Exploring the 2023 U.S. Directive on Autonomy in Weapon Systems: Key Advancements and Potential Implications for International Discussions

Lutiana Valadares Fernandes Barbosa

Abstract: Despite international efforts, no specific regulation exists on Autonomous Weapons Systems (AWS) use in armed conflicts. The Department of Defense (DoD) directives on AWS are essential within the United States and impact international discussions. In 2023, the DoD reviewed the definition of AWS and semi-AWS, replacing the word "human operator" with "operator". We critically present the revision's primary shifts, pushbacks, and good practices.

Keywords: Autonomous Weapons Systems; Directive 3000.09; weapon review; International Humanitarian Law.

Explorando a Diretiva dos Estados Unidos de 2023 sobre Autonomia em Sistemas de Armas: principais avanços e implicações potenciais para discussões internacionais

Resumo: Apesar dos esforços internacionais, não existe regulamentação específica sobre o uso de Sistemas de Armas Autônomos (AWS) em conflitos armados. As diretrizes do Departamento de Defesa (DoD) sobre AWS são essenciais nos Estados Unidos e impactam as discussões internacionais. Em 2023, o DoD revisou a definição de AWS e semi-AWS, substituindo a palavra "operador humano" por "operador". Apresentamos de forma crítica as principais mudanças, retrocessos e boas práticas da revisão.

Palavras-chave: sistemas autônomos de armas; Diretiva 3000.09; revisão de armas; Direito Humanitário Internacional.

I n 2012, the United States published the Department of Defense (DoD) Directive 3000.09, *Autonomy in Weapon Systems* (United States Department of Defense 2012). By then, the debate over Autonomous Weapons Systems (AWS) was at a very initial stage. Three years later, the topic began to be informally discussed at the United Nations (UN) and, in 2017, formally discussed by the Group of Governmental Experts (GGE) on AWS under the auspices of the Convention on Certain Conventional Weapons (CCW) (UNODA 2023).

Throughout the last decade, AWS, which initially resembled characters of science fiction movies, has been used in the international scenario. In 2021, an expert panel report addressed to the UN Security Council acknowledged the deployment of the attack drone AWS STM Kargu-2 (Kargu 2022) in Libya in 2020 and that other AWS might have been used in the ongoing Russian–Ukraine conflict (Kallenborn 2022). The report states: "The lethal autonomous weapons systems were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true 'fire, forget and find' capability" (UNSC 2021, 17).

Despite diplomatic efforts and achievements through the excellent work of the Brazilian Ambassador Flavio Soares Damico, Chair of the 2021-2023 CCW GGE, and predecessors, diplomatic pace is much slower than technological development. While AWS remains without specific regulation in the international arena, few States – such as the U.S. and the United Kingdom (United Kingdom Ministry of Defense 2017) – have their directives on AWS or have made them publicly available. The 2012 U.S. Directive 3000.09 has not only expressed the U.S. position on the debate but also impacted the international discussions, as it was the first State directive on AWS (Insinna & Mehta 2022) and reflected in the U.S. delegation statements at the CCW GGE (United States Delegation 2018, 2). Its concept, for instance, was embraced by international NGOs (Horowitz 2016, 85).

Published in 2012 before the debate at the UN began, Directive 3000.09 has highly influenced international discussions. Following the technological developments since then, in 2023, the DoD published a new Directive 3000.09

Lutiana Valadares Fernandes Barbosa *has been a federal public defender since 2010. She holds a PhD in International Law from Minas Gerais Federal University (UFMG) and a Master's in Law from Columbia University and Pontifical Catholic University of Minas Gerais (PUC-MG). Member of the Ethics Research Group, Human Rights and Artificial Intelligence at the National School of Public Defender's Office (ENADPU). UNESCO researcher on the implementation of the recommendation on AI ethics in Brazil.*

with the potential to impact the ongoing international debate, which takes place mainly under the auspices of the CCW. The novel definition, for instance, is the one contained in the delegations from Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States, as well as Draft articles on AWS prohibitions and other regulatory measures based on International Humanitarian Law (IHL), submitted to the CCW GGE in 2023 (Australia et al. 2023).

The present article analyzes the main novelties of the 2023 Directive 3000.09 and their possible impacts on the international discussion on AWS: substituting the word "human operator" for "operator" in the definition of AWS and semi-AWS; adding a restriction on the requirement of appropriate levels of human judgment and elucidating the definition of failure; adding new requirements for AWS review and deployment; introducing concepts such as transparency, auditability, and explainability; and establishing the AWS Working Group. After critically presenting the primary shifts, this article discusses the good practices and pushbacks to the international community in the final considerations.

Published in 2012 before the debate at the United Nations began, Directive 3000.09 has highly influenced international discussions. Following the technological developments since then, in 2023, the Department of Defense published a new Directive 3000.09 with the potential to impact the ongoing international debate, which takes place mainly under the auspices of the Convention on Certain Conventional Weapons.

Less than one month after the 2023 Directive was issued, the U.S. launched a Political Declaration at the Responsible AI in the Military Domain (REAIM) Conference in the Hague (United States Department of State, 2023). The political declaration enunciates what the U.S. envisions as best practices and shared values on responsible use of AI in the military domain and calls on States to adhere to it. Therefore, considering the pros and cons of the 2023 DoD Directive 3000.09 innovations, we will do so in light of the U.S. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.

THE NEW DEFINITION OF AWS

Until now, there was no internationally agreed definition of AWS, and the 2012 DoD Directive 3000.09 Definition has been widely used in the international debate (Davison 2017). The 2012 DoD Directive defined AWS as:

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 (United States Department of Defense 2012).

The 2023 DoD Directive maintains the structure of the former definition but excludes reference to humans. It defines AWS as:

A weapon system that, once activated, can select and engage targets without further intervention by an operator. This includes, <u>but is not limited to, operator-supervised</u> autonomous weapon systems that are designed to allow <u>operators</u> to override the operation of the weapon system, but can select and engage targets without further <u>operator</u> input after activation (United States Department of Defense 2023a. Our emphasis.)

As observed, the 2023 Directive opted to delete the term "human" from the definition and substitute it for "operator". Despite Scharre's claim that the "core of the definition of an autonomous weapon has *not* changed" (Scharre 2023, our emphasis), this paper claims that, in the context of the development of autonomous technologies, the removal of the word "human" changes the definition. In this sense, Horowitz, the director of the Emerging Capabilities Policy Office at the Pentagon Institute for Security and Technology (IST), has affirmed that "every change we made to the directive was a response to a question or sometimes multiple questions that we got about the original directive" (Institute for Security and Technology 2023).

The draft articles on AWS presented by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States at the meeting of Governmental Experts on Lethal Autonomous Weapons in March 2023 included the novel U.S. DoD definition (Australia et al. 2023).

The language change means an option to accept, or at least open the doors for future acceptance, that non-human operators also perform the actions stated in the Directive. Annex III of the Draft articles on AWS, released after the 2023 Directive, corroborates this interpretation. There is a reference to a human operator, indicating, *a contrario sensu*, that the operator could be a non-human.

6. The IHL requirements and principles including inter alia distinction, proportionality, and precautions in attack must be applied through a chain of responsible command and control by the human operators and commanders who use weapons systems based on emerging technologies in the area of lethal autonomous weapons systems (2019 Report 17d apud Australia et al. 2023).

The 2012 Directive's "human operator" language also conveys that an operator could be or not be a human. In the same sense, the DoD Dictionary of Military and Associated Terms, as of November 2021, defines: "unmanned aircraft — An aircraft that does not carry a human operator and is capable of flight with or without human remote control. Also called UA (JP 3-30)" (United States Office of the Chairman of the Joint Chiefs of Staff 2021). Therefore, an unmanned aircraft could carry a non-human operator.

The novel definition has the side effect of assenting, with a broader distance, between human action and AWS deployment, as a human operator could activate a non-human operator that, in turn, activates the AWS.

The 2023 Directive's glossary defines an operator as "A person who operates a platform or weapon system". The glossary refers not to "human," but to "person". Our interpretation is that the term "person" embraces individuals, natural and legal entities, as defined in legal dictionaries:

Person. 1. A human being. 2. An entity (such as a corporation) that is recognized by the law as having the rights and duties of a human being (Garner 1999, 1162).

In the same sense, the U.S. Code (18 U.S.C. § 2510 (6)), in the chapter dealing with "Wire and Electronic Communications Interception and Interception of Oral Communications", defines "person" as "any employee, or agent of the United States or any State or political subdivision thereof, and any individual, partnership, association, joint stock company, trust, or corporation". Therefore, there is room for interpretation that agents under the 2023 Directive could be legal persons.

As U.S. law currently stands, artificial intelligence (AI) does not bear legal personality, but legal persons can adopt decision-making processes through AI. Despite Scharre stating there is "No need to worry about bots controlling bots!" (Scharre 2023, 6), the Directive updates ensure U.S. global leadership on AWS considering the technological advances (United States Department of Defense 2023b). In 2012, algorithmic impacts on the military domain appeared distant and futuristic (Insinna & Mehta 2022). Considering the actual AI developments and the language changes, there is room for interpretation that this might indeed mean, in the future, "bots controlling bots".

The increased distance from the human element means that under the definition of the 2023 Directive, a human agent might turn on a legal person's decision device that operates through AI, which activates the AWS. Under the 2023 Directive, commanders and operators are still required to exercise "appropriate levels of human judgment over the use of force," as discussed in the next section. However, this human judgment could be of the human that activates a legal person's AI decision-making, for example. Human judgment is required over the use of force and not over the AWS.

We acknowledge the possible view that broadening the scope of AWS definition to those legal persons who can activate does not mean a lower level of human involvement, as the new Directive states that weapon systems activated by whatever persons are considered AWS. However, we disagree with this viewpoint since the novelty implies recognizing the acceptance of AWS activated by non-human operators.

This innovation goes contrary to the claims by International Human Rights Law and International Humanitarian Law organizations such as the

> *There is room for interpretation that agents under the 2023 Directive could be legal persons. As U.S. law currently stands, artificial intelligence (AI) does not bear legal personality, but legal persons can adopt decision-making processes through AI. (...) Considering the actual AI developments and the language changes, there is room for interpretation that this might indeed mean, in the future, "bots controlling bots".*

International Committee of the Red Cross (ICRC 2019), Human Rights Watch (Wareham 2021,1), Stop Killer Robots (Docherty, 2019), and some States that claim for a higher level of human involvement in the process. We also observe that the 2012 Directive comprised fifteen pages and used the word "human" fifteen times, whereas the 2023 Directive used the word "human" twelve times in its twenty-four pages.

RESTRICTIONS ON THE REQUIREMENT OF APPROPRIATE LEVELS OF HUMAN JUDGMENT

Human-machine interaction is a pressing issue in the international discussions at CCW GGE on AWS. While States seem to agree that some human-machine interaction is necessary (ICRC 2019), States broadly diverge on how it should be. Some States claim for meaningful human control (Acheson 2021), which requires a higher degree of human involvement. Other States, such as Israel, Russia, and the U.S., refuse this concept (Acheson 2021, 13). The U.S. advocates for appropriate levels of human judgment:

> *"Appropriate" is a flexible term that reflects the fact that there is not a fixed, one-size-fits-all level of human judgment that should be applied to every context. What is "appropriate" can differ across weapon systems, domains of warfare, types of warfare, operational contexts, and even across different functions in a weapon system. Some functions might be better performed by a computer than a human being, while other functions should be performed by humans (United States Delegation 2018, 2).*

Therefore, appropriate levels of human judgment do not "require manual human control of the weapon system (…) but rather broader human involvement in decisions about how, when, where, and why the weapon will be employed" (Congressional Research Service 2022). Humans deploy AWS to the best of their knowledge. They must not exercise what some interpret as the higher control threshold, as "control is more likely to ensure that humans have the power to reverse a machine's decision on a particular attack" (Human Rights Watch 2016). The 2023 Directive maintains the 2012 requirement of AWS design, training, and testing "to allow commanders and operators to exercise appropriate levels of human judgment over the use of force" (United States Department of Defense 2023a, 3, 6, 10, 11, 15).

In the U.S. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, the U.S. also calls for the appropriate levels of human judgment standards by stating:

E. States should ensure that relevant personnel exercise appropriate care, including appropriate levels of human judgment, in the development, deployment, and use of military AI capabilities, including weapon systems incorporating such capabilities (United States Department of State 2023).

It only calls on human control regarding nuclear weapons:

B. States should maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment (United States Department of State 2023).

International Human Rights Law and International Humanitarian Law organizations such as ICRC (ICRC 2019), Human Rights Watch (Wareham 2021, 1), and Stop Killer Robots oppose the requirement of appropriate levels of human judgment.

Weapons systems that select and engage targets without meaningful human control—known as fully autonomous weapons, lethal autonomous weapons systems, or killer robots—would cross the threshold of acceptability and should be prevented and prohibited through new international law (Docherty 2019).

States such as the UK (United Kingdom Delegation 2018), France (French Delegation 2020), China (China's Delegation 2018), and Korea (Acheson 2021) also claim for meaningful human control.

The Directive innovation is that in the context of deciding formally on the development of AWS, the new Directive limits the scope of appropriate levels of human judgment by adding the phrase "in the envisioned planning and employment processes for the weapon" (United States Department of Defense 2023a, 15).

This new language might require the evaluators of an AWS proposal to declare how they envision AWS use, which could better accomplish the legal review requirement, which is similar to Scharre's interpretation:

This is new. As is the case with other weapons, how autonomous weapons are used can have a significant impact on their safety and even their lawfulness. The weapon should be considered in the context of an intended use (Scharre 2023).

Nonetheless, this better fulfillment of the weapon review requirement is restricted to the anticipated uses. Our perspective is that the new Directive represents a limitation as, before a formal decision on the development of AWS, the requirement is the possibility of commanders and deployers to exercise appropriate levels of human judgment restricted to the envisioned planning and employment of the AWS, and not to all possible foreseen uses of the AWS. This new provision might diminish the requirement of human-machine interaction as, once the AWS is developed, it might be used in contexts different from those envisioned and planned.

AN ELUCIDATING STATEMENT ON THE DEFINITION OF FAILURE

The new Directive added to its definition of failure the following sentence that elucidates what it considers minimizing the probability and consequences of failure: "means reducing the probability and consequences of unintended engagements to acceptable levels while meeting mission objectives and does not mean achieving the lowest possible level of risk by never engaging targets" (United States Department of Defense 2023a).

We acknowledge the possible interpretation that the new sentence restricts what is considered a failure and accepts some risks and consequences of unintended engagement as long as they are within acceptable levels. IHL principle of precaution requires all feasible efforts and not just reducing probabilities to acceptable standards.

Nonetheless, this paper claims that the new provision sets a higher bar than the 2012 Directive, considering that under IHL the general rule is that honest and reasonable accidents are deemed lawful (Milanovic 2020). We argue that the novel Directive restricts the types of accidents that are permissible, namely only those within acceptable levels. Furthermore, the Directive acknowledges that zero failure is unfeasible. Minimizing failures does not mean eliminating them, which is also not required by IHL's principle of precaution in the context of other weapons.

The principle of precaution is a milestone of International Humanitarian Law stated in Article 57 of Additional Protocol I to the Geneva Conventions of 1949. Although not ratified by the United States, it is also a customary International Humanitarian Law rule binding on the country.

According to rule 15 of the ICRC Customary International Humanitarian Law database:

In the conduct of military operations, constant care must be taken to spare the civilian population, civilians and civilian objects. All feasible precautions must be taken to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians, and damage to civilian objects (ICRC n.d.).

Under customary International Humanitarian Law, States must take all feasible precautions to avoid or minimize casualties. Adding content to the feasible precautions in the context of AWS, the 2023 Directive foresees that minimizing failures "means reducing the probability and consequences of unintended engagements to acceptable levels" (United States Department of Defense 2023). If properly implemented, the new Directive should aid the IHL principle of precaution as it adds content.

In line with the 2023 Directive, the U.S. issued in February 2023 the U.S. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, which affirms that military use of AI must follow International Humanitarian Law (United States Department of State 2023):

A principled approach to the military use of AI should include careful consideration of risks and benefits, and it should also minimize unintended bias and accidents (United States Department of State 2023).

The novel Directive seems to have positively influenced the Draft articles on autonomous weapon systems, submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States at the CCW GGE on March 13, 2023, as it states:

Feasible precautions must be taken in planning and conducting attacks to spare, as far as possible, civilians and civilian objects from the loss of life, injury, and damage or

destruction. Feasible precautions are those that are practicable or practically possible, taking into account all circumstances ruling at the time, including humanitarian and military consideration (Australia et al. 2023).

The elucidation on the application of international law as regards failure, brought by the new Directive, might influence other States members of the CCW GGE and add substance to the obligation of precaution in the international context of AWS.

WEAPON REVIEW

The 2023 Directive has innovated in some aspects of its guidelines for review of some AWS, as will be exposed in the following subheadings, mainly as regards a geographic restriction and the review of AWS variants.

The international obligation to review weapons is in Article 36 of Additional Protocol I of 1977 (ICRC 1977). According to it, States must, in "the study, development, acquisition or adoption of a new weapon, means and methods of warfare", assess if its use would "in some or all circumstances be prohibited by international law". However, the U.S. is not a party to this Protocol and does not recognize Article 36 as a customary rule of IHL. The U.S. reviews weapons on policy grounds (Dunlap 2016, 65), assessing "(1) whether the weapon's intended use is calculated to cause superfluous injury; (2) whether the weapon is inherently indiscriminate; and (3) whether the weapon falls within a class of weapons that has been specifically prohibited" (United States Department of Defense, 2015).

Even for States party to the Additional Protocol I, or that recognize the customary obligation of reviewing weapons, the review procedures and standards vary enormously, as they depend on national legislation or directives to assess whether a weapon is legal (Pilloud et al. 1987, 398). Therefore, a weakness of the institute of weapon review is the need for an internationally agreed framework (Amoroso 2020, 253).

AWS brings additional challenges to weapon review due to their autonomous feature and the possibility of in-field machine learning (Crootof 2018, 64). In 2018, the CCW GGE unanimously adopted guiding principles on AWS, three of which are weapon review-related. The principle "d" affirms "(…) in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law". Principle "e" affirms that

"States who are acquiring or developing AWS must consider (…) physical security, appropriate non-physical safeguards (including cybersecurity against hacking or data spoofing), the risk of acquisition by terrorist groups and the risk of proliferation (…)". Finally, principle "f" states, "Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems" (Group of Governmental Experts 2018).

Despite affirming the relevance of weapon review, the guiding principles offer rather vague guidelines. At the Political Declaration at the Responsible AI in the Military Domain, launched at the (REAIM) Conference in the Hague, the U.S. called on States to "(…) take effective steps, such as legal reviews, to ensure that their military AI capabilities will only be used consistent with their respective obligations under international law, in particular international humanitarian law" (United States Department of State 2023). It also recalls the importance of States undertaking rigorous testing and that "Self-learning or continuously updating military AI capabilities should also be subject to a monitoring process to ensure that critical safety features have not been degraded" (United States Department of State 2023).

The novel DoD Directive offers much more granularity and, therefore, not only enunciates U.S. perspective but also influences the international debate, which requires at least the sharing of good practices on this challenging issue.

The novel DoD Directive offers much more granularity and, therefore, not only enunciates U.S. perspective but also influences the international debate, which requires at least the sharing of good practices on this challenging issue.

Weapon Review and Geographic Area and Other Relevant Environmental Constraints

Among the guidelines to review AWS, the 2023 Directive states that competent organs shall verify that the AWS are "designed to complete engagements within a timeframe and geographic area, as well as other relevant environmental and operational constraints, consistent with commander and operator intentions". It further states that if the AWS are "unable to do so, the systems will terminate the engagement or obtain additional operator input before continuing the engagement" (United States Department of Defense 2023).

The 2012 directive required a timeframe limitation (United States Department of Defense 2012) but not a geographic limitation and other relevant environment constraints, representing a positive innovation that adds the likelihood of precision of the AWS, reducing the risk of civilian harms. According to Scharre (2023, 6), "This paragraph means that the DoD cannot field autonomous weapons that are unbounded in time and geography".

This good practice should influence the international community while reviewing AWS. Nonetheless, this same provision of the 2023 Directive demonstrates a trend to lose the bonds between humans and AWS, by removing the term "human". The new guideline to review the AWS requires that, if the system is unable to complete the engagement within the timeframe and geographical limitation, it shall ask for "additional operator input" and no longer "additional human operator input" (United States Department of Defense 2012). As argued in Section II, this operator could be a legal person acting through AI.

Weapon Review and AWS Variants

The 2023 Directive importantly and expressly requires a new review for AWS, that is a variant of former approved AWS, if there are changes in the algorithm, or intended mission set, the operational environment, target sets, and expected adversarial countermeasures (United States Department of Defense 2023). The former Directive had no such provision. The new requirement of review is highly relevant to ensure compliance with IHL if changes occur. Often, AWS encompass machine learning, and its algorithms might learn from experience. In those cases, a new review is crucial. In addition, considering that autonomous devices select and engage targets without the necessity of further intervention by a human being, they bear an inherent component of unpredictability, which might increase to unacceptable levels if the operational environment, mission or target sets, or expected adversarial behavior changes. The new provision adds safety, reliability, and compliance with IHL and IHRL right to life, a good practice that should positively influence other States.

Transparency, auditability, and explainability

A significant shift brought by the 2023 Directive is to incorporate the concepts of transparency, auditability, and explainability, which are on the cutting-edge discussions in the field of autonomous machines and artificial intelligence, as well as at the helm of AWS.

Transparency "refers to the extent to which the system discloses criteria of its functioning. (…) The metaphor for transparency in this sense is the 'why-did-you-do-that?' button: the systems must disclose the criteria, sources, and process (…)" (Spagnolli et al. 2018, 1). It is very relevant to discover how and why failures occur, make AWS understandable to operators, and make accountability possible (Winfield et al. 2021). Although there are different definitions of explainability, we consider it a subset of transparency, meaning transparency is accessible to non-experts (IEEE 2020).

In this regard, on a policy level, the 2023 DoD Directive requires that AWS hardware and software be designed with "(c) technologies and data sources that are transparent to, auditable by, and explainable by relevant personnel". The new Directive also innovated in dealing expressly with AI technologies. It required that the "design, development, deployment, and use of AI capabilities in autonomous and semiautonomous weapon systems will be consistent with the DoD AI Ethical Principles and the DoD Responsible Artificial Intelligence Strategy and Implementation Pathway (…)". In this regard, it also requires, among other things, that IA is traceable:

> *The DoD's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedures and documentation (United States Department of Defense 2023a).*

Transparency, auditability, and explainability are crucial to prevent casualties, improving systems, and ensure accountability when breaches occur. In this regard, the 2023 Directive added protection to those who AWS misdoings, especially civilians, might negatively impact. The U.S. political declarations also called on States to "ensure that military AI capabilities are developed with auditable methodologies, data sources, design procedures, and documentation" (United States Department of State 2023). The Directive should positively influence U.S. policies and be a good practice that might positively impact the international debate.

AUTONOMOUS WEAPONS SYSTEM WORKING GROUP

An important innovation was the creation of an Autonomous Weapons System Working Group composed of Federal Government employees or Service members on active duty, which supports U.S. State organs to consider the DoD interests for

AWS review before formal development and fielding. The Autonomous Weapons System Working Group does not have decision-making power "but is tasked with supporting the decision-makers during the review process" (Paul Scharre 2023). It advises decision-makers if AWS require senior-level approval according to the 2023 Directive and how to identify and address issues during senior-level review (United States Department of Defense 2023a, 19).

A working group on AWS is a good practice that will foster the observance of the 2023 Directive. However, as established in the 2023 Directive, the working group is restricted to federal government employees and service members on active duty and closed to civil society and academia participation, which could potentially add relevant insights to the AWS debate. It acknowledged that AWS are within the States' security interests and are State secrets. However, the Directive could have opened to some civil society and academia participation by foreseen calls for written contribution or foreseen public hearings in which civil society and academia can voice their perspectives.

An important innovation was the creation of an Autonomous Weapons System Working Group composed of Federal Government employees or Service members on active duty, which supports U.S. State organs to consider the DoD interests for AWS review before formal development and fielding.

FINAL CONSIDERATIONS

The 2023 DoD Directive 3000.09 innovations represent both pushbacks and good practices to the international debate. On the downside, the new Directive seems to be sailing in a direction that agrees with further distancing the human element from the deployment of AWS, opposed by International Human Rights Law and International Humanitarian Law organizations such as ICRC (ICRC 2019), Human Rights Watch (Wareham 2021,1), Stop Killer Robots (Docherty 2019), and some States such as France (French delegation 2020), that claim for meaningful human control.

The new Directive definition of AWS excludes the word "human" and substitutes it with "operator," which is not necessarily a human. The Draft articles on AWS reflect the novel definition submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States, meaning the other five States already adhered to it (Australia et al. 2023).

Also, in the helm of weapon review, the new Directive requires that, if the system is unable to complete the engagement within the timeframe and geographical limitation, it shall ask for "additional operator input" (United States Department of Defense 2023a.) substituting the word "human" to "operator" again.

The Directive also maintains the requirement of appropriate levels of human judgment, which, despite requiring some human involvement, is a lower threshold of human-machine interaction compared to other States' claim for meaningful human control, for example (United States Department of State 2023). It also encompasses the negative innovation of limiting the scope of appropriate levels of human judgment by adding the phrase "in the envisioned planning and employment processes for the weapon" (United States Department of Defense 2023a.), meaning that not envisioned or unplanned uses might remain without appropriate levels of human judgment.

Those pushbacks lean towards U.S. military interests rather than ensuring higher standards of protection of Human Rights and International Humanitarian Law and, hopefully, should not be further influencing the international debate. Despite some challenging aspects, the 2023 Directive is an example of good practice. It added content to the feasible precautions by stating that minimizing failures "means reducing the probability and consequences of unintended engagements to acceptable levels" (United States Department of Defense 2023a). Therefore, even accidents must be within acceptable levels.

Regarding weapon review, it foresees the necessity to verify that the AWS is "designed to complete engagements within a timeframe and geographic area, as well as other relevant environmental and operational constraints, consistent with commander and operator intentions" (United States Department of Defense 2023a). It also requires a new review in the case of AWS, a variant of former approved AWS, if there are changes in the algorithm or intended mission set, the operational environment, target sets, and expected adversarial countermeasures. The international community needs to share good practices regarding AWS review. The innovations brought by the 2023 DoD Directive 3000.09 have a great potential to reverberate positively in other States' AWS review processes.

Another positive innovation is AWS's transparency, explainability, and auditability requirements, aligned with the most recent research on autonomous devices. It will be relevant to reduce casualties, make it possible to understand why and who when failures occur, and ensure responsibility.

Creating an AWS working group is a good practice that aids compliance with the Directive. However, the DoD missed the opportunity to ensure

democratic participation in this working group by providing academic and civil society input venues.

Those good practices regarding elucidating the definition of failure, weapon review, requirements of transparency, explainability, and auditability, and creating an AWS working group enhance the protection of Human Rights and International Humanitarian Law. They also have the potential to influence the international community positively and add those topics to the focus of the discussions of the CCW GGE. ⬓

References

Acheson, Ray. 2021. "Editorial: Convergence Against Killer Robots". *CCW Report* 9 (3): 1-3. https://www.reachingcriticalwill. org/disarmament-fora/ccw/2021/laws/ ccwreport/15375-ccw-report-vol-9-no-3.

Anderson, James M. , Benoit Arbour, Roberta Arnold, Thomas Kadiofsky, Tom Keeley, Matthew R. MacLeod, Sean Bourdon, et al. 2015. *Autonomous Systems: Issues for Defense Policymakers*. NATO Supreme Allied Command Technical Report. https://apps.dtic.mil/sti/citations/AD1010077.

Australia, Canada, Japan, Poland, the Republic of Korea, the United Kingdom, and the United States. 2023. *Draft Articles on Autonomous Weapon Systems – Prohibitions and Other Regulatory Measures on the Basis of International Humanitarian Law ('IHL')*. CCW Working Paper 4 (2). https://docs-library.unoda.org/ Convention_on_Certain_Conventional_Weapons_- Group_of_Governmental_Experts_on_Lethal_ Autonomous_Weapons_Systems_(2023)/CCW_G

GE1_2023_WP.4_US_Rev2.pdf.

Boutin, Berenice. 2021. *Legal Questions Related to the Use of Autonomous Weapon Systems*. Asser Institute.

https://www.asser.nl/media/795707/boutin-legal- questions-related-to-the-use-of-aws.pdf.

Carter, Ashton B. 2012. "(DoD) Directive 3000.09, Autonomy in Weapon Systems". *Department of Defense Directive*.

CCW/GGE Chairperson. 2023. *Non-Exhaustive Compilation of Definitions and Characterizations*. CCW/GGE 1.

https://docs-library.unoda.org/Convention_ on_Certain_Conventional_Weapons_- Group_of_Governmental_Experts_on_Lethal_ Autonomous_Weapons_Systems_(2023)/CCW_G GE1_2023_CRP.1_0.pdf.

China's Delegation. 2018. *Position Paper*. CCW Working Paper 7.

https://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.1-WP7.pdf.

Crootof, Rebecca. 2014. "The Killer Robots Are Here: Legal and Policy Implications". *Cardozo Law Review 36 (1837): 1837-1915*. https://ssrn.com/abstract=2534567.

Crootof, Rebecca. 2018. "Autonomous Weapon Systems and the Limits of Analogy". *Harvard National Security Journal 51(9): 51-83*. https://doi.org/10.2139/ssrn.2820727.

Davison, Neil. 2017. "A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law". *UNODA Occasional Papers 30: 5-18*. https://doi.org/10.18356/29a571ba-en.

Docherty, Bonnie. 2019. "Elements of a Treaty on Fully Autonomous Weapons". Stop Killer Robots Briefing Paper.

https://www.stopkillerrobots.org/wp-content/uploads/2020/03/Key-Elements-of-a-Treaty-on-Fully-Autonomous-Weapons.pdf.

Dunlap, Charles James. 2016. "Accountability and Autonomous Weapons: Much Ado about Nothing?" *Temple International & Comparative Law Journal, Forthcoming: 1-16*. https://doi.org/10.2139/ssrn.2764528.

French Delegation. 2020. "Operationalization of the 11 Guiding Principles at National Level". *CCW 06 (08)*. https://documents.unoda.org/wp-content/uploads/2020/07/20200610-France.pdf.

French, Duncan & Tim Stephens. 2014. "International Law Association (ILA) Study Group on Due Diligence in International Law First Report". *Due Diligence*.

Garner, Brian Andrew. 1999. *Black's Law Dictionary*. Eagan: West Group.

General Counsel of the Department of Defense. 2016. *Department of Defense Law of War Manual*. https://dod.defense.gov/Portals/1/Documents/pubs/DoD%20Law%20of%20War%20Manual%20-%20June%202015%20Updated%20Dec%202016.pdf?ver=2016-12-13-172036-190.

Hicks, Kathleen H. 2023a. "(DoD) Directive 3000.09, Autonomy in Weapon Systems". *Office of the Under Secretary of Defense for Policy*. https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

Holland Michel, Arthur. 2020. "The Black Box, Unlocked: Predictability and Understandability in Military AI". *United Nations Institute for Disarmament Research*. https://unidir.org/publication/black-box-unlocked.

Horowitz, Michael C. 2016. "Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems". *Temple International & Comparative Law Journal 30(1): 85-98*. https://sites.temple.edu/ticlj/files/2017/02/30.1.Horowitz-TICLJ.pdf.

Horowitz, Michael. 2023. "Under Control: How Technology Is Shaping DoD's Approach to Autonomous Weapons". *Institute for Security and Technology YouTube video 44:25*. https://www.youtube.com/watch?v=rz19aVYU-Ns.

IEEE. 2020. "Draft Standard for Transparency of Autonomous Systems". *IEEEXplore: 1-70*.

Insinna, Valerie & Aaron Mehta. 2022. "Updated Autonomous Weapons Rules Coming for the Pentagon: Exclusive Details". *Breaking Defense, All Domain Connecting the Joint Force*. https://breakingdefense.com/2022/05/updated-autonomous-weapons-rules-coming-for-the-pentagon-exclusive-details.

International Committee of the Red Cross (ICRC). 2005. *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*. https://www.icrc.org/en/doc/assets/files/other/icrc_002_0811.pdf.

International Committee of the Red Cross (ICRC). 2010. "Protocol Additional to the Geneva Conventions of 12 August 1949". *ICRC Reference*. https://www.icrc.org/en/doc/assets/files/other/icrc_002_0321.pdf.

International Committee of the Red Cross (ICRC). 2019. "Statement of the International Committee of the Red Cross (ICRC)". *Convention on Certain Conventional Weapons*. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2019)/CCW%2BGGE%2BLAWS%2BICRC%2Bstatement%

2Bagenda%2Bitem%2B5e%2B27%2B03%2B2019.pdf.

International Committee of the Red Cross (ICRC). 2023. "Rule 15. Principle of Precautions in Attack". *International Humanitarian Law Database II (5A)* https://ihl-databases.icrc.org/en/customary-ihl/v1/rule15.

ISO/IEC CD 12792. 2023. Information Technology — Artificial Intelligence — *Transparency Taxonomy of AI Systems. (stage 30.60)* https://www.iso.org/standard/84111.html.

Kallenborn, Zachary. 2022. "Russia May Have Used a Killer Robot in Ukraine. Now What?" *Bulletin of the Atomic Scientists*.

https://thebulletin.org/2022/03/russia-may-have-used-a-killer-robot-in-ukraine-now-what/.

Koivurova, Timo & Krittika Singh. 2022. "Due Diligence". *Max Planck Encyclopedias of International Law [MPIL]*. https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e1034?rskey=Zer2AA&result=1&prd=OPIL.

Kwik, Jonathan & Tom Van Engers. 2021. "Algorithmic Fog of War: When Lack of Transparency Violates the Law of Armed Conflict". *Journal of Future Robot Life 2 (1-2): 43-66.* https://doi.org/10.3233/frl-200019.

Majumdar Roy Choudhury, Lipika. 2021. "Letter Dated 8 March 2021 from the Panel of Experts on Libya Established Pursuant to Resolution 1973 (2011) Addressed to the President of the Security Council". *UNSC S/229*. https://undocs.org/S/2021/229.

Milanovic, Marko. 2020. "Mistakes of Fact When Using Lethal Force in International Law: Part I". *Blog of the European Journal of International Law, EJIL: Talk!* https://www.ejiltalk.org/mistakes-of-fact-when-using-lethal-force-in-international-law-pa rt-i/.

n.d. 2011. *Free Dictionary of Law Terms and Legal Definitions*. Nolo Network. https://www.nolo.com/dictionary/p.

n.d. 2016. "Killer Robots and the Concept of Meaningful Human Control". *Memorandum to CCW Delegates, Human Rights Watch*.

https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control.

n.d. 2018. "Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems". *CCW/GGE 1(3)*. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/323/29/PDF/G1832329.pdf?Op enElement.

Office of the Chairman of the Joint Chiefs of Staff. 2021. *DOD Dictionary of Military and Associated Terms*. https://irp.fas.org/doddir/dod/dictionary.pdf.

Pilloud, Claude, Yves Sandoz, Christophe Swinarski, Bruno Zimmermann & International Committee of the Red Cross. 1987. *Commentary on the Additional Protocols of 8 June 1977*. Geneva: Martinus Nijhoff Publishers. https://hdl.loc.gov/loc.law/llmlp.Commentary_GC_Protocols.

Sayler, Kelley M. 2022. *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems*. IN FOCUS, Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11150.

Scharre, Paul. 2019. *Army of None: Autonomous Weapons and the Future of War*. New York: W.W. Norton & Company.

Scharre, Paul. 2023. "DoD Autonomous Weapons Policy". *CNAS Noteworthy*.

https://www.cnas.org/press/press-note/noteworthy-dod-autonomous-weapons-policy.

Spagnolli, Anna, Lily Frank, Pim Haselager & David Kirsh. 2018. "Transparency as an Ethical Safeguard". *Symbiotic 2017: Symbiotic Interactions: 1-6*. https://doi.org/10.1007/978-3-319-91593-7_1.

Stauffer, Brian. 2020. "Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control". New York, N.Y.: Human Rights Watch. https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and.

STM. 2016. "KARGU - Rotary Wing Attack Drone Loitering Munition System". *STM*. https://www.stm.com.tr/en/kargu-autonomous-tactical-multi-rotor-attack-uav.

Taddeo, Mariarosaria & Alexander Blanchard. 2022. "A Comparative Analysis of the Definitions of Autonomous Weapons Systems". *Science*

and Engineering Ethics 28 (37). https://doi.org/10.1007/s11948-022-00392-3.

Tuffi Saliba, Aziz & Lutiana Valadares Fernandes Barbosa. 2020. "Autonomous Weapons Systems and International Law: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains". *Revista de Direito Internacional 17 (3)*. https://doi.org/10.5102/rdi.v17i3.7550.

UNIDIR. 2014. "The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control Might Move the Discussion Forward". *UNIDIR Resources 2*. https://unidir.org/sites/default/files/publication/pdfs//considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf.

United Kingdom. 2018. "Human Machine Touchpoints: The United Kingdom's Perspective on Human Control over Weapon Development and Targeting Cycles". *CCW Working Paper 1*. https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.2-WP1.pdf.

United Kingdom Ministry of Defense. 2017. "Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems". *Development, Concepts and Doctrine Center*.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf.

United States. 2018. "Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems". *CCW Working Paper 4*.

https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.2-WP4.pdf.

United States Department of Defense. 2023b.

"DoD Announces Update to DoD Directive 3000.09, 'Autonomy in Weapon Systems'". *DOD Immediate Release*. https://www.defense.gov/News/Releases/Release/Article/3278076/dod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems/.

United States Department of State. 2023. "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy". *Bureau of Arms Control, Verification and Compliance*.

https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/.

United Nations. "Lethal Autonomous Weapons Systems (LAWS)". *Office for Disarmament Affairs*.

https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/.

Winfield, Alan F. T., Serena Booth, Louise A. Dennis, Takashi Egawa, Helen Hastie, Naomi Jacobs, Roderick I. Muttram, et al. 2021. "IEEE P7001: A Proposed Standard on Transparency". *Frontiers in Robotics and AI 8 (665729)*. https://doi.org/10.3389/frobt.2021.665729.