Autonomous Weapon Systems – Dangers and Need for an International Prohibition

Jürgen Altmann(✉)

Experimentelle Physik III, TU Dortmund, 44221 Dortmund, Germany
juergen.altmann@tu-dortmund.de

Abstract. Advances in ICT, robotics and sensors bring autonomous weapon systems (AWS) within reach. Shooting without control by a human operator has military advantages, but also disadvantages – human understanding of the situation and control of events would suffer. Beyond this, compliance with the law of armed conflict is in question. Would it be ethical to allow a machine to take a human life? The increased pace of battle may overburden human understanding and decision making and lead to uncontrolled escalation. An international campaign as well as IT, robotics and AI professionals and enterprises are calling for an international ban of AWS. States have discussed about limitations in the UN context, but no consensus has evolved so far. Germany has argued for a ban of fully autonomous weapons, but has not joined the countries proposing an AWS ban, and is using a problematic definition.

An international ban could comprise a prohibition of AWS and a requirement that each use of force must be under meaningful human control (with very few exceptions). If remotely controlled uninhabited weapon systems remain allowed, a-priori verification that they cannot attack under computer control is virtually impossible. Compliance could be proved after the fact by secure records of all communication and sensor data and the actions of the human operator.

The AI and robotics communities could make significant contributions in teaching and by engaging the public and decision makers. Specific research projects could be directed, e.g., at dual use, proliferation risks and scenarios of interaction between two fleets of AWS. Because of high military, political and economic interests in AWS, a ban needs support by an alert public as well as the AI and robotics communities.

Keywords: Autonomous weapon system · Combat robot · Armed uninhabited vehicle · Preventive arms control

1 Introduction

In modern armed forces information and communication technologies (ICT) play an ever increasing and by now central role. Advances in ICT have enabled precision-guided missiles in the 1980s and combat drones in the 2000s that in the attack function up to now are being remotely controlled, even though other functions are often carried out autonomously, by the control software on board. Such armed uninhabited air vehicles (UAVs) have proliferated and are deployed by about 30 countries at present. The next step would be machine autonomy also in the weapon use. Such autonomous

C. Benzmüller and H. Stuckenschmidt (Eds.): KI 2019, LNAI 11793, pp. 1–17, 2019.
https://doi.org/10.1007/978-3-030-30179-8_1

weapon systems (AWS) are the subject of intense research and development (R&D) in several countries. A part of such work is devoted to swarms that promise additional military advantages. Artificial intelligence (AI) is seen as an important enabler. AWS could act in all environments: not only in the air, but also on land, on and under water and principally also in outer space (where the conditions for movement and action are very different).

AWS that would select and engage targets by algorithm, without human control, would pose problems in several areas, from ethics via international law to military stability and international security. Accordingly, a scientific debate as well as discussions among states have unfolded. Much literature exists, exemplary overview works are [1] and [2].

The next Sect. 2 describes the trend from remotely controlled armed UAVs to AWS. Then I take a look at military-technology assessment of AWS. Section 4 covers efforts to raise awareness about the problems and discussions among states. Problems with inappropriate definitions in particular by Germany are discussed in Sect. 5. A possible design of an international AWS ban is presented in Sect. 6. Section 7 is devoted to potential contributions by the AI and robotics communities, and Sect. 8 presents a conclusion.

2 From Remotely Controlled Armed Drones to Autonomous Weapons

Armed UAVs have become prominent with the US attacks in the Middle East since 2001, later expanded to regions in Pakistan and Africa. The number of carriers and operations has increased strongly, as has the number of countries with armed drones – today about 30, partly by indigenous development, partly by imports (with China the biggest exporter) [3, 4]. While these UAVs may have automatic functions – not only for flight and trajectory control, some can take off and land autonomously – the attack function is controlled by a human operator via a communication link. Based on background information and real-time video he or she selects a target and attacks ("engages") it. This requires reliable communication that is not endangered in present, strongly asymmetric conflicts, but could be disturbed or interrupted by a more capable adversary. Purely military arguments speak for AWS, however they would bring serious problems in several areas, as shown below.

There are already weapons that attack on their own, without human control. Mines are a very primitive example; some air- or ship-defence systems have an automatic mode where fast incoming projectiles or missiles are detected by radar and an automatic cannon is directed so that its munitions hit the object, or a target-seeking missile is launched against it. These systems are usually called "automatic". They work at short range, in a restricted environment and against a simple class of objects – nevertheless, even here attacks against wrong targets occurred.[1]

[1] Fratricides by US Patriot missiles in the 2003 war against Iraq [53].

Fully autonomous weapons systems, on the other hand, would move in a certain area over a considerable time, search for targets, identify and attack them without human control or oversight. A useful definition was given by the US Department of Defense (DoD):

> "autonomous weapon system. A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation." [5] (emphasis original)[2]

This sensibly includes immobile systems such as fixed armed guardian robots. Autonomy in other functions, such as trajectory control or co-ordination with other systems, is not considered here. The same meaning is also captured by the International Committee of the Red Cross (ICRC), the institution charged with overseeing international humanitarian law (IHL):

> "The ICRC's working definition of an autonomous weapon system is: 'Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.'" [6, p. 4]

Often the degree of human control is indicated with respect to the decision loop that is repeated permanently, in US military parlance the "observe, orient, decide, act" (OODA) loop. A human is "in the loop" if he/she commands the attack. If the weapon system acts on its own and the human only supervises the action, possibly with the ability to intervene, the human is "on the loop". If there is no supervision or influence, the human is "out of the loop".[3]

Already remotely controlled uninhabited weapon systems promise several important advantages to armed forces: soldiers are removed from the place where they apply force and thus are much less exposed to danger (however for a competent adversary the control stations are valuable – and legitimate – targets). The endurance of systems is no longer limited by the human need for rest, since remote operators can work in shifts. When human operators no longer need to be accommodated on board, systems can be made smaller, more light-weight and can go beyond human limits e.g. with respect to acceleration, thus can fly narrower curves. And systems could become cheaper (although this can be questioned with respect to the total cost, e.g. [7]). Politically, the decision to apply force in or against another country is easier to take if soldiers need not be sent there and exposed to danger [8].

Proceeding from remotely controlled to autonomous weapon systems would bring additional advantages in combat: without a permanent communication link the systems are more difficult to detect. Reacting to events can be faster since a distant control

[2] The US-DoD definition of "semi-autonomous weapon", "a weapon system that is intended to only engage individual targets or specific target groups that have been selected by a human operator" [5] is more problematic in that it does not specify how targets or target groups are to be selected.

[3] For a more differentiated autonomy scale with six steps see [55].

station does not need to be involved – the two-way communication time can be as low as a few seconds, but in a fast-paced battle at short distance such a delay can mean the difference between use or loss of one's systems. Not relying on communication also means that a link cannot be jammed or otherwise disrupted. Personnel can be saved when operators no longer need to observe the situation and control the weapons. Autonomy allows attacks by swarms where human control of the individual elements would be practically impossible.

Given these advantages, it is logical that militaries have a high interest in AWS. However, there are counterarguments from a military view: AWS would reduce predictability and control over events in combat. Also, AWS could be hacked or hijacked.

Several precursors of AWS have been developed and some deployed. The Israeli loitering drone Harpy searches for enemy radars, flies into them and explodes. At the Korean demilitarised zone the immobile guardian robots SGR-1A has an option for autonomous shooting, but it seems not to be used. The range of some fire-and-forget missiles is being expanded greatly so that targets can no longer be designated from the launch position, necessitating search in a wider area. Examples with sensors and algorithms for target recognition are the US-Navy Long Range Anti Ship Missile (LRASM) with several 100 km range or the UK Brimstone missile (35 km).

In the series of its "Roadmaps for Unmanned Systems" that the US DoD has issued from 2007 on it has emphasised autonomy from the beginning as one important strand, stating as a general guideline for processor technology:

> "the ultimate goal is to replace the operators with a mechanical facsimile or equal or superior thinking speed, memory capacity, and responses gained from training and experience." [9, p. 53]

In 2012 the US DoD issued a directive "Autonomy in Weapon Systems" that stated conditions for their deployment and use [5]. Specific emphasis was placed on "rigorous hardware and software V[erification] & V[alidation] and realistic system developmental and operational T[est] & E[valuation], including analysis of unanticipated emergent behavior resulting from the effects of complex operational environments". A regression test of the software is to be applied after changes, using automated tools whenever feasible.

Following the Roadmaps, R&D of AWS is continuing, but for the near-term future the US military focuses on "manned-unmanned" or "human-machine teaming" (e.g. [10, p. 19], [11, p. 31]. Prototypes are being developed and demonstrated in several countries, in the area of jet-engine uninhabited combat air vehicles (UCAV) for example the US X-47B with autonomous aerial refuelling and landing and take-off on an aircraft carrier. The UK Taranis was to provide autonomous combat capability originally, but later human control was emphasised; for the British-French successor Future Combat Air System (FCAS), to be deployed from about 2030, autonomous attack is included as an option. At sea the US Defense Advanced Research Projects Agency (DARPA) has the Sea Hunter demonstrator. US Army, Navy and Air Force

have various R&D projects for autonomy in armed unmanned vehicles. The Navy has demonstrated a swarm of motor boats, the Air Force showed a swarm of 103 micro drones released from combat jet aircraft.

Russia is proceeding with armed uninhabited ground vehicles in various sizes and tries to catch up in armed UAVs. Weapons will be under human control for the near future, but autonomous operation by artificial intelligence (AI) is taken into view [12]. In China military R&D aim at intelligent and autonomous weapon systems in all media, with systematic military-civil fusion for fast uses of AI advances in both areas [13]. However, among the three main competitors USA, Russia and China, the latter is the only one that has warned of an arms race and called for a ban on "lethal autonomous weapon systems (LAWS)" [14, pp. 4–7].

3 Military-Technology Assessment of AWS

Questions whether introduction of a new military technology would be good or bad are the subject of military-technology assessment. In this process judgement can be based on the criteria of preventive arms control. They can be arranged in three groups, referring to:

- Existing or intended arms-control treaties; international humanitarian law; weapons of mass destruction.
- Military stability between potential adversaries; arms races; proliferation.
- Human health, environment, sustainable development; societal and political systems; the societal infrastructure.

When a danger in one of these areas is to be feared, considerations about preventive limitations should be done, including methods and means of verification of compliance [15, 16]. AWS raise problems in all fields.

Concerning arms control, existing treaties could be endangered by AWS that would differ from traditional carriers of nuclear or conventional weapons and would thus fall outside of treaty definitions or at least in grey areas, leading to complicated discussions. For example: should small armed UAVs count as combat aircraft under the Treaty on Conventional Armed Forces in Europe (CFE Treaty, 1990)? This problem arises with uninhabited armed vehicles in general, that is already with remotely controlled ones [17].

With respect to the laws of armed conflict, so-called international humanitarian law (IHL), one can state that computer programs will for a long time not be able to reliably discriminate between combatants and non-combatants and to do appropriate assessments of proportionality between damage to civilians or civilian objects on the one hand and the expected military advantage from an attack against a legitimate target on

the other. Even the roboticist who did the most research for IHL-compliant AWS-control software wrote of "daunting challenges" that remain.[4] Another roboticist has criticised this work.[5]

Because AWS proper do not exist yet, arms races and proliferation cannot be observed in reality. But one does not need much imagination to extrapolate from the respective developments in remote-control armed UAVs where arms races and proliferation are going on at high speed. Should one important state start fielding AWS, the others will certainly follow suit, and because of the high military interests the qualitative and quantitative arms race and proliferation will likely proceed much faster than with remote-control armed UAVs [18].

Arms racing can be seen as one dimension of deteriorating military stability, working on a time scale of years and decades. The second dimension concerns crisis instability and escalation, working in much shorter time frames. The general military situation between potential adversaries can be called stable if neither side could gain from attacking another. In particular, if in a severe crisis, when the sides assume that war can begin immediately, pressure exists to strike pre-emptively first, because otherwise one would suffer strong losses, this would be referred to as instable. The fear of crisis instability and escalation into war was a central topic of the debate about nuclear weapons and motivated the nuclear arms-control treaties, but similar considerations hold for conventional forces, too.[6] AWS would create specific problems here [18]. The individual systems as well as swarms would be programmed in advance how to act and react. In particular in duel-type situations at short range this would mean to shoot back immediately on indications of being attacked, because waiting a few seconds for a human to assess the situation and send a counter-attack order can mean the loss of one's systems before they could have launched their weapons. In such a situation, false indications of attack – sun glint interpreted as a rocket flame, sudden and

[4] "The transformation of International Protocols and battlefield ethics into machine-usable representations …", "Mechanisms to ensure that the design of intelligent behaviors only provides responses within rigorously defined ethical boundaries", "The development of effective perceptual algorithms capable of superior target discrimination capabilities …", "The creation of techniques to permit the adaptation of an ethical constraint set and underlying behavioral control parameters that will ensure moral performance …", "A means to make responsibility assignment clear and explicit for all concerned parties …" [41, p. 211f.].

[5] "The work … is, in fact, merely a suggestion for a computer software system for the ethical governance of robot 'behaviour'. This is what is known as a 'back-end system'. Its operation relies entirely on information from systems yet 'to be developed' by others sometime in the future. It has no direct access to the real world through sensors or a vision system and it has no means to discriminate between combatant and non-combatant, between a baby and a wounded soldier, or a granny in a wheelchair and a tank. It has no inference engine and certainly cannot negotiate the types of common sense reasoning and battlefield awareness necessary for discrimination or proportionality decisions. There is neither a method for interpreting how the precepts of the laws of war apply in particular contexts nor is there any method for resolving the ambiguities of conflicting laws in novel situations." [50].

[6] Note that the CFE Treaty in its preamble calls for "establishing a secure and stable balance of conventional forces at lower levels" and for "eliminating disparities detrimental to stability and security". [46] Unfortunately the Treaty is no longer operating with respect to Russia.

unexpected movements of the adversary or a simple malfunction – could trigger escalation. Obviously such interactions could not be tested or trained together beforehand. The outcome of two separate programmed systems interacting with each other cannot be predicted, but if they control armed robots that are directed against each other, fast escalation to actual shooting is plausible. When war would already be occurring, autonomous-weapon interactions could lead to escalation to higher levels, principally up to nuclear-strategic weapons. Similar runaway escalations between different algorithms are being observed in the computer trade at stock exchanges, but after the first severe flash crashes an overarching authority got the possibility to stop the trade, acting as a circuit breaker. No such authority exists in the international system that could interrupt "flash wars".[7]

With respect to the third criteria group, mainly concerned with consequences of new military technology in peace time, the most problematic outcome would be uses of AWS by criminals and in particular terrorists. Such actors could build simple AWS, but not sophisticated ones. The latter would be developed by states, with immensely higher personnel and financial resources, if an AWS arms race between states cannot be prevented. Once built, deployed and exported, however, such systems would probably also get into the hands of non-state actors.

4 Calls, Open Letters and Discussions Among States

Concerned by the increasing use of uninhabited armed systems and by the foreseeable trend to transition to autonomous attack, in 2009 four academics founded the International Committee for Robot Arms Control (ICRAC).[8] In 2012 the international Campaign to Stop Killer Robots was formed that today comprises 106 non-governmental organisations from 54 countries [19]. In 2013 the UN Special Rapporteur on extrajudicial, summary or arbitrary executions for the Office of the High Commissioner for Human Rights called on all states to place a national moratorium on "lethal autonomous robotics" [20]. Many states felt uneasy about the prospect of autonomous killing by machine, so in October 2013 more than 30 countries addressed the problem in the UN General Assembly, and in November 2013 the member states of the Convention on Certain Conventional Weapons (CCW)[9] decided to begin expert discussions on what then became to be called "lethal autonomous weapon systems (LAWS)".

[7] Similar unpredictable, but probably escalatory interactions can be foreseen if offensive cyber operations were done under automatic/autonomous/AI control. Combined with AWS operations the problems could intensify each other.

[8] The author was one of the founders. In the meantime the number of members has grown to 33 [52].

[9] The full name is "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". This framework convention was concluded in 1980 and has five specific protocols, the most relevant in the present context being Protocol IV that prohibits blinding laser weapons [45]. There are 125 member states, including practically all states with relevant militaries [51].

Such discussions were held regularly since 2014 in Geneva, from 2017 on as a formal Group of Governmental Experts (GGE). Invited talks and extensive discussions were held under topics such as autonomy, IHL, ethics, security, a working definition or characterisation of LAWS. Some progress was made: there is general agreement that some form of human control of the use of force is needed. The notion of "meaningful human control" has gained considerable traction.[10] But strong differences of opinion showed up and could not be bridged. By end-2018 28 countries had spoken out for a legally binding prohibition of AWS and the Non-Aligned Movement has called for prohibitions or regulations. But about a dozen states have declared that they do not support negotiating a ban treaty, among them USA, UK, Russia, Israel, Australia and France [21]. Because a consensus rule is used in the CCW, no decision can be taken if only one member state is opposed. Thus a mandate for negotiations on a prohibition protocol is not to be expected. It is open if sidestepping the CCW and negotiating a separate treaty outside, with the help of benevolent states, would be sensible. This approach had led to the Antipersonnel Mine Convention in Ottawa 1997 and the Cluster Munition Convention in Oslo 2008, but military interests in AWS may be much higher. Achieving an international prohibition probably will need more public pressure in the critical countries.

In this respect, various activities in particular from the robotics and AI community have been helpful. In 2012/13 270 computer scientists, engineers, AI experts, roboticists and professionals from related disciplines called for a prohibition on the development and deployment of AWS [22]. An open letter launched in 2015 arguing against an arms race in AI-enabled weapons got more than 4,500 signatures from AI and robotics researchers and above 26,000 others [23, 24]. Particular attention was raised by the video "Slaughterbots" showing possible future swarms of microdrones that seek out individuals and kill them by a shaped explosive applied to the head; due to cheap series production one could even speak of weapons of mass destruction [25].[11] In a 2018 pledge signed by nearly 250 AI-related organisations, many of them renowned hi-tech firms, and more than 3,200 individuals, the signatories have called on governments to prohibit AWS and, absent that, promise that they "will neither participate in nor support the development, manufacture, trade, or use of lethal autonomous weapons" [26]. In Germany in 2019, the Gesellschaft für Informatik has called for a legal ban on lethal autonomous weapons, in particular for action by the Federal Government [27]. Also the Federation of German Industries (BDI) requests a legally binding ban on fully autonomous weapon systems [28]. Particularly remarkable is the opposition of software developers and other employees in big AI/robotics companies (e.g. [29, 30]).

[10] What this can mean in detail is explained in [54].

[11] The Swiss Federal Office for Defence Procurement – armasuisse – has re-enacted the scene and shown that a shaped charge of 3 g explosive can penetrate a skull emulator [47].

5 Problems with Definitions

The discussions about possible limitations hinge on what exactly is meant by an AWS. The understandings of countries differ considerably. The US DoD definition cited in Sect. 2 is succinct and functional. The UK Ministry of Defence (MoD), on the other hand, differentiates between "automatic" and "autonomous" systems:

"Automated system	In the unmanned aircraft context, an automated or automatic system is one that, in response to inputs from one or more sensors, is programmed to logically follow a predefined set of rules in order to provide an outcome. Knowing the set of rules under which it is operating means that its output is predictable. (JDP 0-01.1)
Autonomous system	An autonomous system is capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be. (JDP 0-30)" [31, p. 13]

It is very questionable whether the output of a programmed system, logically following predefined rules, will be predictable always. On the one hand, there are the well-known problems from software complexity, on the other there can be changes in the environment and unforeseen actions of an adversary that can lead to unexpected results that have not shown up in testing or training. The requirements given for calling a system autonomous, on the other hand, raise the bar very high. It is thus easy for the MoD to state: "The UK does not possess fully autonomous weapon systems and has no intention of developing them. Such systems are not yet in existence and are not likely to be for many years, if at all." [31, p. 14] The obvious conclusion is that there is no urgency about AWS, and that automatic systems are not problematic.

Similar questionable understandings are found in the German definitions. The draft definition of the Federal Ministry of Defence (BMVg) states:

> "Lethal Autonomous Weapon System means a weapon system which is primarily designed to directly cause lethal effects to human beings, and which is designed to sense, rationalize, decide, act, evaluate and learn completely independent from human interaction or control.
> …
> Lethal (in the context of this definition) means aiming to cause directly death or fatal injury to a single or several human beings.
> …
> According to this definition, an autonomous system is to be distinguished from a highly/fully automated system which, although not requiring human intervention, operates alone on the basis of a fixed set of inputs, rules, and outputs in a deterministic and largely predictable manner. In particular it is not capable of defining its tasks and goals on its own." [32]

In its proposal at the CCW GGE, the German delegation has used the same understanding:

> **"Statement delivered by Germany on Working Definition of LAWS/'Definition of Systems under Consideration'**
> …
> Germany rejects autonomous weapon systems which are primarily designed to directly cause lethal effects or other damage to human beings, and which are designed to sense, rationalize, decide, act, evaluate and learn completely independently from human interaction or control.
> …
> An understanding of autonomy – and its distinction from automatization – for the working purposes of the GGE on LAWS can be built by referring to autonomy as
>
> – the capacity to perceive (sense and interpret) an environment,
> – evaluate the circumstances of a changing situation without reference to a set of pre-defined goals,
> – reason and decide on the most suited approach towards their realization,
> – initiate actions based on these conclusions,
> – all of the above being executed without any human involvement once the system has been operationalized.
>
> …
> A new quality of technology towards autonomy based on artificial intelligence is reached where systems or individual functionalities have the ability to learn and thus re-define their goals and/or develop new strategies to adapt their respective performance to different situations. Having the ability to learn and develop self-awareness constitutes an indispensable attribute to be used to define individual functions or weapon systems as autonomous.
> **…**
> As a consequence Germany does not intend to develop or to acquire weapon systems that completely exclude the human factor from decisions about the employment of weapon systems against individuals." [33] (emphasis original)

The position of Germany is even more problematic than the UK one. 1. Restricting the notion of LAWS[12] to weapon systems primarily directed against human beings exempts all systems designed to attack military objects. 2. What does "completely excluding the human" mean? An interpretation that a human would still be involved if he or she would activate an AWS that then searches targets on its own is at least compatible with that wording. 3. Demanding the ability to learn and develop self-awareness for calling a system autonomous means that autonomy lies far in the future. But it is just systems without these abilities, but programmed to select and engage targets autonomously, that could be developed and deployed in the next five to ten years and would endanger IHL, would bring an arms race and proliferation and would destabilise the military situation – it is these systems that should be prohibited preventively before deployment would begin.

[12] In military parlance, "lethal" is mostly understood as "destructive", not explicitly as killing people, as e.g. in military notions of "target kill" or "mission kill". The use of the term LAWS for the CCW expert meetings was not intended for exclusion of weapons against materiél or of non-lethal weapons (personal communication from Ambassador Jean-Hugues Simon-Michel of France, first chair of the expert meetings).

Also, Germany has not joined the 28 countries that in the CCW GGE meetings have called for a legally binding prohibition of AWS [21]. This fact and the AWS definition do not fit to the strong statement of Foreign Minister Maas at the 2018 UN General Assembly:

> "Our common rules must keep pace with technological developments. Otherwise, what currently sounds like science fiction may very soon become deadly reality – autonomous weapons systems, or killer robots, that kill without any human control. I ask that you please support, both here in New York and in Geneva, our initiative to ban fully autonomous weapons – before it is too late!" [34]

If Germany were serious about an AWS ban, it could act accordingly in the CCW GGE meeting and join the group of countries demanding a legally binding prohibition. Unilaterally it could change to a simple AWS definition oriented at the critical functions of target selection an engagement (similar to the US DoD and ICRC definitions cited in Sect. 2), could issue guidelines for the Federal Armed Forces and prescribe meaningful human control for all their weapons; these and other proposals have been made by an international work group [35].

6 Possible Design of an International AWS Ban [17, 36]

In order to prevent violations of IHL, arms races and destabilisation by AWS, a legally binding international prohibition of AWS is needed. Systematically, it would fit best into the CCW, it could be added as Protocol VI to this framework convention. Protocol IV banning the use of laser blinding weapons could serve as a role model, but to be effective in the face of much higher military interests in AWS, the prohibition should be comprehensive, that is, hold not only for use, but also for the earlier stages of deployment, testing and development. This negative obligation should be complemented by a positive requirement that for each use of force the selection of the target and the decision to engage are made by a human being who is responsible and accountable. If consensus about adding such a protocol will be impossible in the CCW, then a separate treaty can be used as discussed above.

If AWS will be banned while remotely controlled uninhabited weapons systems remain allowed, a difficult verification problem arises, since the very same weapons systems could attack under remote control or autonomously – there would be no difference that could be observed from the outside. Thus, from a verification standpoint a complete ban of all armed uninhabited vehicles would be best, since then one could check by on-site inspections whether the uninhabited vehicles (e.g. for reconnaissance) have appliances for weapons, such as a machine-gun mount, a bomb bay or hard points for missiles under wings. However, too many countries have armed UAVs already and more will acquire them; these countries will not be convinced easily to give them up again.

Thus the only difference between remotely controlled and autonomous weapon systems would lie in the software. It is inconceivable that states would open their weapon-control software to international inspection, but even if they would, a new version with autonomous-attack function could be uploaded at any time. Thus, a-priori

verification of the inability to attack without human control is practically excluded. The only possibility is to check after the fact that every single attack had been controlled by a human.

To be able to prove this, each warring country would record the sensor and communication data of the remotely controlled combat vehicle and of the ground-control station, as well as the actions of the human operators therein, nationally in a black box. In parallel hash codes of these data would be recorded in a "glass box" and transmitted regularly to a treaty implementing organisation. The organisation would use the hash code to verify that the original data for selected attacks transmitted to it on a later (random or challenge) request are authentic and complete. Because the original data cannot be constructed from the hash code, an adversary could not gain a military advantage even if it had succeeded in getting access to it.

Details of such a method still need to be researched, developed and tested. It would be coupled with inspections and manoeuvre observations and would need more co-operation than customary up to now for inspections, such as under the CFE Treaty. It could provide the necessary transparency for reliable verification but also ensure the needed military secrecy. The effort would not be negligible, but already now similar data are recorded routinely nationally. The glass boxes and corresponding communication would have to be added.

7 Potential Contributions by the AI and Robotics Communities

In the further debate about AWS and about the need for and the possibility of an AWS prohibition, the AI and robotics communities can play an important role. All scientists, engineers and students can follow the national and international developments and discussions and make their voice known – privately as well as with public statements. Whoever has an opportunity to discuss with politicians can use it. Colleagues involved in teaching can include the subject into their curricula. In R&D one can be attentive to potential uses of one's own results for AWS developments.

Some AI or robotics researchers may have the option of devoting a part of their research to the prevention of AWS. Several themes are conceivable, some more interdisciplinary:

- Following general military R&D, dual-use R&D, general AI research with a bearing on AWS.
- Identifying problematic R&D activities, including dual-use research of concern.[13]
- Studying the reliability of machine learning for target recognition, including possible spoofing.
- Looking at the human-machine interface, in particular the concept of explainable AI, for automatic pre-selection of options for human operators.

[13] See the respective discussion in the life sciences (e.g. [49]) and the wider German Leopoldina-DFG "Joint Committee for the Handling of Security-Relevant Research" [48].

- How can international-security aspects be incorporated into ethics codes for responsible research, development and use of AI?
- What are the proliferation risks of component technologies?
- What could hobbyists achieve? What degree of AWS sophistication could non-state actors reach by themselves?
- Doing military-technology assessment of AWS in various scenarios.
- Modelling the interaction of two (or more) AWS fleets or AI-controlled battle-management systems.
- Are "circuit breakers" interrupting "flash wars" conceivable – automatic ones or human ones?
- Commonalities and differences between civilian autonomous vehicles and AWS?
- Verification schemes for an AWS ban; doing a test implementation of the secure recording with hash codes. Can blockchain play a role?
- How about societal verification in case of an AWS ban?
- What confidence and security building measures are possible? Can one develop a code of conduct for states?
- Commonalities and differences between AWS and cyber forces?
- Doing military-technology assessment of other potential military AI uses – what would be dangers, could they be limited preventively?

Results in any of these topics could be highly useful in strengthening the conviction that the introduction of AWS would be detrimental in many respects and that a prohibition would serve the interest of humankind best.

8 Conclusion – Preventive AWS Ban Urgently Needed

If developments with AWS will continue without constraints, three trains will race against each other. The US DoD has proclaimed its "Third Offset Strategy": Military-technological superiority is to be maintained with the help of the five building blocks "learning machines, human-machine collaboration, assisted human operations, human-machine combat teaming, and autonomous weapons" [37].[14] In Russia, Kalashnikov has built a "fully automated combat module" [38], the Kronstadt Group works on AI for military and civilian drone swarms [39], and "If the future went as defense experts are now predicting, Putin said, one day 'wars will be concluded when all the drones on one side are destroyed by the drones of another.'" [40] In China "[t]he PLA's [People's Liberation Army] initial thinking on AI in warfare has been influenced by careful analysis of U.S. military initiatives", in particular the US DoD Third Offset strategy. But "its approach could progressively diverge from that of the United States". "The PLA will likely leverage AI to enhance its future capabilities, including in intelligent and autonomous unmanned systems" [13].

[14] The Trump administration no longer mentions the offset strategy explicitly, but continues emphasising the need to maintain "decisive and sustained U.S. military advantages" or "overmatch" [56, p. 4] , [43, p. 28].

Beyond AWS AI is seen by all three main military countries as providing the means to lead the world or even rule it.[15] The three countries observe each other's military plans, literature and R&D activities intensely. As in the Cold War – where the situation with only two actors was much simpler – fears of falling behind are strong motives to proceed fast, enhanced by secrecy and worst-case assumptions. When one country would introduce AWS, the other two would fast follow up. The initiating act could, however, also come from a globally less relevant state. The world may have a window of only five to ten years to prevent an AWS arms race. Once it had begun in earnest, it would become extremely difficult to reverse, see the arguments presented above with respect to remotely controlled armed uninhabited vehicles.

Preventing the dangers of an AWS arms race with the destabilisation that would come with it requires the relevant states, first and foremost the USA, to recall the insight that national security can only be ensured sustainably by organising international security. For AWS this means an international prohibition. Because there are strong military, economic and political interests in AWS, achieving it needs intense pressure from an enlightened, critical public and support by benevolent states.[16] In this process, the international AI and robotics communities have started to play an important role, and should intensify their efforts.

References

1. Bhuta, N., Beck, S., Geiß, R., Liu, H.-Y., Kreß, C. (eds.): Autonomous Weapons Systems. Law, Ethics, Policy. Cambridge University Press, Cambridge (2016)
2. Scharre, P.: Army of None: Autonomous Weapons and the Future of War. Norton, New York (2018)
3. New America Foundation (2019). https://www.newamerica.org/in-depth/world-of-drones/3-who-has-what-countries-armed-drones. Accessed 16 July 2019
4. Wezeman, P.D., Fleurant, A., Kuimova, A., Tian, N., Wezeman, S.T.: Trends in international arms transfers, 2018, March 2019. https://www.sipri.org/sites/default/files/2019–03/fs_1903_at_2018.pdf. Accessed 16 July 2019
5. US Department of Defense: Autonomy in Weapon Systems (incorporating Change 1, May 8, 2017), 21 November 2012. http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf. Accessed 5 July 2019
6. International Committee of the Red Cross: Ethics and autonomous weapon systems: An ethical basis for human control?, 3 April 2018. https://www.icrc.org/en/download/file/69961/icrc_ethics_and_autonomous_weapon_systems_report_3_april_2018.pdf. Accessed 5 July 2019

[15] Russia: "Whoever becomes the leader in this sphere [AI] will become the ruler of the world." (Putin) [42] China: "[T]he PLA intends to 'seize the advantage in military competition and the initiative in future warfare,' seeking the capability to win in not only today's informatized warfare but also future intelligentized warfare, in which AI and related technologies will be a cornerstone of military power." [57, p. 13] The USA is more circumscript: "The Trump Administration's National Security Strategy recognizes the need to lead in artificial intelligence, and the Department of Defense is investing accordingly." [44].

[16] As in the case of the Anti-personnel Land Mine Convention (1997) by Canada and for the Cluster Munitions Convention (2008) by Norway.

7. Walpole, L.: The True Cost of Drone Warfare?, 8 June 2018. https://www.oxfordresearchgroup.org.uk/blog/the-true-cost-of-drone-warfare. Accessed 16 July 2019
8. Sauer, F., Schörnig, N.: Killer drones – the silver bullet of democratic warfare? Secur. Dialogue **43**(4), 353–370 (2012)
9. US Department of Defense: Unmanned Systems Roadmap 2007-2032 (2007). http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA475002
10. US Department of Defense: Unmanned Systems Integrated Roadmap FY2013-2038 (2013). http://www.dtic.mil/get-tr-doc/pdf?AD=ADA592015. Accessed 5 July 2019
11. US Department of Defense: Unmanned Systems Integrated Roadmap 2017-2042, 28 August 2018. http://cdn.defensedaily.com/wp-content/uploads/post_attachment/206477.pdf. Accessed 5 July 2019
12. Bendett, S.: Russia Is Poised to Surprise the US in Battlefield Robotics, 25 January 2018. https://www.defenseone.com/ideas/2018/01/russia-poised-surprise-us-battlefield-robotics/145439/. Accessed 8 July 2019
13. Kania, E.B.: Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power, 28 November 2017. https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power. Accessed 9 July 2019
14. Allen, G.C.: Understanding China's AI Strategy – Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security, 6 February 2019. https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy. Accessed 18 February 2019
15. Altmann, J.: Präventive Rüstungskontrolle. Die Friedens-Warte **83**(2–3), 105–126 (2008)
16. Altmann, J.: Nanotechnology and Preventive Arms Control (2005). https://bundesstiftung-friedensforschung.de/wp-content/uploads/2017/08/berichtaltmann.pdf. Accessed 16 July 2019
17. Altmann, J.: Arms control for armed uninhabited vehicles: an ethical issue. Ethics Inf. Technol. **15**(2), 137–152 (2013)
18. Altmann, J., Sauer, F.: Autonomous weapon systems. Survival **59**(5), 117–142 (2017)
19. Campaign to Stop Killer Robots (2019). https://www.stopkillerrobots.org/members/. Accessed 11 July 2019
20. Heyns, C.: Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, 9 April 2013. http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf. Accessed 11 July 2019
21. Campaign to Stop Killer Robots: Country Views on Killer Robots, 22 November 2018. https://www.stopkillerrobots.org/wp-content/uploads/2018/11/KRC_CountryViews22Nov2018.pdf. Accessed 12 July 2019
22. Computing experts from 37 countries call for ban on killer robots – Decision to apply violent force must not be delegated to machines, 15 October 2013. https://www.icrac.net/wp-content/uploads/2018/06/Scientist-Call_Press-Release.pdf. Accessed 12 July 2019
23. Autonomous Weapons: an Open Letter from AI & Robotics Researchers, 28 July 2015. https://futureoflife.org/open-letter-autonomous-weapons. Accessed 12 July 2019
24. The 30717 Open Letter Signatories Include (2019). http://futureoflife.org/awos-signatories/. Accessed 12 July 2019
25. Slaughterbots (Video 7:47), November 2017. https://www.youtube.com/watch?v=9CO6M2HsoIA. Accessed 12 July 2019
26. Future of Life Institute (2019). https://futureoflife.org/lethal-autonomous-weapons-pledge/. Accessed 12 July 2019
27. Gesellschaft für Informatik: Tödliche autonome Waffensysteme (LAWS) müssen völkerrechtlich geächtet werden, February 2019. https://gi.de/fileadmin/GI/Allgemein/PDF/GI-Stellungnahme_LAWS_2019-02.pdf. Accessed 12 July 2019

28. Bundesverband der Deutschen Industrie: Künstliche Intelligenz in Sicherheit und Verteidigung, January 2019. https://issuu.com/bdi-berlin/docs/20181205_position_bdi_ki. Accessed 12 July 2019
29. O'Sullivan, L.: I Quit My Job to Protest My Company's Work on Building Killer Robots. American Civil Liberties Union, 6 March 2019. https://www.aclu.org/blog/national-security/targeted-killing/i-quit-my-job-protest-my-companys-work-building-killer. Accessed 12 July 2019
30. Conger, K., Metz, C.: Tech Workers Now Want to Know: What Are We Building This For? New York Times, 7 October 2018. https://www.nytimes.com/2018/10/07/technology/tech-workers-ask-censorship-surveillance.html. Accessed 12 July 2019
31. UK Ministry of Defence: Unmanned Aircraft Systems, August 2017. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/640299/20170706_JDP_0-30.2_final_CM_web.pdf. Accessed 13 July 2019
32. Bundesministerium der Verteidigung, Pol II 5. Definitionsentwurf deutsch/englisch: Letales Autonomes Waffensystem. Personal communication (2014)
33. Germany: Statement delivered by Germany on Working Definition of LAWS/Definition of Systems under Consideration, April 2018. https://www.unog.ch/80256EDD006B8954/(httpAssets)/2440CD1922B86091C12582720057898F/%24file/2018_LAWS6a_Germany.pdf. Accessed 13 July 2019
34. Maas, H. (Minister for Foreign Affairs of Germany): Speech at the general debate of the 73rd General Assembly of the United Nations, 28 September 2018. https://gadebate.un.org/sites/default/files/gastatements/73/de_en.pdf. Accessed 13 July 2019
35. Amoroso, D., Sauer, F., Sharkey, N., Suchman, L.: Autonomy in Weapon Systems – The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy, 23 May 2018. https://www.boell.de/sites/default/files/boell_autonomy-in-weapon-systems_v04_kommentierbar_1.pdf. Accessed 13 July 2019
36. Gubrud, M., Altmann, J.: Compliance Measures for an Autonomous Weapons Convention, May 2013. https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Altmann_Compliance-Measures-AWC_ICRAC-WP2.pdf. Accessed 12 July 2019
37. Work, R.: Deputy Secretary of Defense Speech, 14 December 2015. https://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum. Accessed 9 July 2019
38. Tucker, P.: Russian Weapons Maker To Build AI-Directed Guns, 14 July 2017. http://www.defenseone.com/technology/2017/07/russian-weapons-maker-build-ai-guns/139452/. Accessed 9 July 2019
39. TASS: Russia is developing artificial intelligence for military and civilian drones, 15 May 2017. http://tass.com/defense/945950. Accessed 9 July 2019
40. Sharkov, D.: Vladimir Putin Talks Ruling the World, Future Wars And Life On Mars, 1 September 2017. https://www.newsweek.com/vladimir-putin-talks-ruling-world-future-wars-and-life-mars-658579. Accessed 9 July 2019
41. Arkin, R.C.: Governing Lethal Behavior in Autonomous Robots. Chapman&Hall/CRC, Boca Raton (2009)
42. Russia Today: 'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day, 1 September 2017. https://www.rt.com/news/401731-ai-rule-world-putin/. Accessed 29 November 2017
43. President of the USA: National Security Strategy of the United States of America, December 2017. https://www.whitehouse.gov/wp-content/uploads/2017/12/NSS-Final-12-18-2017-0905.pdf. Accessed 10 July 2019

44. White House: Artificial Intelligence for the American People, 10 May 2018. https://www.whitehouse.gov/briefings-statements/artificial-intelligence-american-people/. Accessed 10 July 2019
45. Protocol on Blinding Laser Weapons (Protocol IV), 13 October 1995. https://www.unog.ch/80256EDD006B8954/(httpAssets)/8463F2782F711A13C12571DE005BCF1A/$file/PROTOCOL+IV.pdf. Accessed 11 July 2019
46. Organization for Security and Co-operation in Europe: Treaty on Conventional Armed Forces in Europe, 19 November 1990. http://www.osce.org/library/14087. Accessed 11 July 2019
47. Drapela, P.: Fake news? Lethal effect of micro drones, 11 April 2018. https://www.ar.admin.ch/en/armasuisse-wissenschaft-und-technologie-w-t/home.detail.news.html/ar-internet/news-2018/news-w-t/lethalmicrodrones.html. Accessed 12 July 2019
48. Scientific Freedom and Scientific Responsibility (2019). https://www.leopoldina.org/en/about-us/cooperations/joint-committee-on-dual-use/. Accessed 12 July 2019
49. World Health Organization: Dual Use Research of Concern (DURC) (2019). https://www.who.int/csr/durc/en/. Accessed 12 July 2019
50. Sharkey, N.E.: The evitability of autonomous robot warfare. Int. Rev. Red Cross **94**, 787–799 (2012)
51. United Nations Office at Geneva: High Contracting Parties and Signatories (2019). https://www.unog.ch/80256EE600585943/(httpPages)/3CE7CFC0AA4A7548C12571C00039CB0C?OpenDocument. Accessed 11 July 2019
52. Members (2019). https://www.icrac.net/members/. Accessed 11 July 2019
53. Report of the Defense Science Board Task Force on Patriot System Performance, January 2005. https://www.acq.osd.mil/dsb/reports/2000s/ADA435837.pdf. Accessed 16 July 2019
54. Sharkey, N.: Staying in the loop. Human supervisory control of weapons. In: Bhuta, N., Beck, S., Geiß, R., Liu, H., Kreß, C. (eds.) Autonomous Weapons Systems. Law, Ethics, Policy, pp. 23–28. Cambridge University Press, Cambridge (2016)
55. US Air Force: Autonomous Horizons – System Autonomy in the Air Force – A Path to the Future, Volume I, Human-Autonomy Teaming, AF/ST TR 15-01. United States Air Force, Office of the Chief Scientist, June 2015. http://www.af.mil/Portals/1/documents/SECAF/AutonomousHorizons.pdf. Accessed 22 July 2019
56. US Department of Defense: Summary of the 2018 National Defense Strategy of the United States of America – Sharpening the American Military's Competitive Edge (2018). https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf. Accessed 10 July 2019
57. He, L. (vice president of the PLA's Academy of Military Science): Establish a Modern Military Theory System with Chinese Characteristics. Study Times, 19 June 2017. Cited by Kania, E.B., Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power, 28 November 2017. https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power. Accessed 9 July 2019