Human Interaction and Emerging Technologies (IHIET-FS 2025), Vol. 196, 2025, 12–22
https://doi.org/10.54941/ahfe1005950

Defining Autonomous Weapon Systems: A Conceptual Overview of Existing Definitory Attempts and the Effectiveness of Human Oversight

Mark Tsagas

University of East London (UEL), London, E15 1NF, United Kingdom

ABSTRACT

Regardless of whether the application of International Humanitarian Law (IHL) to emerging weapon technologies is explicitly or implicitly acknowledged (UNIDIR, 2023), certain legal obligations remain. In particular, adherence to the principles of Distinction, Proportionality, and Precaution in attack continues to be essential. Consequently, there is a need to conduct a thorough inquiry, which this paper seeks to fulfil. This includes determining what a value-neutral definition of Autonomous Weapon Systems (AWS) might entail, alongside a review of the effectiveness of human oversight as a safeguard. Not only does the human-element operate as an integral thematic component of the existing definitions but the value of its position as a pre-requisite for adherence needs to be ascertained.

Keywords: Autonomous weapon systems (AWS), Lethal autonomous weapon systems (LAWS), Automatic, Automated, Artificial intelligence, Human oversight, Generative artificial intelligence

INTRODUCTION

It is only "natural that proponents and opponents of Autonomous Weapon Systems (AWS) will seek to establish a definition that serves their aims and interests. The definitional discussion will not be a value-neutral discussion of facts, but ultimately one driven by political and strategic motivations" (UNIDIR, 2017). Though somewhat of a sullen statement, conceptually echoing the self-serving outcome-oriented nature of 'conflict' espoused by the famous proverb, 'all is fair in love and war' (attributed to the poet John Lyly), the aforementioned stipulation by the United Nations Institute for Disarmament Research (UNIDIR), coupled with the belief that some States are reluctant to engage in a broader definitional exercise, have recently been proven rather prophetic.

In effect, the concepts above have somewhat accurately described the existing state of the discourse surrounding the technology in question. Specifically, the divided stance on definitions can be evidenced in a wide array of publications. These range from and include the 2024 'collation of responses' report compiled by the Implementation Support Unit of the Convention on Certain Conventional Weapons (UNODA, 2024),

the UNODA 2023 report, compiled by the same organisation, focused specifically on individual definitions and characterisations from multiple countries, as well as results from domestic inquiries[1].

The two aforementioned documents shed light on the rationale behind this debate, reverberating the initial quotation's poignancy. The House of Lords Committee (2023) highlighted the UK's lack of an operational definition for Autonomous Weapon Systems and the regulatory challenges this poses. In response, the UK Government acknowledged the Committee's concerns but still declined to adopt an official definition. The reasoning? Definitions often underpin policymaking and could catalyse legal instruments restricting certain systems, thus posing risks to national defence interests (Ministry of Defence, 2024). This stance, albeit not unique to the UK, aptly demonstrates that the suggested apprehension regarding the adverse effects of adopting a wider definition, in so far as bringing the legitimacy or legality of encompassed technologies into question (UNIDIR, 2017), seems to have retained its prevalence.

CONCEPTUALISING AUTONOMOUS WEAPON SYSTEMS

In order to engage with the research questions posed, it is first necessary to briefly illustrate various key facets of the definitory process. While this segment does not propose to engage with the terms in abundant detail, due in no small part to the work of Mariarosaria Taddeo & Alexander Blanchard upon which the submission at hand is seeking to expand, it is nonetheless necessary to offer this contextual backdrop, as a potential means against which the effectiveness of any subsequent rewording can be critiqued. To this effect, it has been stipulated that, in theory, three principal approaches exist that may be utilised to generate an appropriate definition for AWS. Specifically:

Technology-Centric Approach: A more traditional approach with a lens towards providing a technical definition based on descriptions of the physical object focusing on its technical specifications and its intended operating environment.

Human-Centred Approach: This latter approach would seek to define AWS based on their relation to human users. In effect, grounded in legal commitments and norms this method would provide a 'common' language for discussion, with a continued lens on shared objective of maintaining control over such systems, thusly consistent with IHL and able to integrate consideration of ethics and human machine interaction.

Task/Functions Approach: In this instance, the focus is on the delegated functions that would classify a weapon as autonomous; described as the functionalist approach. Though it may be classified as overly inclusive, especially considering the nature of emerging technologies not attached to weapon platforms having shared functions (e.g. target selection), it is not reliant on any sort of technological development. It focuses predominantly

[1]Such as those arising from the House of Lords AI in Weapon Systems Committee Report and the subsequent UK Government response.

on particular functions and is typically sufficient to facilitate broad agreement with the premise (UNIDIR, 2017).

However, in order to truly capture the essence of AWS' current form while still retaining a forward-thinking stance, a blended approach may be warranted. In fact, the definition that is to follow demonstrates aspects from both the human-centred approach, referencing human control, and functionalist approach by demonstrating aspects related to lethality (exertion of kinetic force), identification, selection and attacking of targets and deployment. Although certain aspects of the definition will be subsequently challenged, it is its blended functional definitory approach to the analysis of Autonomous Weapon Systems that sets this particular attempt apart, allowing it to exist as a functional basis for adaptation.

A UNIFORM DEFINITION FOR AWS

As iterated in segments prior, despite the overarching fragmented nature of existing positions on this matter, a genuine attempt was made by both the Oxford Internet Institute and the Alan Turing Institute, working in tandem, to consolidate stances and consequently produce a viable definition. The results of their efforts, culminated into the following definitory statement:

"*An artificial agent which, at the very minimum, is able to change its own internal states to achieve a given goal, or set of goals, within its dynamic operating environment and without the direct intervention of another agent and may also be endowed with some abilities for changing its own transition rules without the intervention of another agent, and which is deployed with the purpose of exerting kinetic force against a physical entity (whether an object or a human being) and to this end is able to identify, select and attack the target without the intervention of another agent is an AWS. Once deployed, AWS can be operated with or without some forms of human control (in, on or out the loop). A lethal AWS is specific subset of an AWS with the goal of exerting kinetic force against human beings*" (Taddeo Mariarosaria & Blanchard Alexander).

As a matter of principle, it is necessary to illustrate that the definition at hand is in fact apt and there is nothing functionally wrong with the wording provided. Furthermore, it is also important to stipulate that the definition above was specifically chosen as the basis for this publication because the original authors have stipulated and evidenced through their analysis that it is indeed a value-neutral definition (ibid), as showcased above.

However, this submission takes the position that the scope of the definitory paragraph above is too narrow, thus excluding forms of autonomous weapon systems that do not adhere to the phrasing regarding 'deployment' and the 'exertion of kinetic force'. Specifically, in relation to the former point of 'deployment', it is submitted that the term in and of itself is better reserved for LAWS, rather than being attributed to AWS as a whole. While other definitions have qualified or at least implied the purpose of deployment to being largely confined to the conducting of physical operations, by virtue of the use of words and phrases such as 'weapons' and 'kinetic contexts' (ibid), the definition provided by NATO (2020), in so far as

describing autonomous systems as those that decide and act to "*accomplish desired goals, within defined parameters, based on acquired knowledge and an evolving situational awareness, following an optimal but potentially unpredictable course of action*" may be a more apt description of how AWS should be construed, in terms of their current application. In essence, while NATO's definition does not refer to AWS explicitly, it does recognise the existence of a wider set of autonomous technologies, which when combined with the suggested existence of multiple categories of military technologies, espoused by O'Hanlon (2020), demonstrates that material deployment may not necessarily be the correct descriptor of their utilisation; with 'employment' serving as a potential suitable replacement term.

In a similar vein, the attribution of the phrase 'exertion of kinetic force' to AWS may have unnecessarily limited the types of technologies that could be included under a wider definition. Specifically, the wording rejects the possible existence of multiple variants of AWS, albeit in an oblique fashion, when in fact a variety of unmanned deployable military technologies have already been documented with the goal of conducting reconnaissance (e.g. Uninhabited Aerial Vehicles (UAV's), juxtaposed against their specifically designed counterparts intended for combat (LAWS) through the exertion of kinetic force, by carrying warheads or other munitions (e.g. Uninhabited Combat Aerial Vehicles (UCAV's) (Sparrow, 2007). Reported recent events, lend additional credence to this claim. Specifically, in relation to presently ongoing conflicts, there has been a reported utilisation of a large amount of 'Unguided (dumb) Bombs'[2]. However, these bombing sprees have been alleged to have been fuelled by Artificial Intelligence-based targeting programmes, such as Lavender and Where's Daddy, advocating for the existence of layered components in such instances of modern warfare[3]. In essence, while an AWS was not the one to utilise force against its identified and acquired targets, it allegedly did play an active role in the process.

Consequently, it is also suggested that the distinction between AWS and LAWS has not been properly constituted, thusly benefitting from further clarification. In fact, while the concept of LAWS has been accurately described as a subset of AWS, designed with a specific purpose of deploying lethal force, the specific attribute (human) of those purported targets has been accepted too readily. In turn, this quick acceptance has resulted in the echoing of one of the principal issues resulting from some current and previously publicised definitory attempts, which conflated the terms AWS and LAWS to a certain extent (Boulanin, 2016). Rather, the concept of 'lethality' will be broadened and subsequently brought in line with more recent definitions that purport to recognise that not only should LAWS not cause loss to civilian life but also avoid damage to civilian objects (UNODA, 2023). Of course,

[2]See for example: Bertrand N and Lillis KB, "Exclusive: Nearly Half of the Israeli Munitions Dropped on Gaza Are Imprecise 'Dumb Bombs,' US Intelligence Assessment Finds" CNN (December 14, 2023) https://edition.cnn.com/2023/12/13/politics/intelligence-assessment-dumb-bombs-israel-gaza/index.html last accessed May 4, 2025.

[3]See for example: Iraqi A, "'Lavender': The AI Machine Directing Israel's Bombing Spree in Gaza" (+972 Magazine, April 3, 2024) https://www.972mag.com/lavender-ai-israeli-army-gaza/ last accessed May 4, 2025.

on this particular point it is worth noting that the outside expansion of the considerations regarding LAWS' 'demographic of targets' does primarily relate to IHL compliance and civilian casualties.

As a result, in order to extend the focus, this paper builds upon the aforementioned proposed wider definition by altering some of the phrasing in order to better reflect existing technologies utilised within security, defence and warfare contexts. Examples of such emerging technologies are 'Project Maven', and the aforementioned 'Where's Daddy' and 'Lavender', to reference but a few; Artificial Intelligence targeting programmes that conceptually do not fit under the moniker of the previous definition, since at the present time they cannot be deployed or independently exercise kinetic force. This approach effectively deals with the proposed framing issues regarding 'lethality' and the omittance of concerns about the development of "less than lethal" weapons that may not be intrinsically anti-personnel but rather anti-material weapons that cause death as secondary or a collateral effect (UNIDIR, 2017). Thusly, this submission's definitory attempts have resulted in the following rewording:

An artificial […] another agent, ***and which is employed with the purpose of identifying, selecting, targeting and/or attacking a target (whether an object or a human being), in a military, security and/or defence context,*** *without the intervention of another agent is an AWS. Once* ***employed****, AWS can be operated with or without some forms of human control (in, on or out the loop).* **Lethal Autonomous Weapon Systems (LAWS) are a specific subset of AWS, capable of being deployed, with the goal of exerting kinetic force or causing material damage against physical targets (whether an object or a human being).**

In the first instance of this reworded definition, it becomes apparent that the term 'deployed' has been replaced with 'employed', the reasoning for which has been explored above. Yet, in addition to this subtle alteration, the concept of 'operational context' was subsequently added in order to limit the breadth of technologies covered under the definition, confining them specifically to instances of warfare, security and defence. The purpose of such a distinction, while perhaps not immediately necessary at the present time, may prove to be of future importance due to the rapidly evolving state of Artificial Intelligence, its prevalence in the public's consciousness, and its steadily increasing use to improve security of infrastructures, communications, adversarial, kinetic and non-kinetic uses (Mariarosaia & McNeish et al., 2023), consequently underpinning national defence and security services. Essentially, the provision of context recognises the dual-use nature of certain technologies and how an overarching definition of AWS without appropriate context may trigger articulation of genuine concern that subsequent regulation of AWS, based on such a definition, may lead to Governments being "*denied technologies and locked out of extremely important high-tech sectors, or that development of civilian applications of increasing autonomy will be harmed*" (UNIDIR, 2017). While it appears to be a valid concern, developing an overarching definition is merely the first step and does not dictate what aspects of the technology States may ultimately choose to regulate or control.

The subsequent wording amendment focuses on defining LAWS, in such a manner that they are viewed as genuinely distinct from AWS. Such a feat entailed a shift of the parameters of identification, selection and attack from operating as a form of mandatory checklist to individualised concepts of identification, and thusly conceptually closer to NATO's stipulations regarding autonomous systems. At the same time, the aspect of 'exercising kinetic force' was decoupled from AWS and relegated to operate specifically as a descriptor for LAWS. Due to this decoupling and with a purview towards the inclusion of 'less than lethal' weapons in this category, the definition for LAWS was expanded to include non-human targets and by consequence the infliction of material damage. Thusly by no longer relying solely on the use of kinetic force as an identifying characteristic, the additions compensate for the inclusion of new non-human targets (buildings etc.) and emerging weapons technologies that may not necessarily rely on kinetic force to inflict harm or damage (e.g. sonic and laser weapons[4]).

As a final point, one is likely to question the efficacy of maintaining the 'attack' descriptor in relation to the AWS, since the subsequent phrase related to the use of 'kinetic force' was extricated. Its continued inclusion was by design and harkens to militaries' interest in a wide variety of increasingly autonomous objects (e.g. supply convoys, surveillance UAVs). The question arises, how would those autonomous objects be protected? A possible answer would be through the inclusion of a self-defence system. However, as the original autonomous object would not be designed to be an offensive system (theoretically sidestepping the definition for LAWS, albeit on a technicality), it may not be captured under the definition of an AWS otherwise, in the absence of the wording (UNIDIR, 2017).

CHALLENGES FOR A WIDER DEFINITION

While the submission at hand sought to develop on the aforementioned definitory attempt for AWS, a challenge could be levied against the proposed rewording. In essence, that the 'new' definition is simply too wide, even with the added contextual elements of warfare, security and defence. In effect, it may be suggested that the rephrasing is self-defeating since it broadens the definitory scope, for the technologies (adhering to more of a functions approach), to such an extent as to render attempts at practical regulation administrably unworkable.

Nevertheless, broadening the scope is presented here as both necessary and justified. Adhering to the suggested existence of multiple categories of military technologies, namely 'sensors', 'computers and communications systems', 'weapons platforms and key enabling technologies of those platforms', and 'other types of weapons and technologies' (Michael O'Hanlon, 2020), to ensure the longevity and widespread applicability of any proposed definition, in the face of rapid technological progress and AI

[4]The feasibility of which was illustrated by the Defence Science and Technology Laboratory and Ministry of Defence. Science D and Laboratory T, "Advanced Future Military Laser Achieves UK First" GOV.UK (January 19, 2024) https://www.gov.uk/government/news/advanced-future-military-laser-achieves-uk-first last accessed May 4, 2025.

integration, the broader spectrum of weapons of war should be reasonably taken into account. Moreover, by remaining value-neutral and grounded in previously outlined parameters, it is suggested to have side-stepped the common pitfall of focusing on futuristic and unrealistic thresholds to the point that it would be rendered meaningless (House of Lords, 2023).

A further point of contention, may be the reference to particular targeting programs as a means of illustrating the breadth of AWS. To this end, it is stipulated that specific descriptors of the utilised examples have not been included in the definitory rewording in an attempt to avoid adhering too closely to the traditional, albeit limited, techno-centric approach to definitions. Rather, the examples were called upon to demonstrate the potential need for a shift away from the concept of 'autonomy of weapon systems' to 'autonomy in weapon systems' (Boulanin, 2016). Yet, it could be stated that the systems in question do not adhere closely to Boulanin's (2016) five specific characteristics of autonomy in weapon systems, (a) mobility, (b) health management, (c) interoperability, (d) battlefield intelligence and (e) use of force. Therefore, even if they could be covered under the AWS definition by virtue of the expanded military technology moniker, their inclusion would still not be warranted due to perceived lack of autonomy. The response to such a challenge would be to stipulate that such programmes can still be deemed as autonomous due to the relative nature of the term, with understandings of it differing depending on the discipline be engineering, robotics or computer science (ibid). This stance could be further validated by implementing the tri-approach to discerning autonomy:

The Human-Machine Command-and-Control Relationship: Namely, how involved is a human in the execution of the task carried out by the machine. In this instance, the technology in question could be stipulated to fall within the scope of 'semi-autonomous' systems, requiring human input at some stage of their target identification and acquisition process. Alternatively, it has been suggested that such AI-based targeting systems can produce outputs independently but are still under the oversight of a human reviewer, validating the generated materials, and therefore may be 'human-supervised autonomous' or 'human on the loop' (aspects of autonomy referenced in both the original definition and subsequent rewording).

The Sophistication of the Machine's Decision-Making Process: This approach refers to self-governance, a system's ability to exercise control over its behaviour and deal with uncertainties in its operating environment, somewhat predicated on the distinction between automatic, automated and autonomous systems. In this instance, while certainly not automatic, whether or not such programmes fit under the moniker of autonomous can be a contentious issue, considering that they do not depend on human oversight or control to function (though it may still be present), their set of rules though pre-defined do not generate consistently predictable outcomes and while their overall activity may be predictable, individuals aspects of their decision making may not be.

The Types of Decisions or Functions being Made Autonomous: With a lens towards the type of decisions or functions made autonomous within a system. Essentially, autonomy is best understood in relation to the type of

tasks on a subsystem/functional level, with certain ones (like targeting for instance as in the case of the referenced systems) operating as greater sources for concern (ibid).

Thusly, at least on a conceptual level, both in terms of weapons system but also autonomy, the expansion of the scope of the definition may be construed as an operational alternate lens on the existing AWS definitory debate.

The existence and subsequent acceptance of this wider definition would only facilitate a global consensus and recognition of the fact that autonomous weapon systems are varied in their form. States would subsequently still need to identify potential problematic applications of the technology and in turn develop appropriate regulatory process (UNIDIR, 2017). An internationally accepted definition is merely the starting point of regulatory discussion rather than a perpetual anchor for consideration.

MEANINGFUL HUMAN CONTROL

The concept of human control, partnering of man and machine, has been a notion explored alongside attempts to define autonomous weapon systems. Aspects of the former also typically exist with no shared definition (UNIDIR, 2017). In effect, human supervision has been enshrined as one of the principal safeguards not only of AWS but of artificial intelligence initiatives as well, rooted in the desire to uphold international humanitarian law and safeguard humanity by having a human agent capable of conducting effective oversight, engage in timely interventions or if necessary, even deactivation (International Committee of the Red Cross, 2021). Notwithstanding its espoused integral stature as a safeguard, given its relevance to the definitory process, an overview of its advocated strength is warranted. Specifically, what purports to be examined in this section are the concepts of responsibility, accountability and the homogenisation of behaviour.

Interestingly, the principle of meaningful human control is suggested to operate on a spectrum of effectiveness, an indication that it can be implemented minimally or maximally. The minimal implementation of human control would entail the existence of a human agent on the loop capable of understanding the functioning of the system with the ability to take the system offline if necessary. Conversely, maximal implementation would require the human agent in charge of the system to combine technical, legal and ethical training to ensure that the decision to allow the system to remain in operation is informed by all relevant dimensions, rather than merely vetting the system's operation (Mariarosaia & McNeish et al., 2023). Yet, it would be prudent to question if even this professed maximal implementation is enough to provide effective oversight, or if additional considerations need to be taken into account to ensure that humans on the loop, no matter their expertise, are not subsequently reduced to operating as another cog in the machine.

In context, the human-technology relations run the risk of becoming what Katharine Hayles calls a "cognitive assemblage". Essentially, humans and technological systems are inter-connected to such an extent that human users may be prompted and habituated into patterns of action or accepting the technology's outputs potentially even to the point of compulsion

(Schwarz, 2021); the homogenisation of behaviour. Although, the compulsive element may be on the extreme end of the scale, nonetheless the prospect does raise interesting concerns regarding the distribution of agency and the control human operators are able to realistically exert over the system they are overseeing. The concept of agency was further explored in relation to "*conditions under which moral inhibitions against violence become weakened*" (Renic & Schwarz, 2023). Essentially, the process of systematic killing can lead to target and agent degradation. With a specific focus on the former two concepts, through the three processes of:

Authorisation (the sanctioned approval of actions by a relevant authority that may separate cognition from effect, subsequently becoming assigned to more abstract goals that transcend the rules of standard morality (ibid), such as acts witnessed in warfare contexts),

Routinisation (the erosion of existing moral concerns and restraints through the reduction of necessary decision making and allowance of the avoidance of the implication of the action, by encouraging focus on details (ibid) rather than on the objective of the technology's outputs), and

Dehumanisation (Target objectification; depriving a victim of their human status to the extent that the principles of morality can be construed as no longer being applicable to their person), it is entirely possible for human agents to no longer be able to effectively execute their roles as overseers. While the majority of these concepts have been closely linked to existing forms of LAWS it is not outside the realm of possibility that this eroding effect can spread to handlers of newly defined forms of AWS.

In turn, this sparks questions regarding the notions of responsibility and accountability. Many scholars have indicated gaps in accountability relationships that will occur with the employment of AWS, leading to a possible accountability vacuum as these systems may diminish the user's moral agency, as explored above, and subsequent responsibility due to how the automated system in charge of the decision-making process is perceived (Verdiesen et al., 2020). To this end, who should be designated as the final loci of responsibility? A variety of candidates can be offered ranging from and including the programmer, a commanding officer, the overseer and perhaps even the machine, all of which have their own merits and demerits (Sparrow, 2016). Yet, the attribution of responsibility will ultimately hinge on the respect of transparency (Mariarosaria et al., 2023), a matter which in and of itself is not inherently transparent, even in acts of legislation. In effect, if one were to turn to the European Union's Artificial Intelligence Act 2024[5] for guidance, it would become apparent that the legislation in question prioritises system level safeguarding and while it does make mention to the type of qualifications human overseers should have, the requirement is fairly modest and mostly suggests 'due consideration' be afforded. In fact, the lack of specific guidance does make it somewhat arduous to appropriately discern the sharpness of the provision, as espoused by Lena (2022).

[5] While the legislation in question is not linked to AWS, it may provide insight into the subject of meaningful human control. A full overview of EU Artificial Intelligence Act 2024 is available through "The AI Act Explorer" https://artificialintelligenceact.eu/ai-act-explorer/ last accessed May 4, 2025.

CONCLUSION

Despite the passage of a number of years, discussions centred around AWS have continued to be informed by polarised views. Collaborative endeavours seem to be few and far between and those that have shown promise (UNODA, 2024) have been subsequently undermined by individual Governments' strategic or political considerations, thusly impeding the progress of generating internationally accepted definitions for both AWS and LAWS. The aim of this article is to offer an alternate vantage point on the subject matter; an interpretation of AWS that does not necessarily conform to all the current readily accepted specifications but rather, still rooted in relevant literature, builds on them without invalidating their contribution to the debate and by continuing to espouse a value-neutral stance. It seeks to do so by widening the definitory scope to include the premise of emerging technologies, while at the same time recognising the need for limits, duly put in place through the addition of context to the offered definition. Furthermore, it distinguishes the importance of human oversight as an integral part of not only the definitory process but also current and future regulatory attempts of AWS. Attempts to define various key aspects of human control are likely to experience a similar trudging journey.

REFERENCES

Boulanin Vincent, "Mapping the Development of Autonomy in Weapon Systems – A Primer on Autonomy", (2016), Stockholm International Peace Research Institute (SIPRI).

Enqvist Lena, "*Human oversight' in the EU artificial intelligence act: What, when and by whom?*", (2023), Law, Innovation and Technology, 15(2), 508–535.

House of Lords, "*Proceed with Caution: Artificial Intelligence in Weapon Systems*", (2023).

International Committee of the Red Cross (ICRC), "ICRC position on autonomous weapon systems", 12th of May 2021.

Ministry of Defence, "The Government Response to the Report by the House of Lords AI in Weapon Systems Committee: 'Proceed with Caution: Artificial Intelligence in Weapon Systems"', (2024).

NATO, "NATO Glossary of Terms and Definitions (English and French)", (2020), AAP-06.

O'Hanlon Michael, "*Forecasting Change in Military Technologies 2020–2040*", (2018), The Brookings Institution.

Renic Neil & Schwarz Elke, "*Crimes of Dispassion: Autonomous Weapons and the Moral Challenge of Systematic Killing*", (2023), Cambridge University Press.

Schwarz Elke, "*Autonomous Weapons Systems, Artificial Intelligence, and the Problem of Meaningful Human Control*", (2021), The Philosophical Journal of Conflict and Violence.

Sparrow Robert, "*Killer Robots*", (2007), Journal of Applied Philosophy, Vol. 24, No. 1.

Taddeo Mariarosaria & Blanchard Alexander, "*A Comparative Analysis of the Definitions of Autonomous Weapons Systems*", Oxford Internet Institute & Alan Turing Institute.

Taddeo Mariarosaria, McNeish David, Blanchard Alexander and Edgar Elizabeth, "*Ethical Principles for Artificial Intelligence in the Defence Domain*" Chapter 7 from "*Artificial Intelligence and International Conflict in Cyberspace*", (2023), Routledge.

United Nations Institute for Disarmament Research (UNIDIR), "*The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches*", 9th of November 2017.

United Nations Institute for Disarmament Research (UNIDIR), "*Proposals Related to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems: A Resource Paper*", 10th of May 2023.

United Nations Office for Disarmament Affairs (UNODA), "*Compilation of replies received to the Chair's guiding questions*" from 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, March 1st 2024.

United Nations Office for Disarmament Affairs (UNODA), "*Non-exhaustive compilation of definitions and characterizations*" from 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, March 10th 2023.

Verdiesen Ilse, Filippo Santoni de Sio & Virginia Dignum, "*Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight*", (2021), Minds & Machines 31, 137–163.