A necessary step back?

Recovering the security perspective in the debate on lethal autonomy

Christian Alwardt · Niklas Schörnig

Received: 19 April 2021 / Revised: 16 December 2021 / Accepted: 20 December 2021 / Published online: 16 February 2022
© The Author(s) 2022

Abstract A few years back, the rapid progress of international efforts to ban lethal autonomous weapon systems (LAWS) left arms controllers amazed: only five years after the founding of the International Committee for Robot Arms Control (ICRAC), the dangers of autonomous weapons were being debated in a UN context, the Convention on Certain Conventional Weapons (CCW), with non-state actors and state actors alike finding common ground in rejecting weapon systems beyond human control. Since then, however, the debate has made little progress, despite increasing pressure by activists and a strong international campaign. In this article, we will argue that the strategies used by campaigners, based on ethical and legal concerns, must be complemented by classic security-related arguments. Unfortunately, key lessons of the Cold War, including the mutual *security* benefits of arms control, seem to have been forgotten. Many concepts that are central to arms control—such as stability and verification—are by no means intuitively understood and must apparently be (re-)“learned”. Some of the world’s most important actors have not been exposed to these concepts, e.g. China, but also other players. Deconstructing military expectations regarding autonomous weapons and focusing on a preventive arms control approach could help the currently stalled process to regain the momentum it needs.

Keywords LAWS · Preventive arms control · Learning · Security

Christian Alwardt
formerly with IFSH, Hamburg, Germany

Niklas Schörnig (✉)
Peace Research Institute Frankfurt (PRIF), Baseler Straße 27–31, 60329 Frankfurt, Germany
E-Mail: schoernig@hsfk.de

Zusammenfassung Noch vor einigen Jahren versetzte die außergewöhnliche Geschwindigkeit, mit der die internationalen Bemühungen um ein Verbot letaler autonomer Waffen voranschritten, Rüstungskontrolleur:innen in großes Erstaunen. Nur fünf Jahre nach der Gründung des *International Committee for Robot Arms Control* (ICRAC) wurden die Probleme und Gefahren autonomer Waffen im Rahmen der UN Waffenkonvention (Convention on Certain Conventional Weapons – CCW) diskutiert. Und ein paar Jahre lang sah es tatsächlich so aus, als könnten sich staatliche und nichtstaatliche Akteure sich hinter dem Ziel vereinen, autonome Waffen ohne bedeutsame menschliche Kontrolle kategorisch abzulehnen. Seit einiger Zeit ist die Debatte aber trotz stärker werdenden Drucks von Aktivist:innen und einer internationalen Kampagne kaum vorangekommen. In diesem Artikel argumentieren wir, dass die Strategie der Gegner:innen autonomer Systeme, die sich im Wesentlichen auf ethische und rechtliche Argumente stützt, um klassische sicherheitspolitische Lehren ergänzt werden muss. Es scheint, dass wichtige Erkenntnisse aus dem Kalten Krieg, einschließlich der Vorteile klassischer Rüstungskontrolle, in Vergessenheit geraten sind und viele der damals bedeutsamen Konzepte, z. B. Stabilität und Verifikation, neu „gelernt“ werden müssen. Einige der wichtigsten Akteure der Welt sind mit diesen Konzepten nicht vertraut, z. B. China (aber auch andere Akteure). Gelingt es, militärische Erwartungen bezüglich autonomer Systeme zu dekonstruieren und sich gleichzeitig auf einen präventiven Rüstungskontrollansatz zu konzentrieren, könnte der derzeit ins Stocken geratene Prozess wieder die nötige Dynamik erlangen.

Schlüsselwörter · LAWS (letale autonome Waffensysteme) · Präventive Rüstungskontrolle · Lernen · Sicherheit

1 Introduction

Rumours had been circulating for some time, but on 7 August 2020 it became official: the American *Defense Advanced Research Projects Agency* (DARPA), tasked with exploring the latest military technologies, announced a computer simulated dogfight between a US Air Force pilot and an artificial intelligence (AI) (DARPA 2020). Many observers had expected the AI to win, but the clear-cut result, with the AI winning five to nil, still came as a surprise. Although this was not a “real” fight, most observers were impressed by the AI’s aggressive and unconventional behaviour (Hitchens 2020). Peter W. Singer, author of one of the most influential books on military robotics, even suggested the possibility of a “Deep Blue vs. @Kasparov63 moment in war” on Twitter,¹ referring, of course, to the first loss suffered by a reigning world chess champion to a computer program in 1996. In the summer of 2021, it was reported that the Chinese military had conducted a similar experiment with a similar result: the human was beaten by the AI (Hambling 2021).

Uncrewed aircraft flown by computer algorithms in the real world, capable of identifying and successfully attacking human opponents, would most likely fall under the category of lethal autonomous weapon systems (LAWS)—a system where

¹ <https://twitter.com/peterwsinger/status/1296540408084824064?s=20>. Accessed 15 February 2021.

in which the “critical functions” of target selection and engagement are combined in a single system without human intervention, a definition used both by the Pentagon (DoD 2012, p. 13) and by the International Committee of the Red Cross (ICRC 2021, p. 2). Rather than being restricted to environments that merely consist of pre-determined or pre-programmed hostile targets, LAWS are capable of operating in “uncontrolled” environments in which they are triggered by their surroundings, leading to “unpredictability in the consequence of their use” (Boulanin et al. 2020, p. ix). As the simulated airspace in recent experiments was highly controlled, any transfer to real-world missions would be a significant leap.

As mentioned above, LAWS are part of a larger, diverse group of technologies that are often described as “new” or “emerging” in the military context. “Emerging technologies” covers everything from modern missile technology, especially hypersonic missiles, to anything that falls under the category of “cyber” and the military use of AI and machine learning. In addition, additive manufacturing (sometimes called 3D printing), new biotechnological processes, nanotechnology and LAWS are often also labelled as “emerging”. Many of these emerging technologies have at least a partly civil origin. Yet despite this uniform classification, these diverse technologies are not best dealt with from a single arms control perspective. Depending on the state of the research, development, testing and fielding, different approaches to controlling the technology are needed.

From a theoretical point of view, the ideal solution when it comes to emerging technologies would of course be to decide on arms control measures long before the technology is ripe for fielding. “Preventive” arms control of this sort would restrict weapon systems even before individual actors have gained a significant advantage (e.g. Altmann 2006; Altmann et al. 1998), with measures aimed “at limiting, interrupting or terminating the relevant research and development processes and/or prohibiting military operations based on their conversion into weapons (systems)” (Neuneck and Mutz 2000, p. 109). In such a scenario, all states would have agreed that keeping the genie in the bottle, rather than freeing it and having to bottle it again later, is the preferable solution. While preventive arms control has many advantages in theory, however, it has rarely been implemented in practice. Prominent exceptions are the Anti-Ballistic Missile Treaty between the USA and the Soviet Union/Russia, which was terminated in 2002, and the banning of the “blinding laser”, a weapon designed to permanently blind rather than kill human opponents but with hardly any military value (Sauer 2021, p. 243). As all relevant states deemed this weapon too cruel to be used in military conflict and saw no military need for it, its use was prohibited by the Convention on Certain Conventional Weapons (CCW), a Geneva-based UN institution that focuses on the humanitarian impact of armament and its compliance with international humanitarian law. Currently, the CCW is again the centre of attention and has been hosting discussions on the potential preventive prohibition of LAWS since 2014, accompanied by a comparably intense public debate fuelled by the Campaign to Stop Killer Robots, an amalgamation of more than 180 individual NGOs.² At the moment, the debate seems to be stalled, prompting the

² <https://www.stopkillerrobots.org/members/>. Accessed 3 September 2021.

question of whether we are currently facing an arms control crisis in the sphere of emerging technologies.

Given the prominence of the debate on LAWS, this article focuses on emerging lethal autonomous weapons technologies and considers the pros and cons brought forward to justify or reject preventive prohibitions. In contrast to most of the existing literature, however, rather than focusing on the common legal and ethical arguments, we approach the issue from a structural and strategic—or security-oriented—perspective. While this perspective is not entirely novel (see Altmann and Sauer 2017), it helps us to critically assess the strategies used by non-state actors in the field in a new light. In contrast to the main body of literature, we will argue that the tremendous pressure exerted by NGOs has both advantages and disadvantages—an argument advanced only recently by Rosert and Sauer (2019). While legal and ethical arguments have helped to convince some states to support a ban, those that are technologically more advanced and focused on their own military advantage remain unimpressed. It would be no exaggeration to say that the process of restricting—or even banning—LAWS is in crisis. We agree that the issue would not be on the international agenda had it not been for the enormous pressure exerted (and solid arguments provided) by critical NGOs, but those very same NGOs have neglected to push security-related arguments when needed. In contrast to Rosert and Sauer, our approach of moderate criticism based on a security perspective arrives at a different, more radical conclusion: rather than focusing on ethical (instead of legal) arguments, we propose a more fundamental shift towards an arms control- and security-based argument that focuses on the major players. This would not only help to broaden our perspective beyond the narrow topic of LAWS but also bring the more fundamental dangers of the acceleration of warfare into clearer focus.

In consequence, we argue, first, that promoting preventive arms control requires criticizing and demystifying expectations regarding the military advantages of specific systems in order to bring major players into the debate. This means that hard security or even military arguments brought forward to justify research and development must be taken seriously and addressed in earnest. We would go even further and stress that in the triangle of law, ethics and security, the latter is the most important issue, as security issues arise independently of any concrete definition of autonomy (see below), do not depend on the lethality of the system, and affect all states, irrespective of their interpretation of international law and their ethical perspective. Unfortunately, security-related arguments are not as intuitively understandable as ethical arguments. During the Cold War, for example, Russian and American decision-makers had to “learn” the logic of arms control, stability and even mutually assured destruction in order to accept arms control as a means of mitigating the danger of a nuclear Armageddon. It was also important to learn the importance of trust-building and verification, leading to the somewhat overused slogan “trust, but verify”. Neglecting arms control in the current debate on LAWS prevents many actors who are unfamiliar with the logic of arms control—or those who have unlearned it—from fully engaging with the argument and deprives us of a valuable opportunity to use all available options to argue for a ban or other precautionary measures. Focusing on security and arms control comes with a price tag, however: we must grapple with the fact that some potential applications of a given

weapon system may follow a reasonable (albeit military) logic and should not be treated lightly, especially when cheating is easy and verifying compliance is hard. This may ultimately lead to the conclusion that a complete ban—while desirable—is currently unfeasible and that compromises must be made. At the very least, the debate must focus on whether and how classic arms control concepts and aims such as stability and verification can foster a tough and resilient arms control regime in the field of autonomous weapons, making it more attractive and overcoming what many perceive as a crisis, or at least a stand-still.

2 War at machine speed: Military expectations and the impact of automated and autonomous weapon systems

As mentioned above, preventive arms control does not have a track record of being particularly successful. As Harald Müller observes, “[a]s long as the potential of a new weapon technology cannot be fully assessed, the military and civilian military planners will not be ready to give up a project in favour of arms control” (Müller 1996, p. 403; our translation). This is particularly true in the realm of military automation, as automated weapon systems inspire military fantasies and awaken our worst fears simultaneously. As we will see below, automated weapons and networked warfare have already led to military operations that are significantly faster, more “efficient” and likely to lead to fewer casualties within one’s own ranks. However, the question of whether lethal and *fully autonomous* weapon systems in a broader sense already exist is a matter of interpretation. Some weapon systems can already be used in a way that satisfies the definition of LAWS cited above (ICRC 2021, p. 5). Terminal defence systems such as Phalanx and Sky Shield can select and engage targets without human intervention, based on pre-set targeting criteria, when switched to the automated mode. Loitering munitions such as the Israeli Harpy and its successor, the Harop—drones that loiter in a battlefield for several hours—can attack pre-selected targets such as radar stations once they detect specific radar emissions. While all of these systems can be understood as autonomous according to the DOD/ICRC definition cited above, they are not *specifically* designed to kill humans, and except for the Harpy/Harop, they currently operate in relatively controlled environments, for example the high sea. While they have been optimized for very specific scenarios, however, they represent important steps towards a possible end to the (possibly irreversible) trend toward automation that began with “simple” and remotely controlled uncrewed weapon systems, e.g. uncrewed aerial vehicles (UAVs). In this article, we therefore understand the automation of weapon systems, or “automated weapon systems”, as a precursor of future, possibly fully autonomous weapon systems with a significantly broader application. Admittedly, this is a grey area, and we can rarely be certain of where, exactly, a particular system is located on the classificatory map, especially when discussing the need for regulation.

The significant rise in the military importance of uncrewed systems over the past two decades can be explained by two factors: first, operating with these systems does not directly endanger the lives of one’s own soldiers (Mandel 2004; Sauer and Schörnig 2012); and second, the lack of an onboard human operator saves space

and weight, allowing for a more compact and efficient system design. Thus, as a main advantage, these systems can be employed in so-called “dangerous, dull and dirty” missions and have both an extended endurance and a higher weapons payload compared to crewed systems of the same size (Krishnan 2009; Singer 2009; Altmann and Sauer 2017, p. 122; Scharre 2018). In the beginning, few tasks were automated, e.g. setting a flight route by GPS coordinates in advance. “Critical functions” (ICRC 2018, p. 4) such as target selection and weapon release were (and still are) given by a human operator via radio, either by direct transmission or satellite link, leading to significant lags of up to several seconds and the risk of loss of signal. Because of this lag, today’s uncrewed weapon systems have difficulty competing with manned systems in contested military domains (Haider 2014). If this is to be overcome and their range of duties and uses in contested environments expanded, more and more tasks must therefore be automated. Such automation usually results in the acceleration of processes, which in itself represents a further military advantage and is indeed the entire point of the technology: “machine speed” has become a key driver of all automation efforts (Doll and Schiller 2019, pp. 3–4).

The maximum process speed is usually tied to the “time” required to make the associated decisions, and thus far humans have been a limiting factor in this regard. Consequently, there have been efforts to increasingly automate military decision-making and to decouple it, at least in part, from the human, thus gaining a speed advantage. Systems designed to defend warships by shooting down incoming missiles cannot wait for human operators to identify and prioritize targets and to press the trigger. Although they are only employed in crises and have a very limited range, they nonetheless represent the most likely path forward: the promise of speed and automation is meant to open up a variety of new military fields of activity, make synergies usable and increase efficiency—always accompanied by a claim to technological leadership and thus military superiority—all of which are preconditions for a worldwide military automation spiral towards war at “machine speed” (Doll and Schiller 2019, p. 4). Transferring any (or at least broader) decision-making power to machines depends on their satisfying the necessary requirements. Despite the challenges, however, military research and development has been highly focused on machine learning and artificial intelligence, with the aim of increasing the quality of machine decision-making (Haner and Garcia 2019).

Further automation, the acceleration of military processes and the prospect of future LAWS have also raised other new military expectations, including the improved mobility and interoperability of weapon systems, efficiency in command and control processes, new fields of military employment and the hope of saving costs, to name but a few (Schörnig 2010, pp. 8–13; Dickow et al. 2015; Altmann and Sauer 2017, p. 119). In addition, autonomous systems do not rely on stable communication links behind enemy lines (Dickow et al. 2015) and should therefore be unhackable and resistant to external interference (Sauer 2021). All this could allow for new military strategies, but at the very least it will have a transformative effect in terms of both military structures and operations (Zacharias 2019; Sauer 2021). Further predicted effects of the increasing automation of weapon systems, in particular processes in the military chain of command and control, include the acceleration of military analysis, improved decision-making, and improved control and execution of certain processes

and battle management as a whole (Schörnig 2014). It is expected that machine-enhanced command and control based on so-called “assistance systems” will lead to greater efficiency, the generation and utilization of new synergies, the supervision and control of a much larger number of systems, and significantly faster military operations. The future employment of such command and control structures will depend to a large extent on computing power and will require shifting a substantial amount of decision-making authority to the machines involved.

On the smaller scale, the interaction and cooperation of individual manned military systems (e.g. combat aircrafts) and highly automated and specialized uncrewed weapon systems (so-called “manned–unmanned teaming”, or MUMT) has been increasingly promoted and put forward by officials (e.g. Barnes and Evans 2010). This could, for instance, take the form of a crewed combat aircraft accompanied by several combat drones acting as forward sensor or attack platforms. Since the pilot is busy controlling her own aircraft, wing drones must therefore operate in a highly automated or autonomous way, especially in stressful combat situations or emergencies.³ In addition, this could take the form of machine–machine interactions, whether involving many individual systems operating or interacting together as a larger entity (e.g. a swarm) or a number of different, highly specialized systems forming a larger system capable of performing new or expanded military tasks. Such swarms could, for example, defeat enemy defence perimeters or attack bunker facilities in an autonomous, self-coordinated action (Hurst 2017) or even be used to conduct a first strike against nuclear assets (Altmann and Sauer 2017, p. 131).

A further likely field of application of LAWS is underwater warfare (Frandrup 2020). Thus far, we lack the technical means to establish reliable communication or control connections underwater over long distances. The ability to shadow and attack an enemy (potentially nuclear-armed) submarine with a small and silent autonomous system is attractive to many countries. Indeed, Russia has already announced the development of a new intercontinental nuclear-powered undersea autonomous vehicle—the Poseidon, designed to carry a 100 MT nuclear warhead, representing a new kind of “doomsday weapon” (Schneider 2019).

In any case, the automation of military processes is also a challenge for the military itself. From a military perspective, it may be desirable, or even necessary, to accelerate operational processes to gain the advantage and stay ahead of the game. On the other hand, militaries have no interest in losing or no longer being able to execute situational control over their own units. Yet precisely this is likely to occur as a result of a spiral towards automation. Forced to act faster and faster, humans will likely unintentionally lose control over automated weapon systems at some point. In other words: the acceleration of warfare goes hand in hand with greater automation, and vice versa, leaving military planners on a slippery slope towards full autonomy and warfare beyond control.

³ Personal conversation with a German Tornado pilot, 2015.

3 The critical perspective

3.1 From scientific criticism to an international campaign: ICRAC, the UN and the Campaign to Stop Killer Robots

Criticism of the increasing automation of warfare and the image of (lethal) autonomous weapon systems roaming the battlefield officially began in 2009 with the founding of the International Committee for Robot Arms Control (ICRAC) by Noel Sharkey (a roboticist), Jürgen Altman (a physicist), and Peter Asaro and Rob Sparrow (both philosophers). ICRAC started out as a network of experts. After ICRAC held its first international conference “Arms Control for Robots—Limiting Armed Tele-Operated and Autonomous Systems” in Berlin in 2010,⁴ and following the publication of a UN report on the trend towards the robotization of the military by Special Rapporteur Philip Alston that same year, the issue was increasingly debated by academic experts. In 2012, Human Rights Watch was the first major NGO to publish a groundbreaking report on the issue, addressing a broader public beyond academic circles (Human Rights Watch 2012). More and more NGOs took up the cause, ultimately founding the Campaign to Stop Killer Robots in 2012 with the aim of putting the issue on the international agenda in a coordinated fashion.⁵ Over the years, the number of participating NGOs rose from 7 to 165 (as of August 2020), and while some observers were surprised by its success (Carpenter 2014), the Campaign became very influential.

From the very beginning, there were three different strands of argumentation brought forward in the critique of (lethal) autonomous weapon systems: (a) arguments based on international security, (b) arguments grounded in international law and (c) arguments focused on ethical considerations. While the three formed a mutually supportive triangle, concerned members of ICRAC focused mainly on international security and international humanitarian law (IHL) (Bahcecik 2019), in particular the principle of distinction between combatants and civilians.

The NGOs’ framing of the issue was set out by UN Special Rapporteur Alston in his August 2010 report, in which he argued that “[u]rgent consideration needs to be given to the *legal, ethical and moral implications* of the development and use of robotic technologies, especially but not limited to uses for warfare” (Alston 2010, p. 21; emphasis by the authors). These arguments seemed to resonate well with NGOs, whereas the security argument was more or less ignored. Some observers have argued that keeping the focus on legal issues (i.e. the distinction principle) and on morality resonated well with older, successful campaigns, especially the campaign to ban anti-personnel landmines in 1999, the campaign to ban cluster munitions in 2008 and the campaign to ban blinding lasers. In consequence, NGOs stuck to their ostensibly winning strategy,⁶ and for a time a positive dynamic emerged. The first official forum for debating the issue of autonomous weapons took the form of an interactive dialogue held by the UN Human Rights Council (HRC)

⁴ Both authors were present at the conference.

⁵ <https://www.stopkillerrobots.org/action-and-achievements/>. Accessed 15 February 2021.

⁶ As described in detail by Rosert und Sauer (2021).

in 2013 (Barbé and Badell 2020, p. 135),⁷ where participants discussed a new UN report with a particular focus on “lethal autonomous robotics” by the new Special Rapporteur, Christof Heyns (Heyns 2013). The Heyns Report again spurred the Campaign, which intensified its lobbying for a ban.

Once the issue had found its way into the General Assembly, a French initiative shifted it from the HRC to the CCW in Geneva, where the participating states decided to launch an informal expert debate on LAWS from 2014 on. While Barbé and Badell argue that this shift from the HRC to the CCW constituted a “reframing” “in terms of security” (Barbé and Badell 2020, p. 136), at least some observers at the time were sceptical about whether the CCW was the right place to debate LAWS, given its strong focus on humanitarian issues and IHL, and were wary of the scarcity of security-related arguments.⁸ From the NGOs’ perspective, this focus and the less restricted format seemed a very good fit for the main thrust of their arguments. The informal meetings gave not only state representatives but also academic experts and NGO members the opportunity to argue their respective cases. In 2016, the fifth CCW review conference decided to continue with a formal and open-ended Group of Governmental Experts (GGE) starting in 2017. The mandate of the GGE was to submit a report to the High Contracting Parties. Whether this report could serve as a basis for a potential ban was unclear, but this didn’t stop the NGOs from celebrating: in 2012, when many state experts had only a vague idea of the problems and dangers of LAWS, no one had expected that a mere five years later an official UN body would actually be discussing the possibility of a preventive ban, which was the NGOs’ ultimate aim. *Prima facie*, critical observers had done everything to the letter.

From 2017/2018 on, after changing the format from informal talks to a GGE, the momentum began to slow. The arguments exchanged were familiar, and no substantial progress had been made. The NGOs grew restless and uneasy (Schörnig 2019); the constant pressure by NGOs, the Campaign and several open letters by scientists and IT celebrities had kept the topic in the public discourse and the media, yet the situation in Geneva, which had started out so dynamically, grew increasingly stuck, with campaigners openly debating the option of abandoning the CCW process (Rosert and Sauer 2021, p. 20).

Today, the number of states that openly support the Campaign’s call for a ban has reached 30—comparatively low given the extensive campaigning. The technologically more advanced states have failed to join the cause—with the debateable exception of China (Kania 2018). Even excluding China, however, other major players such as the US, Russia and India have—for different reasons—been unwilling to start negotiating a ban in earnest. Somewhere in between are Germany and (to a lesser extent) France, both of which have argued for a politically binding

⁷ <http://webtv.un.org/meetings-events/human-rights-council/watch/clustering-id-on-executions-and-idps-9th-meeting-23rd-regular-session-of-human-rights-council/2419860355001#full-text->. Accessed 15 February 2021.

⁸ Personal impression of one of the authors during a CCW preparation meeting at the German Foreign Office with external experts in late 2013 or early 2014. Chatham House Rules apply.

declaration.⁹ While the advocates of a ban view the non-binding agreement as too weak, those interested in pushing the technical limits forward have not supported it either.

Some argue that significant momentum was in fact achieved in 2019 when eleven principles were adopted by the CCW, including, amongst others, the applicability of IHL to the case of LAWS and the principle that “[h]uman responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines” (CCW 2019, p. 9). While some (e.g. the German Foreign office) view the principles as a stepping-stone to further limitations or even an agreement,¹⁰ others have criticized the principles as vague, redundant and in essence unnecessary (Article 36 2019). From the perspective of those aiming at least for a ban on LAWS, the CCW debate seems stuck. What started out as one of the most dynamic campaigns in the history of humanitarian arms control has come to a grinding halt. To understand this turn of events, however, we must first look at the arguments on which the NGOs of the Campaign based their criticism.

3.2 The main arguments of the critics

When it came to the legal issue, the fear that autonomous weapons would be incapable of distinguishing between soldiers and civilians was brought forward from the very beginning of the debate, with Noel Sharkey being one of the most outspoken critics (Sharkey 2009; Human Rights Watch 2012; Carpenter 2014). Doubts were raised not only about the systems’ ability to make this distinction, given their sensory and cognitive limitations, and their capacity to recognize surrender (Sparrow 2015), but also about the “inherent unpredictability” that arises when machine learning and artificial intelligence are applied to critical targeting functions (ICRC 2018, p. 2). Finally, it was (and is) argued that international law is too soft, complex and context-dependent to be reliably implemented in algorithms (Sharkey 2012). In sum: it is questionable whether LAWS will ever be able to comply with the rules of IHL (Geis 2015, p. 14).

While NGOs and critics continue to rely heavily on the assumption of the impossibility of compliance with IHL, some prominent counterarguments are worth mentioning. Ron Arkin, for example, has argued that perfect compliance is indeed impossible but that compliance on a par with or better than human judgement may be possible in the end (Arkin 2009, 2010; Carpenter 2014), a line of reasoning that has been picked up, for example, by the US delegation (Sauer 2021, p. 244). Other proponents of autonomous weapons argue that future systems could be used in limited scenarios, where the violation of IHL is either largely impossible or would be without consequence. Even outspoken critics of LAWS such as Robert Sparrow accept this possibility (Sparrow 2015, p. 701), with autonomous hunting submarines attacking opposing crewed submarines being a case in point. In any case, both sides rely heavily on expectations about future technological developments, and some

⁹ For an overview of the different positions, see Dahlmann et al. (2021, p. 2).

¹⁰ <https://www.auswaertiges-amt.de/en/aussenpolitik/themen/abruerstung/killer-robots/2277026>. Accessed 13 September 2021.

proponents of strict regulation acknowledge that technological improvements could render the argument obsolete (e.g. Rosert and Sauer 2021, pp. 21–21; Sauer 2021, p. 244). Indeed, an interesting question remains open: If it really is impossible to build IHL-compliant LAWS, would there actually be a need for a new ban, as these systems would be banned based on Article 36 of Additional Protocol I anyway?¹¹ On the other hand, if they were to pass an Article 36 review, criticism based on legal grounds would be unfounded.

The second strand of argumentation usually brought forward by proponents of a ban on LAWS is ethical in nature. The central claim is that being killed by an autonomous weapon system violates the very basic principle of human dignity (Asaro 2012; Heyns 2017; ICRC 2018; Amoroso et al. 2018, p. 32). In the beginning the argument was closely connected to IHL, but it soon emerged as a separate argument in its own right. As Special Rapporteur Heyns put it: “Even if it is assumed that LARs [lethal autonomous robots] ... could comply with the requirements of IHL, and it can be proven that on average and in the aggregate, they will save lives, the question has to be asked whether it is not inherently wrong to let autonomous machines decide who and when to kill” (Heyns 2013, p. 17). His own answer was that delegating the decision over life and death to machines “dehumanizes armed conflict even further and precludes a moment of deliberation in those cases where it may be feasible” (Heyns 2013, p. 17). Many others have since taken up the argument, operating under slogans such as “Put Human Dignity First” and contending that “human dignity too often tends to fall by the wayside” (Rosert and Sauer 2019, p. 372). The question whether human dignity is indeed violated by autonomous killing is an important one, as it indeed transcends IHL-based arguments (see Sauer 2021, pp. 253–255). While it is an empirical matter whether an algorithm is capable of distinguishing between soldiers and civilians, it is argued that the very fact that being killed by a machine decision violates the dignity of both civilians and combatants grounds a categorical rejection of LAWS (Rosert and Sauer 2019, p. 372). Unsurprisingly, some have raised doubts about the moral relevance of whether it is a human or a machine that ultimately makes the decision to kill (Birnbacher 2016). Without getting into the debate here, it suffices to point out that certain forms of automated killing, e.g. killing soldiers with landmines or booby traps, and other forms of distanced killing have faced harsh criticism, but rarely on these particular ethical grounds. This raises the question of how much human input is needed to leave dignity intact. This is also reflected in the solutions debated within the CCW, to which we now turn.

3.3 Meaningful human control: Silver bullet, distraction or dead end?

As described above, there is disagreement amongst CCW member states on how to deal with the issue of autonomous weapon systems. There are currently three main camps: on the one side are those who are in favour of a legally binding ban; on the other are those who do not see the (current) need for restrictions beyond what inter-

¹¹ Personal debate with other participants at a German Foreign Office–sponsored workshop on LAWS in Berlin in 2018. Chatham House Rules apply.

national law already regulates anyway; and in between are those who see the need for a non-binding declaration. What all participants do seem to agree on is that there should at least be some form of human control over autonomous weapons. While the US has stressed the need for “appropriate levels of human judgement”,¹² NGOs have countered with the concept of “meaningful human control” (MHC) (Roff and Moyes 2016). The basic idea is that once a weapon system (or its critical functions, to be precise) is under MHC, it can neither surprise the operator nor act in an unwanted way, leaving full responsibility with the human operator. In contrast to pro-forma control, *meaningful* human control entails certain qualities, including cognitive awareness (Roff and Moyes 2016, p. 1) and access to all relevant information on which the decision is based. Not only would MHC guarantee compliance with IHL (at least to the degree that the relevant operator is familiar with the topic), but it would also ensure that a human being is responsible for any kill decision, rendering the ethical problems less relevant. Some advocates of MHC view the concept as a silver bullet, not only solving IHL-related problems and ethical issues but also smashing the Gordian knot of definitional quarrels (Sauer 2021). Instead of defining what has to be banned, the argument goes, agreeing on MHC in all military operations as a new IHL principle would serve as a positive description of what ought to be, thereby closing loopholes and definitional discrepancies (Rosert 2017). It is safe to say that the concept of MHC is indeed an idea on which many states seem to agree, with some arguing that the CCW has become the “crystallization point” of the debate (Dahlmann et al. 2021, p. 3).

While this idea seems tempting at first, problems remain. First, even proponents of the concept admit that “the particulars of the concept have been left open” (Roff and Moyes 2016, p. 1) and that the line to be drawn “will be political, rather than purely technical” (ibid., p. 2). In other words, rather than having to define “autonomy”, states must now define the term “meaningful”. In addition, certain authors have rightly pointed out that in some instances militaries already seem to lack “meaningful” control, at least when control is understood as a relationship between an operator and a weapon. In her analysis of a typical mission execution by a jetfighter pilot, for example, Merel Ekelhof shows that the pilot “will, typically, not visually confirm the target” and that “even in today’s conventional air operations, much of the job of finding, fixing, tracking, targeting, engaging and assessing (the dynamic targeting process) is not part of the operator’s task” (Ekelhof 2019, p. 345). While Ekelhof’s example describes an air-to-surface bombing operation, one could also add air-to-air fights fought beyond the horizon. In other contexts, operators already rely heavily on what is understood as “assistant systems”, and their control over the situation is limited, to say the least (Schörnig 2014). Based on her example, Ekelhof concludes that “meaningful human control, as it is currently understood, may not be the right answer despite significant political support” (Ekelhof 2019, p. 347). As the examples show, however, this has already been superseded by reality, at least in some cases. Some proponents of the concept reply that “there is no one-

¹² <https://geneva.usmission.gov/2016/04/12/u-s-delegation-statement-on-appropriate-levels-of-human-judgment/>. Accessed 15 February 2021. The phrase had already been introduced with the Pentagon Directive 3000.01 in 2011.

size-fits-all standard of meaningful human control” (Sauer 2021, p. 242) since the need for human control is dependent on the operational circumstances. They argue that the “[c]ontrollability of weapons is arguably a proto-norm already” (Sauer 2021, p. 257), that “the emergence of this consensus is ... considerable progress in itself” (Rosert 2017, p. 1), and that making MHC a general, internationally recognized principle of IHL would “be a way out of the impasse” (ibid.).

It is nevertheless doubtful, to say the least, whether states or the military would be willing to again accept human dilatoriness in security-relevant situations where automation, as described above, has already taken root. In other words, there is a danger that states may either view the concept of meaningful control as too radical or undermine the meaning that NGOs and proponents connect to it.

4 Underestimated security policy consequences—a lever for arms control?

As the previous section has shown, while not totally absent in the broader debate (Dickow et al. 2015; Altmann and Sauer 2017; Scharre 2018; Sauer 2021) or even in CCW discussions (see above), legal and ethical issues have put security-related arguments on the back burner. Even those at the centre of the Campaign, such as Jürgen Altman (ICRAC co-founder) and Frank Sauer (ICRAC member), agree that “one particularly crucial aspect has—with exceptions confirming the rule—received comparably little systematic attention: the potential impact of autonomous weapon systems on global peace and strategic stability” (Altmann and Sauer 2017, p. 118). From the perspective of those not involved with the Campaign but who support arms control rather than disarmament (which is the perspective we will adopt below),¹³ and given the problems with the other approaches described above, security-related criticism may be the path that resonates best with those actors who are not convinced by other approaches—especially those countries with a strong and modern technological base and high hopes for automated warfare as described in Sect. 2.

The question then becomes: how likely is it that the military advantages of LAWS will change arms dynamics, leading to new arms races, endangering international stability, raising the risk of unintended conflict escalations and lowering the threshold for war? Both the normative discussion within the GGE and qualitative considerations regarding potential security implications are related to the military advantages that come along with LAWS described above. They are simply different sides of the same coin: the acceleration of military processes and diminishing the need for direct human control. Highly automated weapons and LAWS will transform warfare, making it faster and more efficient on the one hand (as described above) but also blurred and unpredictable on the other.

The most important aims that arms control agreements seek to achieve are containing destabilizing arms dynamics, reducing risk, preventing war and reducing damages and costs (Schelling and Halperin 1961). Historically, stability is connected

¹³ For the difference between an arms control and a disarmament perspective, see for example Müller (2017).

to parity as equal numbers make surprise attacks equally risky for both sides. Certain arms control instruments also aim at deceleration, i.e. by limiting the numbers of troops in a certain area. With verifiable limitations in place, conventional troop concentration will take time, attract attention and allow for preparation, reducing the chances of a surprise attack.

In truth, all of the advantages that automated and autonomous weapon systems offer to the military, as described above, run counter to these ambitions. Faster decision cycles put pressure on the opponent to level the playing field with its own automation efforts, resulting in an arms race (Altmann and Sauer 2017, p. 118). In a strategic situation, where one side is rumoured to be enjoying the advantages of automated warfare or technical military autonomy, rumour alone can be cause enough. Moreover, as the effectiveness of networked systems is less clear in such cases given force-multiplying effects, reducing transparency and raising awareness levels become necessary. These relationships amount to instability caused by an unregulated arms race (Altmann and Sauer 2017, pp. 120–121).

In addition to arms race stability, crisis stability is also of central importance (Altmann and Sauer 2017, p. 121; Sauer 2021, pp. 249–251). In time-sensitive, complex military operations and crisis situations, where a spark is sufficient to cause an explosion, trust in machine-generated situational pictures and analysis can be dangerous and could lead to unintended military actions, casualties or even crisis escalation on a broader scale. Even the widespread use of highly automated weapons systems and LAWS, whose skills are difficult for outsiders to predict, may arouse suspicion and generally increase the risk of misperception and misunderstanding between states in crisis situations and armed conflicts.

Even if decision-making ultimately rests with the human being in one or the other way, computer-generated situational awareness based on enormous amounts of (real-time) data that has been pre-processed and analysed by algorithms would in itself strongly influence any human decision. Given the time required, it would be difficult if not impossible for a human being to comprehend or verify complex situational pictures or suggested actions provided by LAWS in time-critical situations. Furthermore, automated systems may be vulnerable to external jamming and hacking (Dickow et al. 2015; Sauer 2021, p. 247). Finally, highly complex technologies, and especially software, can fall prey to malfunctions or errors that only come to bear in certain situations (Sauer 2021, p. 248), sometimes with devastating consequences. The wisdom of simply “trusting the machine” and shrugging off responsibility for decision-making would therefore seem to be in question.

It is important to note that many of these problems are related not to LAWS in particular but to automated and networked warfare in general. This is important, as the debate thus far, especially within the CCW, has focused almost exclusively on *lethal* autonomous weapon systems. This is understandable insofar as humanitarian and ethical issues almost exclusively arise in the context of lethal weapon systems. On the other hand, this focus has prevented a broader and more general debate on the acceleration of war. While there has been discussion of the impact of AI in command-and-control structures on deterrence, stability and the like (e.g. Horowitz 2018), the two debates have thus far remained separate.

LAWS, however, could add yet another layer to the problem. The potential proliferation of LAWS and the related “technology diffusion” (Sauer 2021, p. 237) will likely prompt their own security policy interactions between states. When confronted with the military superiority that such weapons provide, states will have to consider their responses carefully. Such responses could be asymmetric (e.g. involving other means of attack such as cyber and hybrid warfare or strengthening deterrence, e.g. with nuclear weapons), or they could involve the procurement and deployment of LAWS to match the other side’s capabilities. In view of the associated military advantages and lack of international regulation of autonomous weapon systems, many states could soon find themselves in a dilemma, forced to invest in the development and procurement of these ever-faster weapon systems out of fear of falling behind militarily (United States Air Force Chief Scientist 2010). The direct consequences of such an “automation spiral” could include an arms race and the massive proliferation of these weapons and related technologies. Furthermore, as a result of the proliferation of corresponding dual-use technologies and expertise, even non-state actors could acquire automated or autonomous weapon systems in the future. The threshold for using these systems would be even lower for such actors than for states in cases where conformity to IHL is not desired by the actor.

This perspective, based on security policy interests of broader automation and autonomy in weapon systems, has thus far been overshadowed by normative considerations. The focus has been on moral justification rather than calculating rationality. Even though the preventive containment of these security policy effects and the associated risks is in the common interest of all states, it is questionable whether the direction of the discussion can still be shifted in view of the path taken thus far. Although a cluster of more recent contributions have pointed in this direction (Altmann and Sauer 2017; Sauer 2021), these security policy effects and associated risks seem to have lost their relevance for many states; some even consider them manageable and view technology-driven armament as the preferred solution for maintaining or gaining military superiority. In most cases, the countries that hold this view are leaders in technology and do not want to give up the associated military advantage. The key lesson learned from the Cold War—that in an arms race, only an unstable balance can be achieved—seems to have been forgotten, as has the fact that the cooperative regulation of armament is often in the common interest, creates a stable equilibrium in the long term and comes with significant financial benefits. Neglecting these lessons seems problematic, to say the least.

5 Arms control of LAWS and other “emerging technologies”

Based on common security perceptions, arms control in general is a cooperative and largely legally binding measure between states carried out in their mutual interest. As it is an added value for all actors involved, it does not represent a zero-sum game (Müller 2017). Arms control agreements can be both multilaterally agreed within the international community and concluded bilaterally between states (Schörnig 2017). The sooner the risk potential of emerging technologies and related future weapon systems is recognized internationally, the more effectively the security policy

impacts can be contained at the international level. As mentioned in the introduction, future weapon systems that are already regarded as critical should be regulated preventively, i.e. before they are even developed, procured or used militarily, by means of preventive arms control (e.g. Altmann et al. 1998). Verification measures are the preferred instruments for ensuring the necessary confidence in compliance with arms control agreements. States that are highly concerned about security issues clearly put a premium on the prevention of cheating, that is the secret breach of an arms control treaty. However, it must also be acknowledged that concepts and approaches, how the (non)use of dual-use technologies and software in the military sector can be verified, are still in their infancy. The greater involvement of technicians and computer scientists is required to open up new perspectives and reach the necessary arms control solutions, especially in the case of LAWS (Gubrud and Altmann 2013).

Unfortunately, key lessons of the Cold War, including the mutual *security* benefits of arms control, seem to have been forgotten. Many concepts that are central to arms control—e.g. deterrence, stability, mutually assured destruction and verification—are by no means intuitively understood and will have to be “learned” anew (Nye 1987). Some of the world’s most important actors have not been exposed to these concepts, with China and India being prime examples. Furthermore, as described above, the rise of a strong and powerful civil society has shifted the focus of international arms control debates to humanitarian issues. This is unquestionably a valuable and important development, but it has meant that the security policy perspective has faded into the background. In truth, both perspectives are essential, but at different times in the debate.

We therefore find ourselves in a difficult position regarding preventive control of emerging technologies such as LAWS. As long as technology-pioneering states maintain the lead in the development and use of increasingly autonomous weapon systems, they will likely continue to have only the military promise and advantages of these weapons in mind, and the security policy risks posed by future LAWS will continue to receive little attention. Especially in technologically advanced states, ethical and international law issues are primarily seen as a matter of national consideration, as the CCW discussion has shown thus far. What is needed is greater awareness on the part of technology-pioneering states that other states and non-state actors will one day catch up in terms of the development and use of such weapons, ultimately posing a threat to everyone’s security. This is even more likely given that the key enabling technology in this case is software (rather than hardware), the proliferation of which is notoriously difficult to control. It is unclear whether the threshold for LAWS, understood as systems that can act with minimal human input in a legally permissible way, will ever be reached, but developments in technology have made it possible to come at least very close to creating autonomous weapons. Against this background, and in view of the destabilizing factors described above on the one hand and the prevailing lack of transparency on the other, all states have an interest in regulating automated weapons systems, especially future LAWS. While humanitarian and legal concerns were necessary to kick-off the debate and get the public’s attention, it is now the time to shift towards security related arguments. Security implications—which, while not completely absent, have been overshadowed

by legal and ethical considerations—may have the best chance of laying the ground for a more fruitful discussion of the risks of LAWS and the benefits of arms control in maintaining stability and peace. Civil society should strongly support these efforts.

Arms control measures related to highly automated weapons systems—and LAWS in particular—will therefore have to make reference to the relevant security policy implications and risks, as well as addressing the question of whether and how the classic features of arms control (clear definitions, clear objectives, verifiable compliance) must be adapted or re-thought (Alwardt 2020). One thing is clear: classic arms control alone is no longer sufficient. The re-thinking of arms control will be an ongoing process that has only just begun but that ultimately rests on understanding (and in some cases re-learning) classic concepts. Since providing complete answers to these questions would take us beyond the available space, only two considerations will be offered here.

5.1 Gaining human control by decelerating decision-making and military action

In the debate on MHC, one implicit assumption has been that it decelerates military operations and should be a central element of further restrictions (Boulanin et al. 2020, p. x). The deceleration of military action is not necessarily bound to human control, however, nor is it bound to having control over specific weapon systems. We therefore propose that deceleration be prioritized as the main aim of arms control measures, with MHC serving as a welcome consequence rather than an end in itself. This has several advantages: if the automation and acceleration of decision-making processes and military action are primarily the result of supposed military constraints, it may be easier to reach agreement on an international treaty aimed at decelerating military decision-making and operations or setting binding speed limits for certain military actions. In the field of nuclear weapons, such efforts have been targeted at de-alerting weapon systems, e.g. by separating delivery systems and warheads (EastWest Institute 2009). In a crisis, such measures would provide additional space for human decision-making equal to the time it takes to reassemble the system and prepare it for launch. When it comes to any automated or autonomous weapon system, time-critical decision and operation procedures should in general be slowed down artificially. No military actions should be undertaken on the basis of untested, purely computer-based analyses and recommendations. Only the verifiable deceleration of military action can guarantee MHC, and conversely, the enforcement of human control will necessarily slow down military action. Thus, if all states were to agree to decelerate in certain military fields of action—such as operations with LAWS—this would give human beings the necessary time to make more responsible decisions. Starting with deceleration rather than the “human” factor has the additional benefit of guarding against developments in human enhancement, especially cognitive enhancements, that could contribute to destabilization (Henschke 2017). If certain drugs or implants were to make enhanced humans think and act faster, for example, this would have destabilizing effects despite potentially promoting MHC.

Merely pointing out the need for a new abstract legal principle of “meaningful human control” (Rosert 2017) without defining verifiable parameters may not suffice.

5.2 Regulating the use of LAWS by rules of engagement

The international community could agree on dedicated “whitelist scenarios” where the employment of AWS would be allowed under specific directives, for example in the course of perimeter defence within a clearly defined battlefield environment or in air combat with other uncrewed weapon systems. In turn, the use of any AWS in urban and highly dynamic environments, in crisis-destabilizing scenarios or for the targeted killing of human beings should be strictly prohibited. In order to achieve international acceptance, these scenarios must have universal validity, and the use of AWS must be verifiable and monitored in an appropriated way.

Common to all approaches linking LAWS and arms control is the need for reliable verification, along with the difficulties involved. Thus far, there has been no solution to the verification of software, which in the end will make the difference regarding whether the system in question is teleoperated, “merely” automated or autonomous. The question of verification poses major challenges, in particular in view of the special importance of the quality (as opposed to the quantity) of today’s weapons systems, especially when the quality of an automated weapon system is determined by software that is difficult to assess and control (Schörnig 2015). Although no simple solution would seem to be in sight, there have been promising preliminary considerations regarding how verification could be implemented—considerations which unfortunately have not received the attention they deserve. In 2019, the International Panel on the Regulation of Autonomous Weapons (iPRAW) published a brief overview dealing with the opportunities and challenges associated with the verification of LAWS regulations (iPRAW 2019). A more concrete example with regard to deceleration and restricting autonomy by verifying human control is the approach taken by Gubrud and Altmann (2013), who present the idea of including an internationally certified “glass box” in all uncrewed weapon systems that would record specific operational details such as telemetric data. The data would be recorded, stored and, should doubts about the involvement of human control be raised, analysed by international inspectors.

In summary, along the path of “new arms control” of this sort, there must be international reconciliation on three key issues: (a) the security policy implications of automated and autonomous weapon systems, (b) the relevant factors that give rise to these implications, and, derived from these, (c) the appropriate regulatory and verification instruments that ought to be put in place. This will not be an easy task. More technical expertise is clearly needed, and a longer negotiation process will be necessary if we are to find common answers to these questions within the international community. The launching of such negotiations would be constructive in itself, however, and should be undertaken as soon as possible by both states and civil society. For this, a newly founded forum unrelated to the CCW with its limited focus on lethal autonomy could be helpful.

6 Conclusions: What to do? Dealing with a stalled process

What does all this teach us about the current crisis in arms control? In this article, we have argued that LAWS should be regulated or even banned but that the reasons brought forward by most critics—i.e. violations of ethical principles and potential non-compliance with international law—may not be enough. We agree with Elvira Roster and Frank Sauer that the strategy pursued by the Campaign to Stop Killer Robots has thus far been “less than optimal” (Roster and Sauer 2021, p. 21). Focusing on arguments of international security, stability and the acceleration of warfare may be the best way forward. The intense pressure exercised by major NGOs like Human Rights Watch and the subsequent founding of the Campaign has put the issue of military automation and autonomy on the international agenda and will continue to do so. However, now seems to be the time to find a way to better include those major technological players who still see automated and autonomous weapons as a panacea for their security problems. Focusing on security would also add a further important dimension that has been overshadowed in the current debate: dangerous developments below the threshold of lethal autonomy, i.e. the acceleration of warfare in general.

This leads us to a broader issue: When it comes to most emerging technologies, including the military use of AI in a broader sense, human enhancement, cyber and the like, experts appreciate the need for preventive arms control. When military and civilian planners have high expectations for a specific new technology, however, it will be very difficult to dissuade them from researching, developing and evaluating that technology. Two factors seem to be important for maximizing the chances of preventive arms control: first, generating awareness of specific problems beyond expert circles and scientific journals; and second, deconstructing these expectations from within the system. At least in this realm, focusing on classic security-related arguments seems reasonable—it seems a necessary step back to the past, as our title suggests. It seems imperative to *re-learn* the arms control concepts of the Cold War and to debate them in our current context. What is stability, for example, and why do certain systems contribute to it while others do not? Some actors who have become significantly more important have not been exposed to arms control thinking at all and should be included in academic and diplomatic debates on this theme more intensively. Expert dialogues, track two approaches and smaller fora more conducive to debate may be necessary.

As we have argued above, focusing on acceleration and arms control is likely to promote agreement rather than furthering disagreement. Deriving the concept of MHC from the main goal of deceleration will satisfy both the security-oriented and the legal/ethical-oriented camp. In sum: putting security on the back burner at a time when military strength and armament are viewed as serving the national interest may not be the best way forward and could contribute to deepening the current arms control crisis. Putting security and arms control at the centre comes with a price tag, however. Approaching preventive arms control from a security angle also means accepting the individual security interests of states and addressing them with convincing strategies. This can either mean settling for second-best solutions (i.e. restrictions rather than hard bans) or investing greater effort in coming up with

reliable verification mechanisms. This would seem to be the greatest challenge in the realm of emerging technologies: even where states are politically willing to accept arms control measures, it is doubtful whether they will simply accept other states at their word or be moved by vague principles such as “meaningful human control”. This is not how arms control works. Those states that are genuinely interested in the regulation of emerging technologies need verification to accept arms control, and those that are not hide behind the current lack of verification. Putting much greater effort into developing new, reliable, transparent and effective verification measures seems to be the best way forward.

Acknowledgements We would like to thank Simone Wisotzki and Ulrich Kühn, as well as the two anonymous reviewers for their very helpful and thought-provoking comments. We also would like to thank Carolyn Benson for the very helpful language editing. In both regards, the remaining mistakes are ours alone. Finally, we would like to thank the staff at the ZeFKo who were very patient with us in Corona times.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alston, Philip. 2010. *Interim report of the Special Rapporteur of the Human Rights Council on extrajudicial, summary or arbitrary executions (A/65/321)*. New York: United Nations General Assembly.
- Altmann, Jürgen. 2006. *Military nanotechnology. Potential applications and preventive arms control*. Milton Park: Routledge.
- Altmann, Jürgen, and Frank Sauer. 2017. Autonomous weapon systems and strategic stability. *Survival* 59(5):117–142.
- Altmann, Jürgen, Wolfgang Liebert, Götz Neuneck, and Jürgen Scheffran. 1998. Preventive arms control as a prerequisite for conversion of military-related R&D. In *Conversion of military R&D*, ed. Judith Reppy, Joseph Rotblat, John Holdren, and Vsevolod Avduyevsky, 255–271. London: Macmillan.
- Alwardt, Christian. 2020. Arms control of unmanned weapon systems: facing the challenges. In *A comprehensive approach to countering unmanned aircraft systems. In a comprehensive approach to countering unmanned aircraft systems*, ed. Joint Air Power Competence Centre, 414–432. Kalkar: JAPCC.
- Amoroso, Daniele, Frank Sauer, Noel Sharkey, Lucy Suchman, and Guglielmo Tamburrin. 2018. Autonomy in weapon systems. The military application of artificial intelligence as a litmus test for Germany's new foreign and security policy. https://www.boell.de/sites/default/files/boell_autonomy-in-weapon-systems_v04_kommentierbar_1.pdf. Accessed 15 Feb 2021.
- Arkin, Ronald C. 2009. Ethical robots in warfare. *IEEE Technology and Society Magazine* 28(1):30–33.
- Arkin, Ronald C. 2010. The case for ethical autonomy in unmanned systems. *Journal of Military Ethics* 9(4):332–341.
- Article 36. 2019. Critical commentary on the “guiding principles”. <https://article36.org/wp-content/uploads/2019/11/Commentary-on-the-guiding-principles.pdf>. Accessed 9 Sept 2021.
- Asaro, Peter. 2012. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross* 94(886):687–709.
- Bahcecik, Serif Onur. 2019. Civil society responds to the AWS: growing activist networks and shifting frames. *Global Policy* 10(3):365–369.

- Barbé, Esther, and Diego Badell. 2020. The European Union and lethal autonomous weapon systems: united in diversity? In *European Union contested*, ed. Elisabeth Johansson-Nogués, Martijn C. Vlaskamp, and Esther Barbé, 133–152. Cham: Springer Nature Switzerland.
- Barnes, Michael J., and A. William Evans. 2010. Soldier-robot teams in future battlefields: an overview. In *Human-robot interactions in future military operations*, ed. Michael J. Barnes, Florian Jentsch, 9–29. Surrey: Routledge.
- Birnbacher, Dieter. 2016. Are autonomous weapon systems a threat to human dignity? In *Autonomous Weapon Systems. Law, Ethics, Policy*, ed. Nehal Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu, and Claus Kreß, 105–121. Cambridge: Cambridge University Press.
- Boulanin, Vincent, Neil Davison, Netta Goussac, and Moa Peldán Carlson. 2020. Limits on autonomy in weapon systems. Identifying practical elements of human control. https://www.sipri.org/sites/default/files/2020-06/2006_limits_of_autonomy.pdf. Accessed 13 Sept 2021.
- Carpenter, Charlie. 2014. 'Lost causes'. *Agenda vetting in global issue networks and the shaping of human security*. Ithaca: Cornell University Press.
- CCW. 2019. Meeting of the high contracting parties to the convention on prohibitions or restrictions on the use of certain conventional weapons which may be deemed to be excessively injurious or to have indiscriminate effects: final report. <https://undocs.org/CCW/MSP/2019/9>. Accessed 15 Sept 2021.
- Dahlmann, Anja, Elisabeth Hoffberger-Pippa, and Lydia Wachs. 2021. Autonome Waffensysteme und menschliche Kontrolle. Konsens über das Konzept, Unklarheit über die Operationalisierung (SWP-Aktuell Nr. 31). https://www.swp-berlin.org/publications/products/aktuell/2021A31_AutonomeWaffensysteme.pdf. Accessed 23 Aug 2021.
- DARPA. 2020. Alphadogfight trials foreshadow future of human-machine symbiosis. Virtual finale showcases AI's impressive abilities in simulated F-16 aerial combat. <https://www.darpa.mil/news-events/2020-08-26>. Accessed 15 Feb 2021.
- Dickow, Marcel, Anja Dahlmann, Christian Alwardt, Frank Sauer, and Niklas Schörnig. 2015. First steps towards a Multidimensional Autonomy Risk Assessment (MARA) in weapons systems. https://ifsh.de/file-IFAR/pdf_deutsch/IFAR-WP20.pdf. Accessed 9 Sept 2021.
- DoD—Department of Defense. 2012. Department of Defense Directive Nbr 3000.09. <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>. Accessed 15 Sept 2021.
- Doll, Thomas, and Thomas Schiller. 2019. Artificial intelligence in land forces. <https://www.bundeswehr.de/resource/blob/156026/79046a24322feb96b2d8cce168315249/download-positionspapier-englisch-e-version-data.pdf>. Accessed 9 Sept 2021.
- EastWest Institute. 2009. Reframing nuclear de-alert: decreasing the operational readiness of U.S. and Russian arsenals. https://www.eastwest.ngo/sites/default/files/ideas-files/reframing_dealert.pdf. Accessed 9 Sept 2021.
- Ekelhof, Merel. 2019. Moving beyond semantics on autonomous weapons: meaningful human control in operation. *Global Policy* 10(3):343–348.
- Frandsen, Erich C. 2020. Embracing undersea robots. A US strategy to maintain undersea superiority in an age of unmanned systems. <https://www.atlanticcouncil.org/wp-content/uploads/2020/10/Embracing-Undersea-Robots.pdf>. Accessed 9 Sept 2021.
- Geis, Robin. 2015. The international-law dimension of autonomous weapons systems. <http://library.fes.de/pdf-files/id/ipa/11673.pdf>. Accessed 9 Sept 2021.
- Gubrud, Marc, and Altmann Jürgen. 2013. Compliance measures for an autonomous weapons convention (ICRAC working paper #2). http://icrac.net/wp-content/uploads/2013/05/Gubrud-Altman-Compliance-Measures-AWC_ICRAC-WP2.pdf. Accessed 15 Feb 2021.
- Haider, André. 2014. Remotely piloted aircraft Systems in contested environments. A vulnerability analysis. <http://www.japcc.org/wp-content/uploads/2015/03/JAPCC-RPAS-Operations-in-Contested-Environments.pdf>. Accessed 15 Feb 2021.
- Hambling, David. 2021. Chinese AI learns to beat top fighter pilot in simulated combat. <https://www.forbes.com/sites/davidhambling/2021/06/18/chinese-ai-learns-to-beat-top-fighter-pilot-in-simulated-combat/?sh=19cfd1471b8c>. Accessed 3 Sept 2021.
- Haner, Justin, and Denise Garcia. 2019. The artificial intelligence arms race: trends and world leaders in autonomous weapons development. *Global Policy* 10(3):331–337.
- Henschke, Adam. 2017. 'Supersoldiers': Ethical concerns in human enhancement technologies. <https://blogs.icrc.org/law-and-policy/2017/07/03/supersoldiers-ethical-concerns-human-enhancement-technologies-2/>. Accessed 13 Sept 2021.
- Heyns, Christof. 2017. Autonomous weapons in armed conflict and the right to a dignified life: an African perspective. *South African Journal on Human Rights* 33(1):46–71.

- Heyns, Christof. 2013. Report of the special rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns (A/HRC/23/47). <http://undocs.org/A/HRC/23/47>. Accessed 15 Feb 2021.
- Hitchens, Theresa. 2020. AI slays top F-16 pilot in DARPA dogfight simulation. <https://breakingdefense.com/2020/08/ai-slays-top-f-16-pilot-in-darpa-dogfight-simulation/>. Accessed 15 Aug 2021.
- Horowitz, Michael. 2018. Artificial intelligence, international competition, and the balance of power. *Texas National Security Review* 1(3):36–57.
- Human Rights Watch. 2012. Losing humanity. The case against killer robots. https://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf. Accessed 9 Sept 2021.
- Hurst, Jules. 2017. Robotic swarms in offensive maneuver. *Joint Forces Quarterly* 87(4):105–111.
- ICRC—International Committee of the Red Cross. 2018. Ethics and autonomous weapon systems: An ethical basis for human control? https://www.icrc.org/en/download/file/69961/icrc_ethics_and_autonomous_weapon_systems_report_3_april_2018.pdf. Accessed 9 Sept 2021.
- ICRC—International Committee of the Red Cross. 2021. ICRC position on autonomous weapon systems. ICRC position and background paper. https://www.icrc.org/en/download/file/166330/icrc_position_on_aws_and_background_paper.pdf. Accessed 2 Sept 2021.
- iPRAW—international Panel on the Regulation of Autonomous Weapons. 2019. Verifying LAWS regulation—opportunities and challenges. https://www.ipraw.org/wp-content/uploads/2019/08/2019-08-16_iPRAW_Verification.pdf. Accessed 15 Feb 2021.
- Kania, Elsa. 2018. China's strategic ambiguity and shifting approach to lethal autonomous weapons systems. <https://www.lawfareblog.com/chinas-strategic-ambiguity-and-shifting-approach-lethal-autonomous-weapons-systems>. Accessed 15 Sept 2021.
- Krishnan, Armin. 2009. *Killer robots. Legality and ethicality of autonomous weapons*. Farnham: Ashgate.
- Mandel, Robert. 2004. *Security, strategy, and the quest for bloodless war*. Boulder: Lynne Rienner.
- Müller, Harald. 1996. Von der Feindschaft zur Sicherheitsgemeinschaft – Eine neue Konzeption der Rüstungskontrolle. In *Eine Welt oder Chaos?*, ed. Berthold Meyer, 399–428. Frankfurt a.M.: Suhrkamp.
- Müller, Harald. 2017. Learning unit 1: WMD, conventional weapons and arms control: basic concepts. <https://nonproliferation-elearning.eu/learningunits/arms-control-basics/>. Accessed 15 Feb 2021.
- Neuneck, Götz, and Reinhard Mutz. 2000. *Vorbeugende Rüstungskontrolle*. Baden-Baden: Nomos.
- Nye, Joseph. 1987. Nuclear learning and U.S.-soviet security regimes. *International Organization* 41(3):371–402.
- Roff, Heather, and Richard Moyes. 2016. Meaningful human control, artificial intelligence and autonomous weapons. <https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>. Accessed 9 Sept 2021.
- Rosert, Elvira, and Frank Sauer. 2019. Prohibiting autonomous weapons: put human dignity first. *Global Policy* 10(3):370–375.
- Rosert, Elvira, and Frank Sauer. 2021. How (not) to stop the killer robots: A comparative analysis of humanitarian disarmament campaign strategies. *Contemporary Security Policy* 42(1):4–29.
- Rosert, Elvira. 2017. How to regulate autonomous weapons (PRIF spotlight 6/2017). https://www.hsfk.de/fileadmin/HSFK/hsfk_publicationen/Spotlight0617.pdf. Accessed 15 Feb 2021.
- Sauer, Frank. 2021. Stepping back from the brink: Why multilateral regulation of autonomy in weapons systems is difficult, yet imperative and feasible. *International Review of the Red Cross* 102(913): 235–259.
- Sauer, Frank, and Niklas Schörnig. 2012. Killer drones—The silver bullet of democratic warfare? *Security Dialogue* 43(4):363–380.
- Scharre, Paul. 2018. *Army of none. Autonomous weapons and the future of war*. New York: W.W. Norton & Company.
- Schelling, Thomas C., and Morton H. Halperin. 1961. *Strategy and arms control*. New York: The Twentieth Century Fund.
- Schneider, Mark. 2019. Russia's massive nuclear weapons arsenal is a threat. <https://nationalinterest.org/blog/buzz/russias-massive-nuclear-weapons-arsenal-threat-59947>. Accessed 15 Feb 2021.
- Schörnig, Niklas. 2010. *Robot Warriors. Why the Western investment into military robots might backfire* (PRIF Report No. 100). Frankfurt a.M.: HSFK.
- Schörnig, Niklas. 2014. Automatisierte Kriegsführung – Wie viel Entscheidungsraum bleibt dem Menschen? *Aus Politik und Zeitgeschichte* 64(35–37):27–34.
- Schörnig, Niklas. 2015. From quantitative to qualitative arms control: the challenges of modern weapons development. In *Global trends 2015: prospects for world society*, ed. Michèle Roth, Cornelia Ulbert, and Tobias Debiel, 87–100. Bonn: Stiftung Entwicklung und Frieden/Development and Peace Foundation.

- Schörnig, Niklas. 2017. Rüstung, Rüstungskontrolle und internationale Politik. In *Handbuch Internationale Beziehungen*, ed. Frank Sauer, Carlo Masala, 959–990. Wiesbaden: Springer VS.
- Schörnig, Niklas. 2019. Durchbruch oder Sackgasse? Die Debatte zu letalen autonomen Waffensystemen geht in Genf in die vielleicht letzte Runde. <https://blog.prif.org/2019/03/25/durchbruch-oder-sackgasse-die-debatte-zu-letalen-autonomen-waffensystemen-geht-in-genf-in-die-vielleicht-letzte-runde/>. Accessed 15 Feb 2021.
- Sharkey, Noel. 2009. Weapons of indiscriminate lethality. *FIFF-Kommunikation* 1(09):26–29.
- Sharkey, Noel. 2012. The evitability of autonomous robot warfare. *International Review of the Red Cross* 94(886):787–799.
- Singer, Peter W. 2009. *Wired for war*. New York: Penguin.
- Sparrow, Robert. 2015. Twenty seconds to comply: autonomous weapons systems and the recognition of surrender. *International Law Studies* 91:699–728.
- United States Air Force Chief Scientist. 2010. *Report on technology horizons. A vision for air force science & technology during 2010–2030*. Vol. 1. Washington: Office of the Chief Scientist of the U.S. Air Force. AF/ST-TR-10-01-PR.
- Zacharias, Greg. 2019. Autonomous horizons: the way forward. https://www.airuniversity.af.edu/Portals/10/AUPress/Books/b_0155_zacharias_autonomous_horizons.pdf. Accessed 9 Sept 2021.