Routledge
Taylor & Francis Group

Journal of Strategic Studies

ISSN: 0140-2390 (Print) 1743-937X (Online) Journal homepage: www.tandfonline.com/journals/fjss20

When speed kills: Lethal autonomous weapon systems, deterrence and stability

Michael C. Horowitz

To cite this article: Michael C. Horowitz (2019) When speed kills: Lethal autonomous weapon systems, deterrence and stability, Journal of Strategic Studies, 42:6, 764-788, DOI: 10.1080/01402390.2019.1621174

To link to this article: https://doi.org/10.1080/01402390.2019.1621174

Published online: 22 Aug 2019.

JOURNAL OF STRATEGIC STUDIES
2019, VOL. 42, NO. 6, 764–788
https://doi.org/10.1080/01402390.2019.1621174

Routledge
Taylor & Francis Group

ARTICLE

When speed kills: Lethal autonomous weapon systems, deterrence and stability

Michael C. Horowitz

Political Science, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT
While the applications of artificial intelligence (AI) for militaries are broad, lethal autonomous weapon systems (LAWS) represent one possible usage of narrow AI by militaries. Research and development on LAWS by major powers, middle powers and non-state actors makes exploring the consequences for the security environment a crucial task. This article draws on classic research in security studies and examples from military history to assess the potential development and deployment of LAWS, as well as how they could influence arms races, the stability of deterrence, including strategic stability, the risk of crisis instability and wartime escalation. It focuses on these questions through the lens of two characteristics of LAWS: the potential for increased operational speed and the potential for decreased human control over battlefield choices. It also examines how these issues interact with the large uncertainty parameter associated with potential AI-based military capabilities at present, both in terms of the range of the possible and the opacity of their programming.

KEYWORDS Robotics; arms races; automation; warfare; deterrence; crisis stability

Introduction

The Cuban Missile Crisis is generally thought to be the closest the world came to a global nuclear war during the Cold War. As US President Kennedy and Soviet Premier Khrushchev communicated to try to resolve the standoff in October 1962, the world, and the navies of both sides, waited. The US Navy in particular faced a key dilemma, after being ordered by President Kennedy to establish a naval blockade and prevent further Soviet offensive military deployments. If Soviet naval ships 'ran' the blockade and attempted to supply Cuba with additional weapons, the US Navy would be obligated, based on its normal operating procedures, to respond with force, triggering a war. Moreover, while the US Navy wanted to implement a blockade according to naval regulations, placing the blockade many miles off the Cuban coast, President Kennedy ordered it drawn in closer to Cuba, to give

CONTACT Michael C. Horowitz horom@sas.upenn.edu Political Science, University of Pennsylvania Philadelphia, PA, USA

the Soviet Union more time to 'blink' and decide to not run the blockade. Meanwhile, the Soviet Navy eased closer to Cuba, determined to force the United States to back down and let Soviet naval vessels past even if they possessed offensive weapons.

The US Navy blockade was designed to persuade the Soviet Union that the United States was willing to escalate the conflict if the Soviet Navy did not back off. Meanwhile, the Kennedy administration sought pinpoint control of the US Navy, to ensure they did not escalate unnecessarily. The same was true on the Soviet side. They wanted control of their forces.

This story provides an opportunity to think about how one particular application of artificial intelligence (AI) to militaries, lethal autonomous weapon systems (LAWS), could influence strategic stability and decision-making. Imagine the Cuban Missile Crisis with machine-autonomous US naval ships, able to operate independently after activation. The Kennedy Administration could have ordered the ships into the exact blockade configuration it wanted, and then given them orders to fire on any Soviet naval vessels that attempted to run the blockade.

Deterrence theory might suggest that this would be effective. Deploying-autonomous naval vessels programmed to fire on Soviet naval vessels that attempted to run the blockade would be similar to throwing out the steering wheel in a game of chicken. If the Soviet Union knew that the naval vessels were pre-programmed and could not be controlled any longer by the Kennedy Administration, it would be an extremely credible commitment, which perhaps could have persuaded the Soviet Union to back off earlier.

Closer examination, however, reveals that this episode would not have been so simple. The first problem is uncertainty. Even if the US Navy's vessels were fully autonomous and not reprogrammable, how would the Kennedy Administration have persuaded the Soviet Union that this was the case? Simply telling them, given the lack of trust between the two sides, is unlikely to have worked. And it is not as if the US Navy would have allowed Soviet officials to see the programming code. Therefore, the 'commitment mechanism' advantage of autonomy is lower than it would appear on first glance.

Thus, the reason to deploy autonomous systems would have to be their reliability and effectiveness rather than signalling. And giving up human control to algorithms in a crisis that could end with global nuclear war would require an extremely high level of perceived reliability and effectiveness. Few things are more important to militaries in crisis situations than informational awareness and control over decisions, and there might be fear that autonomous systems are prone to accidents.

This counterfactual illustrates that the development and deployment of LAWS by national militaries, if it occurs, is unlikely to have simple and linear consequences. Instead, human factors, including the psychological desire for control and organisational politics, will strongly shape how militaries think about developing and using LAWS. This will not only influence the potential for arms competition in peacetime, but also deterrence and wartime stability, due to the organisational processes militaries implement for the deployment and use of autonomous systems on the battlefield.

This article draws on research in strategic studies and examples from military history to assess the issues above, as well as the potential for arms control.[1] It focuses on these questions through the lens of key characteristics of LAWS, especially the potential for increased operational speed and, simultaneously, less human control over some battlefield choices. One of the primary attractions of autonomous systems, even compared to remotely piloted systems, is the potential to operate at machine speed. Another potential benefit is the possibility of machine-like accuracy in following programming, but that comes with a potential downside: the loss of control and the accompanying risk of accidents, adversarial spoofing and miscalculation. Even if LAWS malfunction at the same rate as humans in a given scenario, the ability of operators to control the impact of those malfunctions may be lower; this could make LAWS less predictable on the battlefield. The article then examines how these issues interact with the large uncertainty parameter associated with AI-based military capabilities at present, both in terms of the range of the possible and the opacity of their programming.

The results highlight several critical issues surrounding the development and deployment of LAWS.[2] First, the desire to fight at machine speed with autonomous systems, while making a military more effective in a conflict, could increase crisis instability. As countries fear losing conflicts faster, it could generate escalation pressure, including an increased incentive for first strikes. Second, the fear of accidents and losing control of autonomous systems could limit the willingness of militaries to deploy them at times due to a lack of trust in their effectiveness. Third, the dual-use, or even general purpose, character of the basic science underlying many autonomous systems will make the technology relatively hard to control, though whether this is described as diffusion, proliferation or an arms race will depend on political dynamics as much as anything.

Finally, multiple uncertainty parameters concerning LAWS could exacerbate security dilemmas. Uncertainty over the range of the possible concerning the programming of LAWS will increase fear of those systems in the near term, making restraint less likely for competitive reasons. Moreover, the inherent differences between remotely piloted systems and LAWS at the platform level come from software, not hardware, complicating efforts to

[1]Thus, LAWS influence every area identified in the Introduction to this issue. Todd S. Sechser, Neil Narang, and Caitlin Talmadge. 'Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War', *Journal of Strategic Studies* 42/6 (2019), 727-735.

[2]For another perspective, see Jürgen Altmann and Frank Sauer, 'Autonomous Weapon Systems and Strategic Stability', *Survival* 59/5 (2017), 117–142.

build trust. Even though LAWS represent a small subset of the potential ways militaries may employ AI in the coming decades, they are likely to remain a challenging topic to address.

What is autonomy or artificial intelligence?

AI is the use of computing power, in the form of algorithms, to conduct tasks that previously required human intelligence.[3] AI, in this context, is best thought of as an umbrella technology or enabler, like the combustion engine or electricity. Military applications of AI are potentially broad – from image recognition for surveillance to more efficient logistics to battle management.[4] These include both non-kinetic applications, including in the cyber realm, as well as kinetic applications.[5] One potential application of AI is through armed autonomous systems that could be deployed on the battlefield, or what are most popularly called LAWS. These differ from remotely piloted systems where a human, though at a distance, still operates a given vehicle or system.

What is a LAWS? While simple to describe on first glance, and easy to understand in the extreme – an armed robot with extremely broad programming making decisions about engaging in lethal warfare – drawing the line between a LAWS and other weapon systems is complex. In Directive 3000.09, published in 2012, the US Department of Defense defines an autonomous weapon as 'A weapon system that, once activated, can select and engage targets without further intervention by Willi a human operator'.[6] What it means to select and engage a target is not entirely clear, however. For example, homing munitions, which have existed since World War II, select and engage targets, according to a common sense understanding of the terms.[7]

Exactly what functions are autonomous also matters. A system could have automatic piloting, for example, that flies or drives a platform to a

[3]Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. (Englewood Cliffs, Prentice Hall 2009); Calum McClelland, 'The Difference between Artificial Intelligence, Machine Learning, and Deep Learning', *Medium.com*, 4 December 2017, https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991; Nils J. Nilsson, The Quest for Artificial Intelligence: A History of Ideas and Achievements (Cambridge, U.K.: Cambridge University Press 2010); Shane Legg and Marcus Hutter, 'A Collection of Definitions of Intelligence,' Preprint, submitted 25 June 2007, https://arxiv.org/pdf/0706.3639.pdf; Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction, Vol. 1 (Cambridge: MIT Press 1998).

[4]Michael C. Horowitz, 'Artificial Intelligence, International Competition, and the Balance of Power', *Texas National Security Review*, 3/1 (2018), 37–57.

[5]The escalation logic for cyber may also differ from other systems. On emerging technologies and escalation, see Caitlin Talmadge, 'Emerging Technology and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today', *Journal of Strategic Studies* 42/6 (2019), 864-887.

[6]Department of Defense, 'Directive on Autonomy in Weapons Systems, Number 3000.09 , 2012.

[7]Paul Scharre and Michael C. Horowitz, 'An Introduction to Autonomy in Weapon Systems', *CNAS Working Paper*, February 2015, http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v021002.pdf.

target, but still have complete human control over the use of the weapon. That would be a system with a high level of automation, though not a LAWS according to most perspectives. Heather Roff measures the level of autonomy in a weapon system based on three subcomponents: self-mobility, self-direction and self-determination. This helps distinguish systems where there might be autonomy concerning the best way a missile should get to a target, but the target itself is designated by a person from systems where an algorithm might be making higher-level engagement decisions.[8] There are already some applications of limited machine autonomy in military systems, with the most prominent example being the automatic mode present on many Close-In Weapon Systems (CIWS), such as the Phalanx, used to defend ships and incoming missiles from attack.[9]

This article will not resolve the definitional debate surrounding LAWS, which is still ongoing in meetings of the Group of Governmental Experts focused on LAWS in the United Nations Convention on Certain Conventional Weapons. Provisionally, this article adopts the Scharre and Horowitz definition that a LAWS is '[A] weapon system that, once activated, is intended to select and engage targets where a human has not decided those specific targets are to be engaged'.[10] However, moving beyond the close cases (e.g. particular types of missile guidance systems) and considering those weapon systems that clearly use machine intelligence to search for, select, and/or engage targets can help clarify what is at stake in this debate in the first place.[11]

After all, if most militaries most of the time would not have any need for LAWS, or those systems have significant disadvantages relative to remotely piloted military robotics or soldiers on the battlefield, the stakes are lower. In contrast, if the integration of machine intelligence with military systems could give countries or violent non-state actors a significant advantage in how they employ force, it becomes even more crucial to engage the topic.

It is important to note that this article does not address concerns about existential risk related to artificial general intelligence – the fear that a super-intelligence could decide to destroy the human race, either because it decides humans are malign or because humans programme it to achieve a goal it can only accomplish by destroying humans.[12] The existential risk issue associated

[8]Heather M. Roff, *The Forest for the Trees: Autonomous Weapons and 'Autonomy' in Weapons Systems* (Tempe, AZ: Arizona State University 2016).

[9]Rebecca Crootof, 'Autonomous Weapon Systems and the Limits of Analogy', *Harvard National Security Journal* 9 (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820727; Kenneth Anderson, Daniel Reisner, and Matthew C. Waxman, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems', *International Law Studies* 90 (2014), 386–411. Israel also deployed the Harpy, an anti-radar, loitering, munition that could lock on and attack targets based on their radar signature.

[10]Scharre and Horowitz, 16.

[11]Vincent Boulanin and Maaike Verbruggen, 'Mapping the Development of Autonomy in Weapon Systems', *Stockholm International Peace Research Institute*, November 2017: https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems.

[12]Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press 2014).

with AI is not necessarily closely coupled to military applications of AI. If a super-intelligent machine learning system has the ability to take over human society in the interest of a goal – any goal – whether autonomous systems at much smaller orders of magnitude already exist in military systems will likely be unimportant. The super-intelligent system would simply create what it needed.

Why invest in autonomous systems?

Militaries are already increasing their investments in remotely piloted robotic systems. From UAVs such as the MQ-9 Reaper (United States) to uninhabited surface vehicles (USVs) such as the Guardium (Israel) to uninhabited ground vehicles (UGV) such as Platform-M (Russia), militaries around the world are investing in remotely piloted platforms, some of which can carry weapons. In these systems, human control over the use of force is not fundamentally different from the use of force with inhabited systems. In some cases, such as the MQ-9 Reaper, the sensor system a drone pilot uses to launch a weapon might even be the same sensor system a pilot in the cockpit of an inhabited fighter uses. Using remotely piloted systems gives militaries the ability to reduce the risk to their own soldiers while still projecting power in similar ways to how they used force previously.[13] The first places militaries are likely to use kinetic LAWS include relatively 'clear' environments such as air-to-air combat or naval combat, especially in geographic arenas where civilians are extremely unlikely to be present.[14]

Adding autonomy to kinetic systems potentially offers several advantages. Some of these benefits, such as the ability to take soldiers out of the line of fire and protect their lives, resemble remotely piloted robotics. Other features of LAWS would be unique. Most important, autonomous systems powered by machine learning will have a speed-based edge in decision-making and reaction times. AI could essentially enable faster operations by autonomous systems relative to remotely piloted or inhabited systems. An autonomous plane might be more adept at identifying and avoiding air defence threats, for example, or better at predicting and defeating adversaries in an air-to-air dogfight, making it more able to complete its mission. The necessity of fighting at machine speed, as US Deputy Secretary of Defense Robert Work says, is likely to play an important role in driving increasing automation in military systems, even aside from LAWS specifically.[15]

---

[13] Michael C. Horowitz, Sarah Kreps, and Matthew Fuhrmann, 'The Consequences of Drone Proliferation: Separating Fact from Fiction', *International Security* 41:2 (2016), 7–24.

[14] The word kinetic here is deliberate.

[15] Jim Garamone, 'Work Details Multidomain Battlefield of the Future', *U.S. Department of Defense*, 4 October 2016, http://www.defense.gov/News/Article/Article/963806/work-details-multidomain-battlefield-of-the-future.

Second, autonomous systems could be more accurate than humans in following their programming, helping overcome parts of the way fatigue, confusion and the limitations of the human mind can lead to mistakes on the battlefield. For example, an autonomous sniper would not get tired or hesitate in ways that a human might after hours or days waiting in position. This might also relate to planning in tactical situations. Kareem Ayoub and Kenneth Payne argue that ‘A modular AI that can optimise some tactical activity – say, the storming of an enemy position, rapidly coordinating fires and manoeuvre via networked and automated platforms – would outperform a seasoned battalion commander with ease’.[16] The more limited and discrete the tasks, e.g. firing a rifle the same way every time such that a bullet arrives at the same spot every time, the easier for a computer system to outperform a human.[17] The larger the data limitations and complexity of the task, of course, the harder it will be for military applications of AI to be more effective than humans.

Third, autonomous systems could further reduce the labour demands on militaries. While remotely piloted systems take soldiers out of harm’s way, the need for pilots for those remotely piloted platforms, as well as to process the data those platforms gather, does not necessarily reduce the labour demands on modern national militaries. Whether pilots are at a faraway military base or in the cockpit of an inhabited fighter plane, the military still needs to recruit and train those pilots. The larger the number of tasks that an autonomous system can do, however, the greater the potential for AI to reduce the size of the military, thus generating lower human costs. This would likely start in logistical areas and other places where soldiers are already less likely to be at risk, but if the technology advances, it could allow militaries to pull more soldiers out of harm’s way, and reduce the size of their militaries overall.[18]

Autonomous systems could also have practical downsides from a purely military effectiveness perspective (there are a variety of ethical and moral issues associated with autonomous weapons that this article brackets but which are important and worthy of discussion).[19] Normal accidents theory

[16]Kareem Ayoub and Kenneth Payne, ‘Strategy in the Age of Artificial Intelligence’, *The Journal of Strategic Studies* 39/5–6 (2015), 793–819.

[17]For an example of narrow AI, see DeepMind’s AlphaGo. David Silver et al., ‘Mastering the game of GO without human knowledge’, *Nature* 550 (2017): https://www.nature.com/articles/nature24270.

[18]On future battlefields, see John R. Allen, and Amir Husain, ‘On Hyperwar’, *Proceedings of the United States Naval Institute* 143:7 (2017), https://www.usni.org/magazines/proceedings/2017-2007/hyperwar.

[19]Due to space limitations, ethical and moral considerations associated with LAWS are beyond the scope of this article. For a discussion, see, among other articles: Peter Asaro, ‘On banning lethal autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making’, International Review of the Red Cross 94, 886 (2012), 687–709; Heather M Roff, “Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons ” Arizona State University Global Security Initiative Briefing Paper 2016, https://globalsecurity.asu.edu/sites/default/files/files/Control-or-Judgment-Understanding-the-Scope.pdf; United Nations Institute for Disarmament Research. ‘The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values’ No. 3 (2015), http://www.unidir.org/files/publications/pdfs/considering-ethics-and-social-values-en-624.pdf; Michael C. Horowitz, ‘The Ethics and Morality of Robotic Warfare: Assessing The Debate Over Autonomous Weapons’, *Daedalus* 145, 4 (2016), 25–36.

suggests that as system complexity increases, the risk of accidents increases as well and that some level of accidents are inevitable in complex systems.[20]

Autonomous weapon systems are quite complex due the degree of code required to generate algorithms for the use of force given present knowledge and technology.[21] While non-LAWS are also subject to accidents, this raises the question of whether LAWS may be more prone to one or more types of operational risk. LAWS placed in environments beyond the limitations of their programming could malfunction and fail in ways that are dangerous, particularly if they are armed. Even if commanders understand how the systems are supposed to work, and deploy them appropriately, the complexity of the programming raises the prospect of unintended behaviour and accidents.[22] Accidents could become more likely because adversaries will respond to LAWS by engaging in substantial hacking and spoofing efforts designed to prevent them from functioning as intended.[23] Finally, when autonomous systems interact on the battlefield, it might further raise the degree of uncertainty about their behaviour and the risk of accidents.

The potential consequences of these weapons have triggered significant multilateral dialogue in international institutions about the potential for LAWS arms control. Currently, through a Group of Government Experts at the Convention on Certain Conventional Weapons, the United Nations, is discussing the possibility of regulating LAWS.[24] After several years of dialogue, beginning in 2014, state parties are divided. A small group of states favours a complete ban on LAWS. Others, such as France and Germany, are uncertain, and still others, such as the United States and Russia, seem reticent to foreswear capabilities that they view as potentially important for the future conventional battlefield.[25] China's role in these discussions has been particularly interesting – they have signalled their support for a ban in international discussions, but are also heavily investing in military applications of AI.[26] These dialogues have improved mutual understanding of LAWS, and general agreement on the importance of human control in the use of force, but have not resolved complicated definitional issues over what exactly constitutes a LAWS.

[20]Charles Perrow, *Normal Accidents: Living with High Risk Systems* (New York: Basic Books 1984).
[21]Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W.W. Norton & Company 2018); Stephanie Carvin, *Normal Autonomous Accidents* (Norman Paterson School of International Affairs: Carleton University 2018).
[22]Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané, 'Concrete Problems in Ai Safety', *arXiv* (2016): https://arxiv.org/abs/1606.06565.
[23]Scharre. *Army of None*; Carvin, 'Normal Autonomous Accidents'.
[24]See the official history of the debate at: https://www.unog.ch/80256EE600585943/(httpPages)/8FA3C2562A60FF81C1257CE600393DF6?OpenDocument.
[25]'Autonomous weapons and the new laws of war', *The Economist*. 17 January 2019: https://www.economist.com/briefing/2019/01/17/autonomous-weapons-and-the-new-laws-of-war.
[26]Elsa Kania, 'China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapon Systems', *Lawfare*, 17 April 2018, https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems.

The arms control dilemma, in this case, is that the more the possession of LAWS improves the ability of a military to fight and win the nation's wars, the harder it will become for the international community to effectively regulate them. That does not make regulation impossible, to be clear. But the dual-use character of AI means some types of LAWS could be accessible to many militaries, not only major powers, and a broader set of states would have something to lose, from a capabilities perspective, through regulation. It also gives a broader number of states potential interest in regulation.[27]

Uncertainty surrounding the definition of a LAWS, though not a central focus of this article, could also make traditional arms control more difficult. From nuclear treaties like the Limited Test Ban Treaty to the Ottawa Convention, successful arms control agreements have generally tackled discrete technologies. The breadth of the category of AI and difficulties in defining what constitutes a LAWS at the margins are making reaching agreement on a definition of LAWS challenging at the international level.

One possibility is that, rather than the creation of a formal institution, norms on use could develop. Some norms already seem to be developing in the use of drones – for example, as the Turkey/Russia encounter in 2016 shows, countries are already treating the shooting down of a drone as less serious than the shooting down of an inhabited aircraft. The United States also attempted to promote norms of responsible behaviour with regards to drone use with the October 2016 rollout of the Joint Declaration for the Export and Subsequent Use of Armed or Strike-enabled UAVs.[28] The slow introduction of some types of LAWS could allow for norms to develop in a way that encourage safe and responsible use. But, what if technology advances in a way that leads to the rapid introduction of LAWS? The current GGE on LAWS is probably not enough to generate norms on their own.

Regardless, these ongoing arms control discussions make it all the more important to develop a firm analytical grasp on how LAWS may shape the security environment. For if fears of arms races, crisis instability and other negative consequences are unwarranted, the impetus for regulation would decline. Alternatively, the more plausible the consequences of LAWS, the more important international discussions become.

Impact on the security environment

Given these potential incentives to use, or not use, autonomous weapons systems, how might their development shape the international security

[27]On dual use technology in another arena, see Tristan Volpe, 'Dual-Use Distinguishability: How 3D-Printing Shapes the Security Dilemma for Nuclear Programs', *Journal of Strategic Studies* 42/6 (2019), 814-840.

[28]U.S. Department of State, 'Joint Declaration for the Export and Subsequent use of Armed or Strike-Enabled UAVs', 7 October 2016: https://pl.usembassy.gov/joint/(Accessed 10 March 2018).

environment? While much remains unpredictable by drawing on historical examples, specifying particular types of LAWS, and evaluating areas where militaries may invest in autonomy, investigating potential consequences becomes more plausible. This section explores the development of LAWS, including the potential for arms competition, the deployment of LAWS, and the way their development and deployment could shape deterrence and crisis stability.

There are other ways that LAWS could influence conflict as well, of course, including war initiation and how increasingly automated systems shape how states view war costs, but those are mostly out of the scope of this article.

Autonomous Weapon System Deployment

The history of cruise and ballistic missile development – and specifically the desire for faster and more accurate missiles – illustrates why countries might seek to accelerate the development and deployment of LAWS. In the Cold War, for example, increased missile accuracy was mostly viewed as a good in itself, or a natural development.[29] The promise of being able to operate faster than adversaries on the battlefield, or get inside their decision loops, are also constant goals of planners, as is the ability to use force more precisely. In combination, at the outset, these autonomous characteristics of weapon systems should make their development and deployment more likely.

A key factor that will influence the development and especially deployment of LAWS is trust in the system. As referenced above, the automatic mode on the Phalanx and related CIWS systems provides an approximation of a LAWS. Moreover, Iron Dome (Israel) and the PAC-3 Patriot (US and others) integrate autonomy into targeting and firing decisions. A track record of effectiveness can lead operators to trust them too much through a process called automation bias, as they outsource judgment to algorithms.[30] For example, in 2003, use of the Patriot missile system led to two fratricides in the US-led invasion of Iraq. In both cases, the autonomous characteristics of the system proved too brittle for the real ambiguities of combat, and operators trusted the sensors too much.[31]

[29]Donald A. MacKenzie, *Inventing accuracy: a historical sociology of nuclear missile guidance* (Cambridge, MA: MIT Press 1993). On missile accuracy, see Keir A. Lieber, and Daryl G. Press, 'The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence', *International Security* 41/4 (2017), 9–49.

[30]Mary L. Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems', *AIAA 1st Intelligent Systems Technical Conference*, 2004, 557–562.

[31]John K. Hawley, 'Patriot Wars, Automation and the Patriot Air and Missile Defense System', *Center for a New American Security*, 25 January 2017: https://www.cnas.org/publications/reports/patriot-wars.

For systems without proven track records, the combination of the military desire for control and concern about accidents might drive limitations in the deployment of more advanced LAWS, especially in more complex battle-spaces. Given that the extensive deployment of LAWS would inherently change the character of human control on the battlefield, militaries will want to do extensive testing, potentially even more than they do for other weapon systems, to ensure they will operate as programmed in potential usage scenarios.[32] While the Patriot example above suggests automation bias that lead to 'too much' trust in automation, the opposite can occur for less proven systems, as some research on remotely piloted systems suggests. For example, Schneider and McDonald note that ground controllers tend to trust manned aircraft more than unmanned systems even controlling for the capabilities of the aircraft. They write that 'What the JTACs [Joint terminal attack controllers] wanted was a "warm fuzzy." They wanted to interact with a pilot that had skin in the game because they were flying close enough to the ground to put them in danger'.[33] Military leaders may trust LAWS less than their remotely piloted counterparts, and certainly less than inhabited systems.

Moreover, just as with non-LAWS, militaries will be wary of deploying LAWS they believe adversaries can easily counter or that are likely to malfunction. The potential for accidents on the part of LAWS, particularly if others deploy them, since LAWS might be particularly prone to accidents when interacting with other systems, could lead to mutual restraint in some cases. Each side might not deploy a LAWS if it believes the other side might deploy them, and that their system will be unpredictable if the other sides deploys its LAWS. This is related to the way the acquisition of chemical weapons by both sides led to stalemates in World War I and helped discourage battlefield use in World War II in the European theater.[34] Both sides had chemical weapons capacity but did not use them because they knew the other side could respond in kind, and that it would lead to results that were, at best, a stalemate and, at worst, unpredictable (due to the way wind and other factors can rapidly shift the effect of a chemical weapons attack).

The very unpredictability of LAWS may make them a modern-day version of Schelling's 'threat that leaves something to chance'.[35] One could therefore imagine scenarios where the threat to deploy a LAWS, precisely because of uncertainty over how it would function, might influence how actors consider

[32]There is the possibility that fear of the development of LAWS by other states could lead states to actually do less testing and analysis than necessary before deployment autonomous weapons. This is discussed below in the section on arms races.

[33]Julia Macdonald and Jacquelyn Schneider, 'Battlefield Responses to New Technologies: Views from the Ground on Unmanned Aircraft', *Security Studies* 28/2 (2019), 216-249.

[34]Jeffrey W. Legro, *Cooperation under Fire: Anglo-German Restraint During World War II* (Ithaca, NY: Cornell University Press 1995).

[35]Thomas C. Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press 1960).

whether and how to fight. The potential target of an LAWS attack would have to weigh the probability they think the weapon will function as intended and the impact of the weapon if it does function as intended. This also presumes, of course, that if the weapon functions as intended that it would give the user an edge (since there would be almost no reason to deploy an LAWS if it did not give a military an edge). In that scenario, the user is threatening the target with a weapon that could potentially make a difference on the battlefield, but might not function as advertised.

This potential logic of LAWS deployment restraint would not apply, however, in several situations. Groups such as the Islamic State, countries such as North Korea, or even countries such as Russia could think about the use of force in ways that lessen restaints on using LAWS. Militant groups could try to build simple LAWS from commercial machine learning algorithms. The way the Islamic State has modified consumer drones to launch grenades or serve as guided weapons illustrates the possibilities.[36] A higher degree of unpredictability for these systems also might not be as much of a concern for a group such as the Islamic State, since an outcome where a system malfunctions and it ends up killing civilians instead of adversaries would be less concerning. Similarly, North Korea may be less worried about malfunctions that kill civilians than most states, given that the regime is already one of the worst human rights violators in the world.

Finally, given its conventional military inferiority in comparison with the United States, Putin's Russia may also be willing to take more risk when it comes to LAWS. Russia's military has concerns about a military effectiveness gap relative to the United States driven by both operational capacity and technology. This gap creates an incentive for Russia to automate and helps explain why Russia is investing heavily in military robotics and autonomous systems, though their actual progress is less clear.[37] These actors would still have strong interests in their LAWS functioning effectively, though. They would not want to deploy systems that might fail in ways that killed their own forces unnecessarily, for example.

Another situation in which actors might show less restraint in using LAWS even if they are less reliable than non-autonomous weapons, is rarer – all-out war. A country that believes they are at risk of losing a war might become more risk acceptant. Militaries might therefore deploy a LAWS that, if it functioned properly, might help a country win an important battle even if that system also had a higher failure risk than other potential

[36]Ben Watson, 'The Drones of ISIS', *Defense One*, 12 January 2017, https://www.defenseone.com/technology/2017/01/drones-isis/134542/.

[37]Alina Polyakova, 'Weapons of the Weak: Russia and AI-driven asymmetric warfare', *Brookings Institution*, 15 November 2018, https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/; Samuel Bendett, 'Here's How the Russian Military Is Organizing to Develop AI', *Defense One*, 20 July 2018, https://www.defenseone.com/ideas/2018/07/russian-militarys-ai-development-roadmap/149900/.

systems. Given the prior expectation of a loss, there would be more upside to using a less predictable weapon.

More generally, the interaction between the potential for accidents and LAWS deployment will also depend on how the weapons are programmed to behave if they fail. A weapon that failed safely by shutting off would be more attractive than a weapon that failed by firing wildly at friend and foe. This could influence the design of LAWS in the first place, since responsible militaries will be worried about malfunctioning weapons endangering their own troops, getting back to the trust issue raised above.

Arms Races

A persistent concern about the development of LAWS is that their debut in the international system could trigger arms races. A letter signed by thousands of scientists in 2015 stated that 'If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow'.[38]

Unpacking this issue requires understanding what constitutes an arms race. As Huntington wrote, '[A]n arms race is defined as a progressive, competitive peacetime increase in armaments by two states or coalition of states resulting from conflicting purposes or mutual fears. An arms race is thus a form of reciprocal interaction between two states or coalitions'.[39] How valid is the fear of an arms race in the application of machine learning to military systems? This fear, after all, is one key motivator behind the ongoing GGE talks on LAWS in Geneva.[40]

Arms races or proliferation?

It is important to distinguish arms races from the proliferation of military technologies. The scientists in the example above compared an LAWS arms race to the spread of Kalashnikovs, but there was never an arms race in Kalashnikovs – they just spread because they were cheap, easy to produce and useful.[41] Huntington's classic work on arms races distinguishes between proliferation and arms races. Countries can increase their arms acquisitions

[38]Future of Life Institute, 'Autonomous Weapons: An Open Letter from Ai & Robotics Researchers', 28 July 2015, https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1. Also see Ben Garfinkel and Allan Dafoe. 'How Does the Offense-Defense Balance Scale', *Journal of Strategic Studies* 42/6 (2019), 736-763.

[39]Samuel P. Huntington, 'Arms Races: Prerequisites and Results', *Public Policy* 8 (1958), 41–86. Also see Lewis F. Richardson, *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War* (New York: Boxwood Press 1960).

[40]Also see Altmann and Sauer, 'Autonomous Weapon Systems and Strategic Stability'.

[41]Future of Life Institute, 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers'.

due to an 'absolute need' that exists 'regardless of the actions of other states', or for economic reasons.[42]

Huntington argues that many things described as arms races are simply general build-ups due to military necessity (or for economic reasons), rather than a specific build-up due to a particular disagreement between states.[43] For example, after their debut in World War I, countries around the world acquired tanks in the 1920s and 1930s. Tanks then became a critical part of ground warfare in World War II and subsequent conventional ground combat operations. Yet, few would call the spread of tanks in the 1920s and 1930s an arms race. It was simply proliferation that was not possible to stop, as large-calibre guns, caterpillar tracks and the combustion engine were available to countries around the world.[44]

To the extent that those concerned about an arms race in autonomous weapon technologies are actually concerned with proliferation, some degree of proliferation may be inevitable simply due to the underlying factors involved in the production of LAWS. As Horowitz argues, military capabilities diffuse faster when, on the technology acquisition side, there is underlying commercial demand and the unit costs are low.[45] Commercial markets are already driving the integration of AI into several areas of the United States and global economies, through deep learning and machine learning applications.[46] From Google search to Macy's shopping assistance for customers, AI is increasingly embedded in commercial sectors of modern society.[47]

Narrow applications of AI could become increasingly integrated into most economic sectors, with many algorithms, once developed, now available to many actors.[48] Export controls are unlikely to stop basic narrow AI capabilities from spreading, even if more advanced applications are beyond the capacity of most companies and governments; government regulations generally lag emerging technologies. Finally, AI innovation is occurring around the world, not just in the United States or even the West.

Machine learning capabilities designed for commercial purposes could also have spillovers with useful applications to the military realm. This would

[42] Huntington, 'Arms Races: Prerequisites and Results', 41–42.
[43] Ibid., 42.
[44] Barry R. Posen, *The Sources of Military Doctrine: France, Britain, and Germany between the World Wars* (Ithaca: Cornell University Press 1984).
[45] Michael C. Horowitz, *The Diffusion of Military Power: Causes and Consequences for International Politics* (Princeton, NJ: Princeton University Press 2010).
[46] 'March of the Machines: What History Tells Us About the Future of Artificial Intelligence – and How Society Should Respond', *Economist*, June 25 2016, http://www.economist.com/news/leaders/21701119-what-history-tells-us-about-future-artificial-intelligenceand-how-society-should.
[47] Jacques Bughin et al., 'Artificial Intelligence: The Next Digital Frontier', *McKinsey Global Institute Discussion Paper*, June 2017, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.
[48] Tim Hwang, 'Computational Power and the Social Impact of Artificial Intelligence', 23 March 2018, http://dx.doi.org/3147910.3142139/ssrn.3147971; Martin Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future* (New York: Basic Books 2015).

reverse the Cold War dynamic in the West, where US civilian economic innovations such as GPS often spun out of military development programmes.[49] AI, as described above, is more an enabler such as the combustion engine than a weapon. It is therefore different than a platform like stealth technology, which really only has military purposes.

Less wealthy and powerful countries, as well as militant groups, could exploit commercial applications of machine learning to build simple LAWS. This is part of the concern driving the GGE process to regulation autonomous weapons in Geneva. Some are already plausible today. A determined militant group or nation-state could mount a machine gun on a tracked vehicle, connect a heat sensor and write software code such that the machine gun will fire at anything that has the heat signature of a human being. While such a system would be indiscriminate and violate the Law of War in nearly all use cases, it is buildable today.

Moreover, if machine learning can aid tasks such as connecting weapons to tracking devices and target identification, the resulting automation would ease the challenges of system integration, or the bringing together of component subsystems into a single, functioning operation. Consequently, it could be possible to modify commercial AI to aid in military operational planning, making it easier for many non-state actors as well as nation-states to conduct ever-more complex operations. In combination, proliferation and the risk of broad access to some types of LAWS motivates many that are apprehensive about LAWS in the international community, which is part of why the CCW advanced the initial LAWS discussions into a GGE.

On the other hand, extensive diffusion of algorithms with relevance for militaries may not occur.[50] The proliferation case made above depends on the portability of machine learning algorithms from the civilian to the military realm, or large nascent investment by many actors. Investments in the commercial AI sector far outpace those in the military sector,[51] and due to the trust issues raised above, some militaries may actually hesitate to invest heavily in AI. The more differences there are, in any given case, the more reprogramming will be necessary, which could increase barriers to entry.[52] It is also possible that simpler LAWS will proliferate quickly, while advanced systems, especially those that contain any type of planning function, only have military applications, or require especially high degrees of compute,

[49]Stephen G. Brooks, *Producing Security: Multinational Corporations, Globalization, and the Changing Calculus of Conflict* (Princeton, NJ: Princeton University Press 2005).

[50]For example, Ayoub and Payne (17–18) argue that the barriers to entry for military applications for AI could be higher than anticipated.

[51]Horowitz, 'Artificial Intelligence, International Competition, and the Balance of Power', The distinguishability of military and civilian investments also matters.

[52]Ibid.

may only be accessible to large states with significant information technology infrastructures.[53]

Autonomous arms race dynamics

If many actors may have the ability to produce, adapt or import some types of LAWS, will that lead to arms races, as opposed to proliferation? Note that while talking about AI developments in a macro sense as an arms race is conceptually flawed, because of the numerous potential applications of very different types of AI across wide military and civilian spectrums, it is theoretically possible that arms races could occur across particular dimensions of AI connected to specific military missions. Specific examples of LAWS could, in theory, be such a dimension.

Arms races do happen, after all, even given the definitional caveats raised above. During the Cold War, for example, the United States and Soviet Union engaged in an arms race over nuclear weapons, with each side seeking to build more sophisticated nuclear weapons and delivery systems to gain an edge.[54] Russian President Vladimir Putin infamously discussed leadership in AI in explicitly competitive terms in 2017 when he stated that AI '[C]omes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world'.[55]

All arms races share an underlying political dynamic whereby fear of developments by one or multiple other actors, and the inability to verify that those actors are not developing particular capabilities, fuels more intense development of new weapon systems than would happen otherwise.[56] An arms race in the area of machine autonomy would be no different in that dimension. The root would be inherently political.[57] Actors would also have to believe that they would gain an advantage from the developing LAWS, or least be at a significant disadvantage if they did not develop those weapon systems.

Jervis argues that arms races occur due to a security dilemma when states have the ability to measure each other's capabilities, but not their intentions.[58] The opacity surrounding LAWS development might generate increased risk for arms competition because of potential opacity about capabilities, in addition to the 'normal' opacity that exists about intentions. First, it will be extremely difficult for states to credibly demonstrate

[53]Ibid. Hwang, 'Computational Power and the Social Impact of Artificial Intelligence'.
[54]Matthew Evangelista, *Innovation and the Arms Race: How the United States and the Soviet Union Develop New Military Technologies* (Ithaca: Cornell University Press 1988).
[55]James Vincent, 'Putin Says the Nation That Leads in AI "Will Be the Ruler of the World"', *The Verge*, 4 September 2017, https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world.
[56]Colin S. Gray, *House of Cards: Why Arms Control Must Fail* (Ithaca: Cornell University Press 1992).
[57]Assume that it is humans making the decision about whether to build LAWS.
[58]Robert Jervis, 'Cooperation under the Security Dilemma', *World Politics*, 30/2 (1978): 186–214.

autonomous weapon capabilities. The difference between a remotely piloted system and an autonomous system is software, not hardware, meaning verification that a given country is operating an autonomous system at all would be difficult. Second, uncertainty about the technological trajectory of machine learning and specific military applications means that countries might have significant uncertainty about other countries' capabilities. Thus, countries might invest a lot in AI applications to military systems due to fear of what others are developing.

The heightened role of uncertainty about what other countries are developing would make an LAWS arms competition different than many historical arms races – for example, the Anglo-German naval arms race prior to World War I. In the Anglo-German naval arms race case, both sides could see the ships being produced by the other side because those ships left port for testing, and were subject to reporting by spies who could observe construction patterns.[59] Even though there was some uncertainty about the specific capabilities of battleships and battlecruisers, each side could count the number and size of the guns deployed on each ship.

Third, the rules of engagement for LAWS would also likely be unknown – and use of an LAWS by a state in one engagement might not generate predictability, since a state could change the programming of the system prior to the next engagement. Thus, opacity surrounding AI capabilities could, potentially, lead to worse case assumptions about capability development by potential adversaries, thus making arms race dynamics more likely.

Research on bargaining and war also suggests that uncertainty about capabilities makes it harder for countries to come to agreements when they enter into disputes. Private information about military capabilities means both sides can believe they are likely to win if a dispute escalates.[60] The dispute then becomes harder to resolve and more likely to escalate. To the extent that machine learning systems generate more uncertainty due to their opacity, an arms race over machine learning systems might therefore be somewhat more likely to escalate. The extent of the effect would be difficult to determine, however.

Another risk is that competitive dynamics mean countries will accelerate their weapons development cycles and deploy LAWS before fully testing them, due to a fear of falling behind. This would essentially resolve the trust dilemma in a way that makes accidents more likely.[61] Given concern that LAWS could be more prone to accidents, such a development would be especially dangerous. This risk of an LAWS arms race causing countries to

[59]On key moments in the Anglo-German arms race, see John H. Maurer, 'Churchill's naval holiday: Arms control and the Anglo-German naval race, 1912–1914, *The Journal of Strategic Studies* 15/1 (1992), 102–127.
[60]James D. Fearon, 'Rationalist Explanations for War', *International Organization* 49/3 (1995).
[61]Scharre, *Army of None*.

take short cuts in weapons development seems unlikely, however. Militaries want weapons they can control (excluding potential exceptions noted above), and they are unlikely to approve of deploying weapon systems they view as less able to accomplish a mission, or more likely to put their own forces in danger, than alternatives. Thus, the incentive to deploy effective systems will hedge at least somewhat against short-cuts in the weapons development process. Moreover, public awareness about the risks of AI could play a role in shaping how militaries consider deploying AI systems, even in a competitive scenario. Fear of AI should lead most militaries to be more careful, rather than less careful, in the testing and development of LAWS. The pressure to stay ahead will compete with the pressure to deploy effective systems.

Deterrence and Crisis Stability

How might the deployment of LAWS influence deterrence and the prospect for wartime escalation, including with nuclear-armed countries?

The relationship between speed and crisis stability in a world of deployed LAWS represents one of the clearest risk factors associated with autonomous weapons. The United States and the Soviet Union avoided nuclear war during the Cold War in part due to the development of mutually assured destruction, a situation where each side believed that, even if it struck first, the target would still have enough nuclear forces remaining to destroy the aggressor. The countries developed complicated and overlapping systems for command and control, as well as different types of nuclear strike systems.[62] Ballistic missiles, for example, represented the 'autonomous' weapons of their day, because they could not be recalled, which was unique at the time. There was also a trade-off between perceived attack capabilities and perceived strategic stability. Ballistic missiles with multiple independently targeted re-entry vehicles could allow countries to maximise damage in a first strike, but those very capabilities also made them disruptive to strategic stability.[63]

The Soviet Union also allegedly deployed an automated system called 'Perimeter', known as the Dead Hand system, in response to fears of decapitation. Evidence from former Soviet military and nuclear officials suggest that the Soviets designed the system to enable retaliation against a US nuclear first strike even if Soviet command and control was decapitated. Soviet leadership could active Perimeter in a crisis if they feared they might lose active control of their nuclear arsenal due to a US strike.[64]

[62]Robert Powell, 'Crisis Stability in the Nuclear Age', *American Political Science Review* 83/1 (1989), 61–76.
[63]Altmann and Sauer, 'Autonomous Weapon Systems and Strategic Stability'.
[64]Bruce G. Blair, 'Russia's Doomsday Machine,' *The New York Times*, 8 October 1993, http://www.globalzero.org/files/bb_russias_doomsday_machine_10.08.1998.pdf; and Nicholas Thompson, 'Inside the Apocalyptic Soviet Doomsday Machine,' *WIRED*, 1 September 2009, https://www.wired.com/2009/09/mf-deadhand/.

The speed associated with LAWS could potentially threaten first strike stability in a crisis. The ability to fight at machine speed means a state could win faster – but it also means that state could lose faster. Countries could fear that an aggressor, using LAWS or related systems operating at machine speed, could quickly knock out their command and control capabilities, eliminating their ability to retaliate (regardless of whether one or both sides has nuclear weapons). This fear would create incentives for many of the least stable military postures developed during the Cold War, including strategic weapons on high alert, launch on warning postures, and others. A country fearing it might not have the ability to respond in time if its command and control capabilities are devastated by machine-speed attack could also have incentives for pre-delegation. Autonomous weapon systems could therefore place pressure on escalation control mechanisms.

LAWS, if they prove effective battlefield weapons, could also threaten deterrence by undermining a nuclear deterrent itself.[65] Imagine, for example, undersea or above-ground swarms of autonomous systems with the ability to target ballistic missile submarines or ICBM silos. Some also fear a situation where undersea LAWS track adversary submarines. Fear that those tracking systems could undermine the sea-based deterrent of a country, especially a nuclear-armed country, could in theory create first strike incentives as well. These scenarios currently seem very unlikely for a variety of technical reasons, including power restrictions and communications challenges underwater.[66]

Now, there is nothing necessarily unique about the weapons being autonomous in this scenario – fast weapon systems that can threaten command and control systems can place pressure on strategic stability in general. For example, precisely because hypersonic missiles could hit over-the-horizon targets in a fraction of the time it would take existing ballistic or cruise missiles, many analysts believe they would undermine strategic stability.[67] For this situation to come about, autonomous tracking systems that could attack would have to be credible, already observed by the target, and something the target would not have the ability to defend themselves from. The uncertainty about survivability that LAWS would create in this situation could be mitigated by defensive systems, in theory.

[65]This is also raised by Altmann and Sauer, 'Autonomous Weapon Systems and Strategic Stability', 131.

[66]Other emerging technologies may also influence preemption incentives. In the cyber arena, see Jacquelyn Schneider, 'The Capability/Vulnerability Paradox and Military Revolutions: Implications for Computing, Cyber, and the Onset of War', *Journal of Strategic Studies* 42/6 (2019), 841–863. On the stability of sea-based deterrents, see Lieber and Press, 'The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence'; Ryan Snyder and Benoît Pelopidas, 'Correspondence: New Era or New Error? Technology and the Future of Deterrence', *International Security* 43 3 (Winter 2018/9), 190–193.

[67]Amy F. Woolf, 'Conventional Prompt Global Strike and Long-Rang Ballistic Missiles: Background and Issues', *Congressional Research Service*, 6 April 2018: https://fas.org/sgp/crs/nuke/R41464.pdf.

The high degree of uncertainty about LAWS, especially for first moving states, before actors in the international system gain a more complete understanding of the realm of the possible with LAWS, could also impact deterrence in two ways. First, as discussed above, there could be uncertainty over whether the weapon will function as programmed. That generates uncertainty an aggressor might try to use to coerce a target.

A second type of uncertainty could also create instability. Software will determine how an autonomous weapon will function, including the rules of engagement, and countries are not going to allow potential adversaries to read their software code. In crisis situations, there could be greater uncertainty that opponents lack experience in dealing with, relative to facing human opponents.

Consider the motivating example for this article, the Cuban Missile Crisis. Imagine that the US Navy ships were deployed in a picket line, but the Soviet Union did not know how the ships were programmed to respond. The ships might be programmed to allow Soviet ships running the blockade to proceed to Cuba, or they might fire on the Soviet ships. Moreover, given that humans would not be in the loop, there would be no chance for either frailty on the part of a ship commander to not use force, or over-aggressive behaviour. It is possible that this uncertainty would have made the Soviet Navy more likely to back down, though it is hard to say. But the level of uncertainty would potentially have been higher.

Of course, there is already uncertainty already concerning the rules of engagement for human-piloted systems during crisis situations. However, uncertainty about an autonomous system would not simply be about whether LAWS would follow commands, but about what the autonomous system was programmed to do. This is another reason why decision makers might therefore want an override switch. Given the challenges of signaling that a weapon is actually autonomous, the terminal impact of LAWS on deterrence could wash out. That being said, the discussion above illustrates risks that countries will have to manage.

Conclusion

This article assesses the growing integration of AI in military systems with an eye towards the impact on crisis stability, specifically how countries think about developing and deploying weapons, as well as when they are likely to go to war, and the potential for arms control.[68] Contrary to some public concern and media hype, unless AI capabilities reach truly science fiction levels, their impact

[68]Arms control surrounding emerging technologies can be possible, but is challenging. See Heather Williams, 'Asymmetric Arms Control and Strategic Stability: Scenarios for Limiting Hypersonic Glide Vehicles', *Journal of Strategic Studies* 42/6 (2019), 789-813.

on national and subnational military behaviour, especially interstate war, is likely to be relatively modest. Fundamentally, countries go to war for political reasons, and accidental wars have traditionally been more myth than reality.[69] The effects for subnational use of AI could be more significant, especially if military applications of AI make it easier for autocrats to use military force to repress their population with a reduced number of loyalists.

The commercial spread of machine learning in the private sector means some form of spillovers to military applications will be inevitable. The desire for faster decision-making, concern about the hacking of remotely piloted systems, and fear of what others may be developing could all incentivise the development of some types of LAWS. However, awareness of the potential risk of accidents regarding these systems, as well as the desire for militaries to maintain control over their weapons to maximise their effectiveness, will likely lead to caution in the development and deployment of systems where machine learning is used to select and engage targets with lethal force.

One of the greatest risks regarding applications of AI to military systems likely comes from opacity concerning those applications, especially as it interacts with the potential to fight at machine speed. Unlike missiles or bombers, it will be difficult for countries to verify what, if any, AI capabilities potential adversaries have. Even an international agreement restricting or prohibiting the development of LAWS would be unlikely to resolve this concern. Fear would still exist. Given that uncertainty makes disputes harder to resolve, this could have an impact.

These factors make international regulation potentially attractive, in theory, but challenging in application, because the very thing about LAWS that might make international regulations on LAWS attractive – their ability to enable faster and more devastating attacks, as well as the risk of accidents – may also make those regulations harder to implement and increase the risks if cheating occurs. But discussions at the CCW are ongoing, and may yet yield progress, or at the least agreement on considering safety and reliability issues when evaluating the development and use of autonomous systems. Humanity's worst fears about an intelligent machine turning against it aside, the integration of machine learning and military power will likely be critical area of inquiry for strategic studies in the years ahead.

Acknowledgments

The author would like to thank Neil Narang, Todd Sechser, Caitlin Talmadge, the anonymous referee, the editors, and workshop participants at UVA and Stanford for their insights. All errors are the responsibility of the author.

[69]Dan Reiter, 'Exploding the Powder Keg Myth: Preemptive Wars Almost Never Happen', *International Security* 20/2 (1995), 5–34.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This research was supported, in part, by a grant from Carnegie Corporation of New York.

Notes on contributor

Michael C. Horowitz is Professor of Political Science and the Associate Director of Perry World House at the University of Pennsylvania. His publications include the book *The Diffusion of Military Why Power Leaders Fight* (Princeton, NJ: Princeton University Press: 2010).

ORCID

Michael C. Horowitz http://orcid.org/0000-0002-3067-7157

Bibliography

Allen, John R. and Amir Husain, 'On Hyperwar', *Proceedings of the United States Naval Institute* 143 7 (2017), https://www.usni.org/magazines/proceedings/2017-2007/hyperwar

Altmann, Jürgen and Frank Sauer, 'Autonomous Weapon Systems and Strategic Stability', *Survival* 59/5 (2017), 117–42. doi:10.1080/00396338.2017.1375263

Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané, 'Concrete Problems in Ai Safety', *arXiv*, (2016): https://arxiv.org/abs/1606.06565.

Anderson, Kenneth, Daniel Reisner, and Matthew C. Waxman, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems', *International Law Studies* 90 (2014), 386–411

Asaro, Peter. 'On Banning Lethal Autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making', *International Review of the Red Cross*, 94/886 (2012), 687–709.

'Autonomous weapons and the new laws of war', *The Economist*, January 17, 2019. https://www.economist.com/briefing/2019/01/17/autonomous-weapons-and-the-new-laws-of-war

Ayoub, Kareem and Kenneth Payne, 'Strategy in the Age of Artificial Intelligence', *The Journal of Strategic Studies* 39/5-6 (2015), 793–819

Blair, Bruce G. 'Russia's Doomsday Machine', *The New York Times*, October 8, 1993, http://www.globalzero.org/files/bb_russias_doomsday_machine_10.08.1998.pdf

Bostrom, Nick, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press 2014)

Boulanin, Vincent and Maaike Verbruggen, 'Mapping the Development of Autonomy in Weapon Systems', *Stockholm International Peace Research Institute*, November

2017: https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems

Brooks, Stephen G., *Producing Security: Multinational Corporations, Globalization, and the Changing Calculus of Conflict* (Princeton, NJ: Princeton University Press 2005)

Bughin, Jacques et al., 'Artificial Intelligence: The Next Digital Frontier', *McKinsey Global Institute Discussion Paper*, June 2017, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

Carvin, Stephanie, *Normal Autonomous Accidents* (Norman Paterson School of International Affairs: Carleton University 2018)

Crootof, Rebecca, 'Autonomous Weapon Systems and the Limits of Analogy', *Harvard National Security Journal* 9 (2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820727

Cummings, Mary L., 'Automation Bias in Intelligent Time Critical Decision Support Systems', *AIAA 1st Intelligent Systems Technical Conference*, 2004: 557–62.

Department of Defense, 'Directive on Autonomy in Weapons Systems, Number 3000.09', (Department of Defense, 2012).

Evangelista, Matthew., *Innovation and the Arms Race: How the United States and the Soviet Union Develop New Military Technologies* (Ithaca: Cornell University Press 1988)

Fearon, James D., 'Rationalist Explanations for War', *International Organization* 49/3 (1995)

Ford, Martin., *Rise of the Robots: Technology and the Threat of a Jobless Future* (New York: Basic Books 2015)

Future of Life Institute, 'Autonomous Weapons: An Open Letter from Ai & Robotics Researchers', July 28, 2015: https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1.

Garamone, Jim, 'Work Details Multidomain Battlefield of the Future', *U.S. Department of Defense*, October 4, 2016, http://www.defense.gov/News/Article/Article/963806/work-details-multidomain-battlefield-of-the-future.

Garfinkel, Ben and Allan Dafoe. 'How Does the Offense-Defense Balance Scale', *Journal of Strategic Studies* 42/6 (2019), 736–763.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans, 'When Will AI Exceed Human Performance? Evidence from AI Experts', *Journal of Artificial Intelligence Research*, (Forthcoming): https://arxiv.org/abs/1705.08807.

Gray, Colin S., *House of Cards: Why Arms Control Must Fail* (Ithaca: Cornell University Press 1992)

Hawley, John K., 'Patriot Wars, Automation and the Patriot Air and Missile Defense System', *Center for a New American Security*, January 25, 2017: https://www.cnas.org/publications/reports/patriot-wars.

Horowitz, Michael C., *The Diffusion of Military Power: Causes and Consequences for International Politics* (Princeton, NJ: Princeton University Press 2010)

Horowitz, Michael C. 'The Ethics and Morality of Robotic Warfare: Assessing The Debate Over Autonomous Weapons', *Daedalus*, 145/4 (2016), 25–36.

Horowitz, Michael C., 'Artificial Intelligence, International Competition, and the Balance of Power', *Texas National Security Review* 3/1 (2018), 37–57

Horowitz, Michael C., Sarah Kreps, and Matthew Fuhrmann, 'The Consequences of Drone Proliferation: Separating Fact from Fiction', *International Security* 41 2: (2016), 7–24. doi:10.1162/ISEC_a_00257

Huntington, Samuel P., 'Arms Races: Prerequisites and Results', *Public Policy* 8 (1958), 41–86

Hwang, Tim, 'Computational Power and the Social Impact of Artificial Intelligence', March 23, 2018, http://dx.doi.org/3147910.3142139/ssrn.3147971

Jervis, Robert, 'Cooperation under the Security Dilemma', *World Politics* 30/2 (1978), 186–214

Kania, Elsa, 'China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapon Systems', *Lawfare*, April 17, 2018. https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems

Legg, Shane and Marcus Hutter, 'A Collection of Definitions of Intelligence,' Preprint, submitted June 25, 2007, https://arxiv.org/pdf/0706.3639.pdf

Legro, Jeffrey W., *Cooperation under Fire: Anglo-German Restraint during World War II* (Ithaca, NY: Cornell University Press 1995)

Lieber, Keir A. and Daryl G. Press, 'The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence', *International Security* 41/4 (2017), 9–49. doi:10.1162/ISEC_a_00273

Macdonald, Julia and Jacquelyn Schneider, 'Battlefield Responses to New Technologies: Views from the Ground on Unmanned Aircraft' *Security Studies*, 28/2 (2019), 216–249

MacKenzie, Donald A., *Inventing Accuracy: A Historical Sociology of Nuclear Missile Guidance* (Cambridge, MA: MIT Press 1993)

'March of the Machines: What History Tells Us about the Future of Artificial Intelligence—And How Society Should Respond', *Economist*, June 25, 2016, http://www.economist.com/news/leaders/21701119-what-history-tells-us-about-future-artificial-intelligenceand-how-society-should.

Maurer, John H., 'Churchill's Naval Holiday: Arms Control and the Anglo-German Naval Race, 1912–1914', *The Journal of Strategic Studies* 15/1 (1992), 102–27. doi:10.1080/01402399208437475

McClelland, Calum, 'The Difference between Artificial Intelligence, Machine Learning, and Deep Learning', *Medium.com*, December 4, 2017, https://medium.com/iotforall/the-difference-between-artificial-intelligence-machine-learning-and-deep-learning-3aa67bff5991.

Nilsson, Nils J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge, UK: Cambridge University Press 2010).

Perrow, Charles, *Normal Accidents: Living with High Risk Systems* (New York: Basic Books 1984)

Posen, Barry R., *The Sources of Military Doctrine: France, Britain, and Germany between the World Wars* (Ithaca: Cornell University Press 1984)

Powell, Robert, 'Crisis Stability in the Nuclear Age', *American Political Science Review* 83:1 (1989), 61–76. doi:10.2307/1956434

Reiter, Dan, 'Exploding the Powder Keg Myth: Preemptive Wars Almost Never Happen', *International Security* 20/2 (1995), 5–34. doi:10.2307/2539227

Richardson, Lewis F., *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War* (New York: Boxwood Press 1960)

Roff, Heather M., *The Forest for the Trees: Autonomous Weapons and "Autonomy" in Weapons Systems* (Tempe, AZ: Arizona State University 2016)

Roff, Heather M. 'Meaningful Human Control or Appropriate Human Judgment? The Necessary Limits on Autonomous Weapons.' Arizona State University Global

Security Initiative Briefing Paper 2016. https://globalsecurity.asu.edu/sites/default/files/files/Control-or-Judgment-Understanding-the-Scope.pdf

Russell, Stuart and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 3rd (Englewood Cliffs, NJ: Prentice Hall 2009)

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

Scharre, Paul and Michael C. Horowitz, 'An Introduction to Autonomy in Weapon Systems', *CNAS Working Paper*, February 2015, http://www.cnas.org/sites/default/files/publications-pdf/Ethical%20Autonomy%20Working%20Paper_021015_v021002.pdf.

Schelling, Thomas C., *The Strategy of Conflict* (Cambridge: Harvard University Press 1960)

Schneider, Jacquelyn, 'The Capability/Vulnerability Paradox and Military Revolutions: Implications for Computing, Cyber, and the Onset of War', *Journal of Strategic Studies* 42/6 (2019), 841–863.

Sechser, Todd S., Neil Narang, and Caitlin Talmadge. 'Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War', *Journal of Strategic Studies* 42/6 (2019), 727–735.

Silve, David et al., 'Mastering the Game of GO without Human Knowledge', *Nature* 550 (2017), https://www.nature.com/articles/nature24270

Snyder, Ryan and Benoît Pelopidas, 'Correspondence: New Era or New Error? Technology and the Future of Deterrence', I*nternational Security*, 43/3 (Winter 2018/9), 190-193.

Sutton, Richard S. and Andrew G. Barto. *Reinforcement Learning: An Introduction*, Vol. 1 (Cambridge: MIT Press, 1998).

Talmadge, Caitlin. 'Emerging Technology and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today', *Journal of Strategic Studies* 42/6 (2019), 864–887.

Thompson, Nicholas, 'Inside the Apocalyptic Soviet Doomsday Machine', WIRED, September 1, 2009, https://www.wired.com/2009/09/mf-deadhand/.

United Nations Institute for Disarmament Research. 'The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values' No. 3 (2015), http://www.unidir.org/files/publications/pdfs/considering-ethics-and-social-values-en-624.pdf

Vincent, James. 'Putin Says the Nation That Leads in AI 'Will Be the Ruler of the World,' *The Verge*, Sept. 4, 2017, https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world.

Volpe, Tristan. 'Dual-Use Distinguishability: How 3D-Printing Shapes the Security Dilemma for Nuclear Programs', *Journal of Strategic Studies* 42/6 (2019), 814–840.

Watson, Ben, 'The Drones of ISIS', *Defense One*, January 12, 2017, https://www.defenseone.com/technology/2017/01/drones-isis/134542/.

Williams, Heather. 'Asymmetric Arms Control and Strategic Stability: Scenarios for Limiting Hypersonic Glide Vehicles', *Journal of Strategic Studies* 42/6 (2019), 789–813.

Woolf, Amy F., 'Conventional Prompt Global Strike and Long-Rang Ballistic Missiles: Background and Issues', *Congressional Research Service*, April 6, 2018: https://fas.org/sgp/crs/nuke/R41464.pdf.