit, freedom from care, the condition that Cicero called the "object of supreme desire," or "the absence of anxiety upon which the happy life depends." One of the principal synonyms for "securitas," in the Lexicon Taciteum, is "Sicherheitsgef?hl": the feeling of being secure.26 The word later assumed a different and opposed meaning, still in relation to the inner condition of the spirit: it denoted not freedom from care but carelessness or negli gence. Adam Smith, in the Theory of Moral Sentiments, used the word "security" in Cicero's or Seneca's sense, of the superiority to suffering that the wise man can find within himself. In the Wealth of Nations, security is less of an inner condition, but it is still a condition of individuals. Smith indeed identifies "the liberty and security of individuals" as the most important prerequisites for the development of public opulence; security is understood, here, as freedom from the prospect of a sudden or violent attack on one's person or property.27 It is in this sense the object of expenditure on justice, and of civil government itself.28 There is no reference to security, by contrast, in Smith's discussion of expenditure on de fense ("the first duty of the sovereign, that of protecting the society from the violence and invasion of other independent soci eties").29 The only security mentioned is that of the sovereign or magistrate as an individual, or what would now be described as the internal security of the state: Smith argues that if a sovereign has a standing army to protect himself against popular discontent, then he will feel himself to be in a condition of "security" such that he can permit his subjects considerable liberty of political "remonstrance."30

#### To determine what is best for security we should use a criterion of ‘Realism’

Walt 18 Stephen M. Walt is the Robert and Renée Belfer professor of international relations at Harvard University."The World Wants You to Think Like a Realist". Pulbished by Foreign Policy on May 30, 2018. Available here: (https://foreignpolicy.com/2018/05/30/the-world-wants-you-to-think-like-a-realist/) - AP

In short, it is still highly useful to think like a realist. Let me explain why. Realism has a long history and many variants, but its core rests on a straightforward set of ideas. As the name implies, realism tries to explain world politics as they really are, rather than describe how they ought to be. For realists, power is the centerpiece of political life: Although other factors sometimes play a role, the key to understanding politics lies in focusing on who has power and what they are doing with it. The Athenians’ infamous warning to the Melians captures this perfectly: “The strong do what they can, and the weak suffer what they must.” Quentin Tarantino couldn’t have put it any better. For realists, states are the key actors in the international system. There is no central authority that can protect states from one another, so each state must rely upon its own resources and strategies to survive. Security is a perennial concern — even for powerful states — and states tend to worry a lot about who is weaker or stronger and what power trends appear to be. Cooperation is far from impossible in such a world — indeed, at times cooperating with others is essential to survival — but it is always somewhat fragile. Realists maintain that states will react to threats first by trying to “pass the buck” (i.e., getting someone else to deal with the emerging danger), and if that fails, they will try to balance against the threat, either by seeking allies or by building up their own capabilities. Realism isn’t the only way to think about international affairs, of course, and there are a number of alternative perspectives and theories that can help us understand different aspects of the modern world. But if you do think like a realist — at least part of the time — many confusing aspects of world politics become easier to understand. If you think like a realist, for example, you’ll understand why China’s rise is a critical event and likely to be a source of conflict with the United States (and others). In a world where states have to protect themselves, the two most powerful states will eye each other warily and compete to make sure that they don’t fall behind or become dangerously vulnerable to the other. Even when war is avoided, intense security competition is likely to result.

### Accountability

#### Lethal Autonomous Weapons can more accurately prevent civilian deaths and simultaneously identify biases in their deployment

Lewis 20 Larry Lewis was a senior advisor for the State Department on civilian protection in the Obama Administration, the lead analyst and co-author for the Joint Cviilian Casualty Study, and spearheaded the first data-based appraoch to protecting civilians in conflict. "Killer robots reconsidered: Could AI weapons actually cut collateral damage?" Published by the Bulletin of the Atomic Scientists on January 10, 2020. Available here: (https://thebulletin.org/2020/01/killer-robots-reconsidered-could-ai-weapons-actually-cut-collateral-damage/) - AP

The United States, Russia, and China are all signaling that artificial intelligence (AI) is a transformative technology that will be central to their national security strategies. And their militaries are already announcing plans to quickly move ahead with applications of AI. This has prompted some to rally behind an international ban on autonomous, AI-driven weapons. I get it. On the surface, who could disagree to quashing the idea of supposed killer robots? Well, me for starters. The problem with an autonomous weapons ban is that its proponents often rely on arguments that are inaccurate both about the nature of warfare and about the state of such technology. Activists and representatives from various countries have been meeting at the United Nations for six years now on the issue of lethal autonomous weapons. But before calling for society to ban such weapons, it behooves us to understand what we are really talking about, what the real risks are, and that there are potential benefits to be lost. In short, we need to talk about killer robots, so we can make an informed decision. Unfortunately, for many people, the concept of autonomous weapons consists of Hollywood depictions of robots like the Terminator or RoboCop—that is, uncontrolled or uncontrollable machines deciding to wreak havoc and kill innocents. But this picture does not represent the current state of AI technology. While artificial intelligence has proved powerful for applications in banking, in medicine, and in many other fields, these are narrow applications for solving very specific problems such as identifying signs of a particular disease. Current AI does not make decisions in the sense that humans do. Many AI experts such as the authors of Stanford University’s One Hundred Year Study on Artificial Intelligence don’t think so-called general AI–the kind envisioned in science fiction that’s more akin to human intelligence and able to make decisions on its own—will be developed any time soon. The proponents of a UN ban are in some respects raising a false alarm. I should know. As a senior advisor for the State Department on civilian protection in the Obama administration, I was a member of the US delegation in the UN deliberations on lethal autonomous weapons systems. As part of that delegation, I contributed to international debates on autonomous weapons issues in the context of the Convention on Certain Conventional Weapons, a UN forum that considers restrictions on the design and use of weapons in light of the requirements of international humanitarian law, i.e, the laws of war. Country representatives have met every year since 2014 to discuss the future possibility of autonomous systems that could use lethal force. And talk of killer robots aside, several nations have mentioned their interest in using artificial intelligence in weapons to better protect civilians. A so-called smart weapon—say a ground-launched, sensor-fused munition— could more precisely and efficiently target enemy fighters and deactivate itself if it does not detect the intended target, thereby reducing the risks inherent in more intensive attacks like a traditional air bombardment. Activists hold a banner for the Campaign to Stop Killer Robots. Many activists hope the United Nations enacts a ban on lethal autonomous weapons systems. Credit: Campaign to Stop Killer Robots (Creative Commons). I’ve worked for over a decade to help reduce civilian casualties in conflict, an effort sorely needed given the fact that most of those killed in war are civilians. I’ve looked, in great detail, at the possibility that automation in weapons systems could in fact protect civilians. Analyzing over 1,000 real-world incidents in which civilians were killed, I found that humans make mistakes (no surprise there) and that there are specific ways that AI could be used to help avoid them. There were two general kinds of mistakes: either military personnel missed indicators that civilians were present, or civilians were mistaken as combatants and attacked in that belief. Based on these patterns of harm from real world incidents, artificial intelligence could be used to help avert these mistakes. RELATED: Memorial Days: the racial underpinnings of the Hiroshima and Nagasaki bombings Though the debate often focuses on autonomous weapons, there are in fact three kinds of possible applications for artificial intelligence in the military: optimization of automated processing (e.g., improving signal to noise in detection), decision aids (e.g., helping humans to make sense of complex or vast sets of data), and autonomy (e.g., a system taking actions when certain conditions are met). While those calling for killer robots to be banned focus on autonomy, there are risks in all of these applications that should be understood and discussed. The risks fall in one of two basic categories: those associated with the intrinsic characteristics of AI (e.g., fairness and bias, unpredictability and lack of explainability, cyber security vulnerabilities, and susceptibility to tampering), and those associated with specific military applications of AI (e.g., using AI in lethal autonomous systems). Addressing these risks—especially those involving intrinsic characteristics of AI—requires a collaboration among members of the military, industry, and academia to identify and address areas of concern. Like the Google employees who pushed the company to abandon work on a computer vision program for the Pentagon, many people are concerned about whether military applications of artificial intelligence will be fair or biased. For example, will racial factors lead to some groups being more likely to be targeted by lethal force? Could detention decisions be influenced by unfair biases? For military personnel themselves, could promotion decisions incorporate and perpetuate historical biases regarding gender or race? Such concerns can be seen in another area where AI is already being used for security-related decisions: law enforcement. While US municipalities and other governmental entities aren’t supposed to discriminate against groups of people, particularly on a racial basis, analyses such as the Department of Justice investigation of the Ferguson, Mo., Police Department illustrate that biases nonetheless persist. Law enforcement in the United States is not always fair. A number of investigations have raised concerns that the AI-driven processes used by police or the courts—for instance, risk assessment programs to determine whether defendants should get paroled—are biased or otherwise unfair. Many are concerned that the pervasive bias that already exists in the criminal justice system introduces bias into the data on which AI-driven programs are trained to perform automated tasks. AI approaches using that data could then be affected by this bias. Academic researchers have been looking into how AI methods can serve as tools to better understand and address existing biases. For example, an AI system could pre-process input data to identify existing biases in processes and decisions. This could include identifying problematic practices (e.g., stop and frisk) as well as officers and judges who seem to make decisions or arrests that may be compromised by bias, reducing the role data from these processes or people has in a risk assessment. There are also ways to adjust the use of AI tools to help ensure fairness: for example, treating cases from different groups in a manner that is consistent with the way a particular group believed not to be subject to bias is treated. In such a way, AI—so often believed to be hopelessly bound to bias—can in fact be a tool to identify and correct existing biases. A Google office building. In a 2018 letter, Google employees cited fears of bias in AI to pressure the company to abandon an AI project to help the Pentagon analyze video taken by drones. Credit: The Pancake of Heaven! (Creative Commons). Similarly, the Pentagon could analyze which applications of artificial intelligence are inherently unsafe or unreliable in a military setting. The Defense Department could then leverage expertise in academia and industry to better characterize and then mitigate these types of risks. This dialogue could allow society to better determine what is possible and what applications should be deemed unsafe for military use. The US Defense Department’s 2018 AI strategy commits it to lead internationally in military ethics and AI safety, including by developing specific AI applications that would reduce the risk of civilian casualties. There’s no visible evidence yet of the Defense Department starting an initiative to meet this commitment, but other nations have begun practical work to develop such capabilities. For example, Australia is planning to explore this technology to better identify medical facilities in conflict zones, a much needed capability given the many such attacks in recent years. The Pentagon has taken some steps to prioritize AI safety. For example, the Defense Advanced Research Projects Agency, also known as DARPA, has a program that aims to develop explainable AI. AI systems can make decisions or produce results even while the how and the why behind those decisions or results is completely opaque to a human user. Steps to address this “black box” problem would be welcome, but they fall short of what is possible: a comprehensive approach to identify and systematically address AI safety risks. When it comes to lethal autonomous weapons, some say the time for talking is over and it’s time to implement a ban. After all, the argument goes, the United Nations has been meeting since 2014 to talk about lethal autonomous weapons systems, and what has been accomplished? Actually, though, there has been progress: The international community has a much better idea of the key issues, including the requirement for compliance with international law and the importance of context when managing the human-machine relationship. And the UN group of government experts has agreed to a number of principles and conclusions to help frame a collective understanding and approach. But more substantive talking is needed about the particulars, including the specific risks and benefits of autonomous weapons systems. And there is time. In 2012, the Pentagon created a policy on autonomous weapons (Directive 3000.09) requiring a senior level review before development could begin. Still, after eight years, not one senior level review has yet been requested, showing that the fielding or even the development of such capabilities is not imminent. Artificial intelligence may make weapons systems and the future of war relatively less risky for civilians than it is today. It is time to talk about that possibility.

#### This ability to identify excessive use of force and war crimes is currently being developed – meaning instituting it into LAWs is increasingly possible

Hao 20 Karen Hao is the artificial intelligence senior reporter for MIT Technology Review. "Human rights activists want to use AI to help prove war crimes in court." Published by MIT Technology Review on June 25, 2020. Available here: (https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/) - AP

In 2015, alarmed by an escalating civil war in Yemen, Saudi Arabia led an air campaign against the country to defeat what it deemed a threatening rise of Shia power. The intervention, launched with eight other largely Sunni Arab states, was meant to last only a few weeks, Saudi officials had said. Nearly five years later, it still hasn’t stopped. By some estimates, the coalition has since carried out over 20,000 air strikes, many of which have killed Yemeni civilians and destroyed their property, allegedly in direct violation of international law. Human rights organizations have since sought to document such war crimes in an effort to stop them through legal challenges. But the gold standard, on-the-ground verification by journalists and activists, is often too dangerous to be possible. Instead, organizations have increasingly turned to crowdsourced mobile photos and videos to understand the conflict, and have begun submitting them to court to supplement eyewitness evidence. But as digital documentation of war scenes has proliferated, the time it takes to analyze it has exploded. The disturbing imagery can also traumatize the investigators who must comb through and watch the footage. Now an initiative that will soon mount a challenge in the UK court system is trialing a machine-learning alternative. It could model a way to make crowdsourced evidence more accessible and help human rights organizations tap into richer sources of information. The initiative, led by Swansea University in the UK along with a number of human rights groups, is part of an ongoing effort to monitor the alleged war crimes happening in Yemen and create greater legal accountability around them. In 2017, the platform Yemeni Archive began compiling a database of videos and photos documenting the abuses. Content was gathered from thousands of sources—including submissions from journalists and civilians, as well as open-source videos from social-media platforms like YouTube and Facebook—and preserved on a blockchain so they couldn’t be tampered with undetected. Along with the Global Legal Action Network (GLAN), a nonprofit that legally challenges states and other powerful actors for human rights violations, the investigators then began curating evidence of specific human rights violations into a separate database and mounting legal cases in various domestic and international courts. “If things are coming through courtroom accountability processes, it’s not enough to show that this happened,” says Yvonne McDermott Rees, a professor at Swansea University and the initiative’s lead. “You have to say, ‘Well, this is why it’s a war crime.’ That might be ‘You’ve used a weapon that’s illegal,’ or in the case of an air strike, ‘This targeted civilians’ or ‘This was a disproportionate attack.’” In this case, the partners are focusing on a US-manufactured cluster munition, the BLU-63. The use and sale of cluster munitions, explosive weapons that spray out smaller explosives on impact, are banned by 108 countries, including the UK. If the partners could prove in a UK court that they had indeed been used to commit war crimes, it could be used as part of mounting evidence that the Saudi-led coalition has a track record for violating international law, and make a case for the UK to stop selling weapons to Saudi Arabia or to bring criminal charges against individuals involved in the sales. So they decided to develop a machine-learning system to detect all instances of the BLU-63 in the database. But images of BLU-63s are rare precisely because they are illegal, which left the team with little real-world data to train their system. As a remedy, the team created a synthetic data set by reconstructing 3D models of the BLU-63 in a simulation. Using the few prior examples they had, including a photo of the munition preserved by the Imperial War Museum, the partners worked with Adam Harvey, a computer vision researcher, to create the reconstructions. Starting with a base model, Harvey added photorealistic texturing, different types of damage, and various decals. He then rendered the results under various lighting conditions and in various environments to create hundreds of still images mimicking how the munition might be found in the wild. He also created synthetic data of things that could be mistaken for the munition, such as a green baseball, to lower the false positive rate. While Harvey is still in the middle of generating more training examples—he estimates he will need over 2,000—the existing system already performs well: over 90% of the videos and photos it retrieves from the database have been verified by human experts to contain BLU-63s. He’s now creating a more realistic validation data set by 3D-printing and painting models of the munitions to look like the real thing, and then videotaping and photographing them to see how well his detection system performs. Once the system is fully tested, the team plans to run it through the entire Yemeni Archive, which contains 5.9 billion video frames of footage. By Harvey’s estimate, a person would take 2,750 days at 24 hours a day to comb through that much information. By contrast, the machine-learning system would take roughly 30 days on a regular desktop. BLU-63 munitions strewn across rocks. The real image shown in the analysis at the top of the article. VFRAME Human experts would still need to verify the footage after the system filters it, but the gain in efficiency changes the game for human rights organizations looking to mount challenges in court. It’s not uncommon for these organizations to store massive amounts of video crowdsourced from eyewitnesses. Amnesty International, for example, has on the order of 1 terabyte of footage documenting possible violations in Myanmar, says McDermott Rees. Machine-learning techniques can allow them to scour these archives and demonstrate the pattern of human rights violations at a previously infeasible scale, making it far more difficult for courts to deny the evidence.

#### LAWs can help us to hold people more accountable for war crimes by providing information currently unavailable – outright bans prevent this

Müller 16 Vincent C. Müller is a Professor of Philosophy at Eindhoven University of Technology, University Fellow at Leeds, the president of the European Association for Cognitive Systems, and chair of the euRobotics topics group on 'ethical, legal and socio-economic issues'. "Autonomous Killer Robots Are Probably Good News." IN "Drones and Responsibility: Legal, Philosophical, and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons." Published by Ashgate in 2016. Available here: (https://philpapers.org/archive/MLLAKR.pdf) - AP

3.2.3. Narrowing the responsibility gap The responsibility framework outlined above shows how responsibility should be ascribed for many of the wrongful killings that could be committed by killer robots. The technology gives rise to a related and further beneficial effect, which is often not noted. Holding someone accountable for their action, e.g. for actual conviction for a war crime requires reliable information—which is often unavailable. The ability to acquire and store full digital data records of LAWS’ action and pre-mission inputs allows a better determination of the facts, and thus of actual allocation of responsibility, than is currently possible in the ‘fog of war’. As well as allowing allocation of responsibility, the recording of events is also likely to diminish the likelihood of wrongful killings. There is already plenty of evidence that, for example, police officers who have to video their own actions are much less likely to commit crimes. So, killer robots would actually reduce rather than widen responsibility gaps. 3.2.4. Regulation and standards The foregoing has the following implication: moral interest should be focused on the determination of the technical standards of reliability which robots—including killer robots—should meet. The recent EU ‘RoboLaw’ report makes a parallel point, in arguing that we should resist the urge to say that ‘robots are special’ in terms of responsibility. Rather, we should adopt a functional perspective and see whether the new technology really does require new legal regulation, and in which areas (based on Bertolini 2014; Palmerini et al. 2014: 205f). This seems to be a move in the right direction: We already devise automated systems (e.g. automated defence of ships against air attacks) where the ‘rules of engagement’ are put into software. The same ‘due care’ is to be expected for the manufacture and use of LAWS. Just like for civil autonomous cars, we need to specify standards that LAWS manufacturers must abide by. These standards must ensure that the robot acts according to the principles of distinction and proportionality (this is already possible now if one thinks of targeting tanks, ships, planes or artillery, for example). Both manufacturing and distributing LAWS that do not abide by these standard would be a war crime. If a killer robot is manufactured with due care according to these standards but commits a war crime, due to use in situations for which it was not designed or licensed, the crime is the responsibility of the soldier/user. The responsible person for a particular command or action can be identified in the military chain of command – this is a deeply entrenched tradition. Finally, if the soldiers can show that they exercised due care, then the deaths are accidents. Regulation of LAWS thus requires two components. First, there are the technical standards of reliability which LAWS must meet; pertinently, what degrees of reliability LAWS must meet in terms of distinction and proportionality in their decisions to attack. Second, there are the legal instruments by which accountability is to be exercised over those who fail to manufacture, distribute or deploy LAWS in accordance with those standards. Each dimension—that of technical standards and of law— should be subject to enforcement at the international and national levels. The proposed policy structure can thus be schematically presented as a matrix: Autonomous Killer Robots Are Probably Good News 11/16 Legal and technical regulation International National Legal International Humanitarian Law Criminal Law Technical Technical standards for performance Control regimes for technical standards; special standards

### US Hegemony

#### The development of Lethal Autonomous weapons is inevitable – refusing research places countries at a major disadvantage and can exacerbate AI’s worst aspects

Kirsch 18 Andreas Kirsch is currently a Fellow at Newspeak House in London after having worked at Google and DeepMind in Zurich and London as a software and research engineer. "Autonomous weapons will be tireless, efficient, killing machines—and there is no way to stop them." Published by Quartz on July 23, 2018. Available here: (https://qz.com/1332214/autonomous-weapons-will-be-tireless-efficient-killing-machines-and-there-is-no-way-to-stop-them/) - AP

The world’s next major military conflict could be over quickly. Our human soldiers will simply not stand a chance. Drones and robots will overrun our defenses and take the territory we are standing on. Even if we take out some of these machines, more of them will soon arrive to take their place, newly trained off our reactions to their last offense. Our own remote-controlled drones will be outmaneuvered and destroyed, as no human operator can react quickly enough to silicon-plotted attacks. This isn’t a far-off dystopian fantasy, but a soon-to-be-realized reality. In May, Google employees resigned in protest over the company helping the US military develop AI capabilities for drones. (The company ultimately decided to shelve the project.) More recently, 2,400 researchers vowed not to develop autonomous weapons. Many AI researchers and engineers are reluctant to work on autonomous weapons because they fear their development might kick off an AI arms race: Such weapons could eventually fall into the wrong hands, or they could be used to suppress the civilian population. How could we stop this from happening? The first option is developing a non-proliferation treaty to ban autonomous weapons, similar to the non-proliferation treaty for nuclear weapons. Without such a treaty, the parties voluntarily abstaining from developing autonomous weapons for moral reasons will have a decisive disadvantage. Nobody mourns them or asks for their bodies to be returned from war. That’s because autonomous weapons have many advantages over human soldiers. For one, they do not tire. They can be more precise, and they can react faster and operate outside of parameters in which a human would survive, such as long stints in desert terrains. They do not take years of training and rearing, and they can be produced at scale. At worst they get destroyed or damaged, not killed or injured, and nobody mourns them or asks for their bodies to be returned from war. It is also easier to justify military engagements to the public when autonomous weapons are used. As human losses to the attacker’s side are minimal, armies can keep a low profile. Recent engagements by the US and EU in Libya, Syria, and Yemen have focused on using drones, bombing campaigns, and cruise missiles. Parties without such weapons will have a distinct handicap when their soldiers have to fight robots. But even if all countries signed an international treaty to ban the development of autonomous weapons, as they once did for nuclear non-proliferation, it would be unlikely to prevent their creation. This is because there are stark differences between the two modes of war. There are two properties that make 1958’s nuclear non-proliferation treaty work quite well: The first one is a lengthy ramp-up time to deploying nuclear weapons, which allows other signatories to react to violations and enact sanctions, and the second one is effective inspections. To build nuclear weapons, you need enrichment facilities and weapons-grade plutonium. You cannot feasibly hide either and, even when hidden, traces of plutonium are detected easily during inspections. It takes years, considerable know-how, and specialized tools to create all the special-purpose parts. Moreover, all of the know-how has to be developed from scratch because it is secret and import-export controlled. And even then, you still need to develop missiles and means of deploying them. But it’s the opposite with autonomous weapons. To start, they have a very short ramp-up time: Different technologies that could be used to create autonomous weapons already exist and are being developed independently in the open. For example, tanks and fighter planes have lots of sensors and cameras to record everything that is happening, and pilots already interface with their plane through a computer that reinterprets their steering commands. They just need to be combined with AI, and suddenly they have become autonomous weapons. AI research is progressing faster and faster as more money is poured in by both governments and private entities. Progress is not only driven by research labs like Alphabet’s DeepMind, but also by game companies. Recently, EA’s SEED division began to train more general-purpose AIs to play its Battlefield 1 game. After all, AI soldier don’t need to be trained on ground: Elon Musk’s OpenAI has published research on “transfer learning,” which allows AIs to be trained in a simulation and then adapted to the real world. It’s much harder to spot an AI for autonomous weapons than it is to spot the creation of a nuclear weapon. This makes effective inspections impossible. Most of the technologies and research needed for autonomous weapons are not specific to them. In addition, it’s much harder to spot an AI for autonomous weapons than it is to spot the creation of a nuclear weapon. AIs can be trained in any data center: They are only code and data, after all, and code and data can now easily be moved and hidden without leaving a trace. Most of their training can happen in simulations on any server in the cloud, and running such a simulation would look no different to outside inspectors from predicting tomorrow’s weather forecast or training an AI to play the latest Call of Duty. Without these two properties, a treaty has no teeth and no eyes. Signatories will still continue to research general technologies in the open and integrate them into autonomous weapons in secret with low chances of detection. They will know that others are likely doing the same, and that abstaining is not an option. So, what can we do? We cannot shirk our responsibilities. Autonomous weapons are inevitable. Both offensive and defensive uses of autonomous weapons need to be researched, and we have to build up deterrent capabilities. However, even if we cannot avoid autonomous weapons, we can prevent them from being used on civilians and constrain their use in policing. There can be no happy ending, only one we can live with.

#### While the US currently is leading in AI ongoing work especially militarily must be done to maintain strategic advantage

Sherman 19 Justin Sherman was a Cybersecurity Policy Fellow at New America. "Reframing the U.S.-China AI "Arms Race." Published by New America on March 6, 2019. Available here: (https://www.newamerica.org/cybersecurity-initiative/reports/essay-reframing-the-us-china-ai-arms-race/) - AP

Artificial Intelligence and State Power Artificial intelligence is poised to contribute greatly to bolstering a developed nation’s economy. Accenture Research and Frontier Economics predict, based on research in 12 developed countries, that AI could “double annual economic growth rates” in 2035 while also increasing labor productivity by up to 40 percent.49 McKinsey Global Institute predicts AI may deliver $13 trillion in global economic activity by 2030.50 PricewaterhouseCoopers puts that figure even higher at up to $15.7 trillion in global GDP growth by 2030, much of which will be due to productivity increases.51 These estimates are varied, but they all rightfully predict enormous economic growth due to an explosion in AI uses worldwide.52 However, these gains will not be evenly spread. As research from McKinsey Global Institute articulates, “leaders of AI adoption (mostly in developed countries) could increase their lead over developing countries,” and “leading AI countries could capture an additional 20 to 25 percent in net economic benefits, compared with today, while developing countries might capture only about 5 to 15 percent.”53 With the United States and China already representing the largest economies in the world, maximizing uses of AI within either nation could lead to massive gains in state power and influence on the global stage. “After all,” writes political scientist Michael Horowitz, “countries cannot maintain military superiority over the medium to long term without an underlying economic basis for that power.”54 Further, there is in part a question of pure economic power: If Chinese companies don’t just develop better but also use that AI more profitably than American firms, China benefits economically and by extension has more resources to build state power generally. That the United States currently has significant AI talent does not mean an American edge in AI development is decisive and everlasting. Militarily speaking, artificial intelligence is also revolutionary for state military power. The People’s Liberation Army (PLA) in China views AI as a revolutionary factor in military power and civil-military fusion,55 just as the U.S. Department of Defense has similarly recognized how advances in artificial intelligence “will change society and, ultimately, the character of war.”56 China is investing in this future. The PLA has already funded a number of AI military projects as part of its 13th Five-Year Plan, spanning command decision-making, equipment systems, robotics, autonomous operating guidance and control systems, advanced computing, and intelligent unmanned weapon systems.57 In 2017, President Xi Jinping called for the military to accelerate AI research in preparation for the future of war.58 There has even been a report of the Beijing Institute of Technology recruiting high-talent teenagers for a new AI weapons development program.59 The Chinese government is undoubtedly preparing to maximize its AI development in the service of maximizing its military power. That the United States currently has significant AI talent does not mean an American edge in AI development is decisive and everlasting. The United States has started to do the same, in some respects: It has established a Defense Innovation Board for ethics of AI in war,60 as well as a Joint Artificial Intelligence Center to develop “standards….tools, shared data, reusable technology, processes, and expertise” in coordination with industry, academia, and American allies.61 DARPA, the Defense Advanced Research Projects Agency, currently has 25 programs in place focused on artificial intelligence research, and in September 2018, its director announced a plan to spend up to $2 billion over the next five years on more AI work.62 But there is still much to be done, as I’ll address in the last section. Even within the U.S. military’s approaches to artificial intelligence, as one West Point scholar notes, “the military is facing some hard questions about how it will adapt its culture and institutions to exploit new technologies—and civilians face a tough job ensuring they answer them effectively.”63 There are certainly military leaders aware of this fact—in announcing the $2 billion in AI funding, DARPA’s director depicted it “as a new effort to make such systems more trusted and accepted by military commanders”64—yet the road ahead will have its challenges. In general, the U.S. defense apparatus’ willingness to engage in cultural and operational shifts will greatly influence how successfully AI is integrated into the United States military. It’s also important to note that China’s government and its private companies will likely be less constrained by ethical and legal norms when developing AI than will their American counterparts.65 Faster deployment of and greater experimentation with AI may result, even though this may lead to perhaps chaotic or more unpredictable deployments of artificial intelligence—or, perhaps, plainly unethical uses of AI. This leads into the second main reason why U.S.-China AI competition still matters.

#### Military AI Development could prove instrumental in deterrence capabilities – even in scenarios that AI fizzles out the most strategic option is to balance it with other technologies, not remove it.

Kallenborn 19 Zachary Kallenborn is a freelance researcher and analyst, specializing in Chemical, Biological, Radiological, and Nuclear (CBRN) weapons, CBRN terrorism, drone swarms, and emerging technologies writ large. "WHAT IF THE U.S. MILITARY NEGLECTS AI? AI FUTURES AND U.S. INCAPACITY." Published by War On the Rocks on September 3, 2019. Available here: (https://warontherocks.com/2019/09/what-if-the-u-s-military-neglects-ai-ai-futures-and-u-s-incapacity/) - AP

AI could threaten the credibility of the U.S. nuclear deterrent. Although constant, real-time tracking of all nuclear submarines is difficult to imagine due to the massive size of the oceans, technology improvements and some luck could allow an adversary to know the locations of second-strike platforms for long enough to eliminate them in a first strike. Swarms of undersea drones and big data analysis offer great potential for new and improved anti-submarine platforms, weapons, and sensor networks. Already, some missile defenses use simple automation that could be improved with AI. Drones can also help track missiles, serve as platforms to defeat them, or simply collide with incoming missiles and aircraft. AI improvements generally enable more advanced robotic weapons, more sophisticated swarms, and better insights into data. Of course, the long history of failed attempts and huge costs of missile defense suggest elimination of nuclear deterrence is highly unlikely, but all of these developments could add up to serious risks to the reliability of nuclear deterrence. In such a world, a United States without robust military AI capabilities is extremely insecure. The United States has neither conventional superiority nor a reliable nuclear deterrent, and must drastically rethink American grand strategy. U.S. extended deterrence guarantees would be far less effective and some states under the umbrella would likely seek their own nuclear weapons instead. South Korea and Saudi Arabia would likely become nuclear weapons states due to their established civilian nuclear programs, high relative wealth, and proximity to hostile powers in possession or recent pursuit of nuclear weapons. The United States could expand its nuclear arsenal to mitigate the harms of a less reliable deterrent, but that would require abandoning the New Strategic Arms Reduction Treaty and other arms control treaties. Ensuring national security would mean avoiding conflict or focusing on homeland defense — rather than a forward defense posture with forces stationed on the Eurasian landmass — to increase adversary costs. Diplomacy, soft power, and international institutions remain key to national security. However, a soft-power strategy would be extremely challenging. The factors that could inhibit development of AI — domestic dysfunction, high debt, and international isolation — would cause considerable harm to U.S. soft power. American soft power is arguably already in decline and funding for the State Department and U.S. Agency for International Development have been cut considerably. Likewise, any abandonment of arms control treaties to support the nuclear arsenal would cause further damage. In short, in AI Trinity, a United States without AI is no longer a serious global power. AI Fizzle The year is 2040 and dreams of a robotic future remain a fantasy. During the early 2020s, implementation of machine learning and data analysis techniques expanded, creating some organizational and logistical efficiencies and reduced costs. But those changes were not transformative. Some states developed AI-powered autonomous platforms, but the battlefield impact was limited. A well-placed jammer or microwave weapon could defeat even large masses of autonomous systems. The possibility of AI Fizzle has not been given enough serious consideration. Future AI may not handle battlefield complexities well enough to prove useful. True robotic dominance may require human levels of AI, which will likely take over 80 years or more given how little neuroscientists know about the human brain. Just autonomously distinguishing between non-combatant and combatants is unlikely in the near term. Advances in AI may also slow. During the 1980s, AI research entered a so-called “winter” in which research funding cuts, rejections of older AI systems, and market shifts resulted in a lull in breakthroughs and public interest. Particular AI techniques may also go through dark periods, as during the 1970s and 1990s when funding and innovations in neural networks dried up. Some already predict a coming AI winter. In this world, the costs of U.S. limited development of AI are minimal and may be a net positive. Resources and leadership attention spent on encouraging AI may be directed to other, ultimately more impactful capabilities. For example, gaps in the suppression of enemy air defense mission could prove more consequential than AI in the short run. Challenges unrelated to technology, such as defense mobilization, may matter most. Other emerging technologies, such as 3-D printing and nanotechnology, also may prove more transformative than AI. 3-D printing may revolutionize manufacturing and nanotechnologies may lead to extremely low-cost sensors, self-healing structures, and ultra-light materials. In this scenario, if the United States focuses on these technologies while adversaries focus on AI, the United States would gain first-mover advantages and a more robust capability. Alternatively, no single emerging technology may prove transformative. Various emerging technologies may provide real, but not major benefits. If so, then the United States should find the right combination of technologies that best support security needs, and apply and integrate them into the defense establishment. A balance ought to be struck between emerging and established technologies — sometimes tried and true is best

### Bans Fail

#### Lack of transparency in AI systems causes bans to fail by limiting the ability to verify whether particular technologies violate the ban

Morgan et al 20 Forrest E. Morgan is a senior political scientist at the RAND Corporation and an adjunct professor at the University of Pittsburgh Graduate School of Public and International Affairs. Ben Boudreaux is a professor at Pardee RAND Graduate School and a policy researcher at RAND working in the intersection of ethics, emerging technology, and national security. Andrew Lohn is an engineer at the RAND Corporation and a professor of public policy at the Pardee RAND Graduate School. Mark Ashby, Research Assistant at RAND Corporation. Christian Curriden is a defense analyst at the RAND Corporation. Kelly Klima is associate program director of the Acquisition and Development Program (ADP) for the Homeland Security Operational Analysis Center (HSOAC). Derek Grossman is a senior defense analyst at RAND focused on a range of national security policy and Indo-Pacific security issues. "Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World." Published by the Rand Corporation in 2020. Available here: (https://apps.dtic.mil/sti/pdfs/AD1097313.pdf) - AP

A significant number of countries supports a new legally binding treaty that would ban the development and use of LAWS. The Campaign to Stop Killer Robots has circulated a list of 26 states that support a ban. However, most of the states supporting a new legal instrument are developing countries that do not possess sophisticated AI technology sectors or have military forces with extensive AI capabilities. Meanwhile, most of the major military powers perceive significant value in military AI and do not wish to create new international constraints that could slow its technological development. The United States, the United Kingdom, Russia, and other countries hold that existing international law, including LOAC, already provides significant humanitarian protections regarding the use of LAWS, and thus no new treaty instruments are necessary. China, conversely, has proposed a ban on LAWS modeled on the UN protocol prohibiting the use of blinding laser weapons, but it seems to define LAWS so narrowly that a ban on this class of weapons would not apply to systems currently under development. China has also suggested that concepts such as meaningful human control should be left up to sovereign determination, rather than defined through international processes. Thus, it appears that China’s professed support for a new legal instrument would not actually constrain the development or use of military AI. In addition, some states have questioned what verification and monitoring measures would be associated with any new international ban. Given the inherent lack of transparency of many AI systems, states have expressed concern that signatories to any ban might not live up to their international commitments. As a result, many governments, including those of France, Germany, and other European states, have supported simply developing a nonbinding political declaration that would articulate the importance of human control being designed into and exercised across the acquisition, development, testing, and deployment life cycle of military AI systems. A nonbinding declaration or code of conduct of this sort would be easier to reach than a new treaty, but other states have expressed doubt that it would be useful, since it could not be enforced. Given the resistance of several major military powers and the need for their acquiescence to a new treaty, the international community is not likely to agree to a ban or other regulation in the near term. However, there is a view broadly resonant among many countries, including the United States, key allies, and important stakeholders, such as the International Committee on the Red Cross, that further international discussion regarding the role of humans in conducting warfare is necessary.

#### Bans are Unnecessary – Automated weapon systems are still governed by current international legal standards which prevent the abuses highlighted in the affirmative

Scheffler and Ostling 19 Sarah Scheffler, Ph.d Student at Boston University Computer Science Department. Jacob Ostling, Boston University School of Law and Brown Rudnick LLP. "Dismantling False Assumptions about Autonomous Weapon Systems." Published by the Association for Computing Machinery, Published in 209. Available here: (https://cs-people.bu.edu/sscheff/Autonomous\_Weapons\_False\_Assumptions.pdf) - AP

Five broad principles limit the use of weapons under international law: (1) unnecessary or superfluous suffering; (2) military necessity; (3) proportionality; (4) distinction; and (5) command responsibility or “accountability”.36 Each of the five are recognized to some extent as customary international law (“CIL”), so they are binding on states to some extent regardless of treaty status.37 Thus, if a weapon can never be used in a manner that comports with each standard, then it is per se unlawful.38 There is nothing inherent to AWS that prevents them from abiding by each of these principles, but each principle does impose limits on their use as applied. 39 Even if AWS as a class of weapon system are not per se illegal, the principles of IHL, especially distinction, proportionality, and accountability do impose substantive restrictions on the development and deployment of AWS.40 We describe the restrictions (or lack thereof) on AWS for each of the five principles. 3.1 Suffering The prohibition on unnecessary or superfluous suffering outlaws weapons which cause suffering to combatants with no military purpose.41 The principle is codified in AP 1, Art.35(2), and as applied, is concerned with a weapon’s effect on combatants (i.e. poisoning), not the platform used to deliver that weapon.42 While an AWS has the potential mis-judge the amount of suffering it will cause, nothing in the use of an AWS as a delivery platform modifies the harm inflicted by a particular type of weapon.43 Thus, AWS could comply with the rule by employing any traditionally legal weapon. 3.2 Military necessity Military necessity requires that a weapon provide an advantage for legitimate military objectives.44 This principle is augmented by the rule of precaution in attack, codified in Article 57 of Additional Protocol One, 45 but also reflecting CIL, which requires that attackers exercise “constant care . . . to spare the civilian population.”46 In particular, with respect to the means of warfare used, Article 57 requires that attackers use the means least likely to harm civilians, unless doing so would sacrifice some military advantage.47 Consequently, to satisfy both rules, AWS must avoid civilian casualties at least as well as existing weapon systems or provide some other military advantage unavailable from those systems. This is a low threshold in practice, because autonomy does offer numerous advantages. Autonomous systems reduce the need for communication between a system and a human pilot or operator.48 This is especially useful compared to remote-controlled systems in environments where communication is denied or difficult,49 but even in environments with assured communications, autonomy frees up communication bandwidth for other uses and allows reaction speeds quicker than communication latency would permit.50 Unlike humans, who suffer cognitive fatigue after time has passed and suffer from stress, autonomous machines generally continue functioning at full potential for as long as they remain turned on.51 They may directly use sensors more advanced than human senses, and they can integrate many different data sources effectively.52 The quick reaction time of autonomous systems may be useful in applications where human reaction time is insufficient to address incoming threats.53 Autonomous systems could further allow for reduction of expensive personnel such as pilots and data analysts.54 Finally, unmanned autonomous weapons need not prioritize their self-preservation, enabling them to perform tasks that might be suicidal for manned systems.55 3.3 Proportionality Proportionality, as codified in Article 51(5)(b) of Additional Protocol One, prohibits attacks “which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.”56 IHL does not offer any numerical definition for what constitutes an “excessive” ratio of civilian casualties, nor has any international consensus developed.57 In practice, the standard has been treated as requiring that civilian casualties be “reasonable” in relation to military advantage.58 This reasonableness standard may be satisfied by selecting targets based on assessments which use numerical values to weigh risks to civilians, such as the U.S. Collateral Damage Estimation Methodology.59 Today’s AWS can only meet half of this standard. There is no technical barrier preventing an AWS from programatically determining acceptable distributions of collateral damage using existing frameworks, however, “military advantage” is considered a subjective case-by-case evaluation.60 There are some initial attempts at technical methods for assessing proportionality from first principles, but they leave much to be desired both in terms of practicality and validity.61 In the near term, human commanders can perform this assessment and pre-specify conditions under which the AWS can act without violating proportionality.62 Moreover, environments without civilians (e.g., undersea) offer venues where the proportionality assessment is likely to be fairly straightforward. Thus, AWS can satisfy proportionality. 3.4 Distinction Distinction, codified in Article 48 of Additional Protocol One, requires that parties to a conflict “at all times distinguish between the civilian population and combatants and between civilian objects and military objectives.”63 Distinction only renders a weapon per se unlawful if it is incapable of being directed at a specific military objective, although such a weapon may be unlawful as applied for failure to distinguish between combatants and civilians during use.64 Attackers are further required to err on the side of caution where there is doubt as to whether a target is a civilian or a combatant.65 Today’s AWS are are generally considered unable to distinguish between combatants and non-combatants in urban warfare conditions or other environments that mix combatants and non-combatants,66 but even absent the ability to distinguish at this high level, AWS could legally be used in scenarios where there are no civilians present (e.g. undersea submarines or missile defense).67 3.5 Command accountability Lastly, command responsibility, also known as “accountability,” sets forth the requirement that superiors be held liable for war crimes committed by their subordinates, if they knew or should have known of the crimes, and failed to take reasonable measures to prevent those crimes.68 Legal systems which impose responsibilities on superior officers to uphold the law are generally sufficient to satisfy the command responsibility requirement.69 Human Rights Watch (HRW) has argued that AWS violate the principle of command responsibility “[s]ince there is no fair and effective way to assign legal responsibility for unlawful acts committed by fully autonomous weapons.”70 This is incorrect; humans decide how to deploy AWS, and may be properly held liable for failing to do so in accordance with the law.71 As long as human commanders choose when and how to deploy AWS, responsibility for the weapon’s actions rests with the commander, not the weapon itself. IHL dictates a requirement that commanders must actively ensure that weapons they employ adhere to LOAC.72 The DoD has also imposed this responsibility unambiguously, stating that “[p]ersons who authorize the use of . . . autonomous weapons must do so with appropriate care and in accordance with the law of war, applicable treaties, weapon system safety rules, and applicable rules of engagement (ROE).”73 The threat of repercussions to those who would recklessly deploy AWS is a substantive restriction on their use.

#### Banning the weapons themselves does nothing – automated targeting will aid in combat violence and atrocity regardless of the automated attack

Michel 20 Arthur Holland Michel is the co-director of the Center for the Study of the Drone at Bard College. "The Killer Algorithms Nobody’s Talking About." Published by Foreign Policy on January 20, 2020. Available here: (https://foreignpolicy.com/2020/01/20/ai-autonomous-weapons-artificial-intelligence-the-killer-algorithms-nobodys-talking-about/) - AP

This past fall, diplomats from around the globe gathered in Geneva to do something about killer robots. In a result that surprised nobody, they failed. The formal debate over lethal autonomous weapons systems—machines that can select and fire at targets on their own—began in earnest about half a decade ago under the Convention on Certain Conventional Weapons, the international community’s principal mechanism for banning systems and devices deemed too hellish for use in war. But despite yearly meetings, the CCW has yet to agree what “lethal autonomous weapons” even are, let alone set a blueprint for how to rein them in. Meanwhile, the technology is advancing ferociously; militaries aren’t going to wait for delegates to pin down the exact meaning of slippery terms such as “meaningful human control” before sending advanced warbots to battle. To be sure, that’s a nightmarish prospect. U.N. Secretary-General António Guterres, echoing a growing chorus of governments, think tanks, academics, and technologists, has called such weapons “politically unacceptable” and “morally repugnant.” But this all overlooks an equally urgent menace: autonomous systems that are not in themselves lethal but rather act as a key accessory to human violence. The debate over so-called killer robots overlooks an equally urgent menace: autonomous systems that are not in themselves lethal but rather act as a key accessory to human violence. Such tools—let’s call them lethality-enabling autonomous systems—might not sound as frightening as a swarm of intelligent hunter drones. But they could be terrifying. At best, they will make conflict far more unpredictable and less accountable. At worst, they could facilitate ghoulish atrocities. Many such technologies are already in use. Many more are right around the corner. And because of our singular focus on headline-grabbing killer robots, they have largely gone ignored. Militaries and spy services have long been developing and deploying software for autonomously finding “unknown unknowns”—potential targets who would have otherwise slipped by unnoticed in the torrent of data from their growing surveillance arsenals. One particularly spooky strand of research seeks to build algorithms that tip human analysts off to such targets by singling out cars driving suspiciously around a surveilled city. Other lethality-enabling technologies can translate intercepted communications, synthesize intelligence reports, and predict an adversary’s next move—all of which are similarly crucial steps in the lead-up to a strike. Even many entry-level surveillance devices on the market today, such as targeting cameras, come with standard features for automated tracking and detection. For its part, the U.S. Department of Defense, whose self-imposed rules for autonomous weapons specifically exempt nonlethal systems, is allowing algorithms dangerously close to the trigger. The Army wants to equip tanks with computer vision that identifies “objects of interest” (translation: potential targets) along with recommendation algorithms—kind of like Amazon’s—that advise weapons operators whether to destroy those objects with a cannon or a gun, or by calling in an airstrike. All of these technologies fall outside the scope of the international debate on killer robots. But their effects could be just as dangerous. The widespread use of sophisticated autonomous aids in war would be fraught with unknown unknowns. An algorithm with the power to suggest whether a tank should use a small rocket or a fighter jet to take out an enemy could mark the difference between life and death for anybody who happens to be in the vicinity of the target. An algorithm with the power to suggest whether a tank should use a small rocket or a fighter jet to take out an enemy could mark the difference between life and death for anybody who happens to be in the vicinity of the target. But different systems could perform that same calculation with widely diverging results. Even the reliability of a single given algorithm could vary wildly depending on the quality of the data it ingests. It is also difficult to know whether lethality-enabling artificial intelligence—prone as computers are to bias—would contravene or reinforce those human passions that all too often lead to erroneous or illegal killings. Nor is there any consensus as to how to ensure that a human finger on the trigger can be counted on as a reliable check against the fallibility of its algorithmic enablers. As such, in the absence of standards on such matters, not to mention protocols for algorithmic accountability, there is no good way to assess whether a bad algorithmically enabled killing came down to poor data, human error, or a deliberate act of aggression against a protected group. A well-intentioned military actor could be led astray by a deviant algorithm and not know it; but just as easily, an actor with darker motives might use algorithms as a convenient veil for an intentionally insidious decisions. Automation’s vast potential to make humans more efficient extends to the very human act of committing war crimes. If one system offers up a faulty conclusion, it could be easy to catch the mistake before it does any harm. But these algorithms won’t act alone. A few months ago, the U.S. Navy tested a network of three AI systems, mounted on a satellite and two different airplanes, that collaboratively found an enemy ship and decided which vessel in the Navy’s fleet was best placed to destroy it, as well as what missile it should use. The one human involved in this kill chain was a commanding officer on the chosen destroyer, whose only job was to give the order to fire. Eventually, the lead-up to a strike may involve dozens or hundreds of separate algorithms, each with a different job, passing findings not just to human overseers but also from machine to machine. Mistakes could accrue; human judgment and machine estimations would be impossible to parse from one another; and the results could be wildly unpredictable. These questions are even more troubling when you consider how central such technologies will become to all future military operations. As the technology proliferates, even morally upstanding militaries may have to rely on autonomous assistance, in spite of its many risks, just to keep ahead of their less scrupulous AI-enabled adversaries. Once an AI system can navigate complicated circumstances more intelligently than any team of soldiers, the human will have no choice but to take its advice on trust And once an AI system can navigate complicated circumstances more intelligently than any team of soldiers, the human will have no choice but to take its advice on trust—or, as one thoughtful participant at a recent U.S. Army symposium put it, targeting will become a matter of simply pressing the “I-believe button.” In such a context, assurances from top brass that their machines will never make the ultimate lethal decision seem a little beside the point. Most distressing of all, automation’s vast potential to make humans more efficient extends to the very human act of committing war crimes. In the wrong hands, a multi-source analytics system could, say, identify every member of a vulnerable ethnic group.

## Extensions

### Bans Fail

#### Even if a Ban would be positive – definitional issues and variations create significant legal challenges to bans

Rosert and Sauer 20 Elvira Rosert is a Junior Professor for International Relations at Universität Hamburg and the Institute for Peace Research and Security Policy in Hamburg. Frank Sauer is a Senior Researcher at Bundeswehr University Munich. His work covers nuclear issues, terrorism, and cyber-security, as well as emerging military technologies. "How (not) to stop the killer robots: A comparative analysis of humanitarian disarmament campaign strategies." Published by Contemporary Security Policy, published online on May 30, 2020. Available here: (https://www.tandfonline.com/doi/full/10.1080/13523260.2020.1771508) - AP

Conclusion In this article, we set out to answer how an international, legally binding regulation of LAWS can be brought about. Humanitarian advocacy campaigns wield significant influence in general; the Campaign to Stop Killer Robots does so in particular. We thus focused on its strategy in engaging the international community at the CCW in Geneva, the epicenter of the debate surrounding a possible regulation of LAWS. We found the campaign’s strategy to be less than optimal. As our comparative analysis of three humanitarian disarmament processes revealed, the campaign against LAWS is modeled after past successes, despite weapon autonomy differing from blinding lasers or landmines in several important ways. These differences limit the portability of some tried-and-tested strategy components. Actor-related components such as awareness-raising, dissemination of expertise, and coalition-building are similar in the three campaigns against LAWS, BLW and APL, and appear to be conducive to the goal of a ban on LAWS too. However, rehashing the issue- and institution-related components of the BLW and APL campaign strategies creates weak spots in the case of LAWS. The “killer robots” frame, for instance, while attempting to convey a simple and dramatic message, also renders the issue futuristic and, thus, less urgent. The prominent focus on the indiscriminateness of LAWS is an attempt to activate an argument that proved powerful against APL and CM, but might turn out to be obsolete in the case of LAWS due to technological improvements. Most importantly, LAWS are portrayed as a category of weapons, which is not accurate because weapon autonomy is an elusive function in a human-machine system. Lastly, due to the lack of critical mass and “champion state” leadership, the LAWS process is not (yet) ripe for a venue shift. Our approach confirms the necessity of the components we studied, but it cannot (nor was it designed to) specify their relative importance or identify sufficient combinations. That said, our findings do highlight an aspect that deserves further attention: the fit of the framing to the issue. This insight is less trivial than it may seem. In theoretical literature thus far, the framing’s fit has been considered almost exclusively with regard to different audiences and normative environments. The issue itself has remained neglected. Having shown how a mismatch between the framing and the issue’s key characteristics can compromise a campaign’s message, we suggest exploring the relevance of the framing/issue fit in additional cases. Our findings also suggest that modifying the substance of the argument, the expected regulatory design, and the institutional factors would increase the likelihood of the KRC’s strategy achieving its stated goal. In terms of substance, the most straightforward argument against LAWS is not a legal but an ethical one, namely, the argument that delegating life and death decisions to machines infringes upon human dignity. We therefore propose moving further away from the KRC’s initial messaging, which was heavily focused on the indiscriminateness of LAWS, their incompatibility with IHL, and the plight of civilians. Shifting toward more fundamental ethical concerns will, first, make the case against LAWS less susceptible to consequentialist counter-positions (which argue that the illegality of LAWS will be remedied by technological progress). Second, it makes it more likely that the general public will react viscerally and reject LAWS more sharply (Rosert & Sauer, 2019; Sharkey, 2019, p. 83). In terms of regulatory design, the complex and polymorphic nature of weapon autonomy represents a special challenge. “Ban killer robots” sounds straightforward, but it is not as cut-and-dried as “ban anti-personnel landmines” due to the sheer amount of endless variations on what “killer robots” might look like. The LAWS debate within the CCW is thus less firmly grounded in existing IHL principles and more prone to definitional struggles. Therefore, it is encouraging that the CCW’s focus is currently shifting from a categorical definition of LAWS toward the role of the “human element,” that is, the creation of conditions to retain meaningful human control. We strongly suggest doubling down on the corresponding regulatory option, namely, codifying meaningful human control as a principle requirement in IHL (Rosert, 2017). The KRC has already begun to embrace this idea of a positive obligation in its latest working paper on the key elements of a future treaty (Campaign to Stop Killer Robots, 2019a). Lastly, instead of taking the issue outside the UN framework, we suggest exploring the option of a venue reform that targets the CCW’s mode of decision-making. Consensus, while being the traditional rule in Geneva, is not necessarily required under the convention. The CCW could, theoretically, resort to majority voting (United Nations, 1995, pp. 454–455). Hence, the LAWS issue has not only the potential to come to fruition and result in a legally binding instrument from the CCW, but also to induce institutional change and restore the CCW’s originally intended function—a development from which future norm-setting processes would benefit as well.