Section III: The Accountability Conundrum

Problem Statement: Recognition of AI's Capabilities and Risks

The contextualization in Section I has delved into some of the main characteristics of Artificial Intelligence technologies with regards to their application in Autonomous Weapons Systems.[126]

However, for the purposes of this Section, it is important to circle back and analyze some of their most relevant features.[127]

First, these technologies have the potential to act unpredictably.[128] It has become increasingly normal for AIs to rely on machine learning, which is in essence a form of soft-

[126] *See above:* Autonomous Weapon Systems (AWS).

[127] Ryan Abbot & Alex Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction,* 53 University of California at 330-332, (2019) [*hereinafter:* Abbot & Sarch, *Punishing AI*].

[128] See, *e.g.*, Taha Yasseri, *Never Mind Killer Robots — Even the Good Ones Are Scarily Unpredictable*, PHYS.ORG (Aug. 25, 2017) available at: <*https://phys.org/news/2017-08-mind-killer-robots-good-scarily.html*> accessed 11 February 2021; Janelle Shane, *Why Did the Neural Network Cross the Road?*, AI WEIRDNESS (2018), <*https://aiweirdness.com/post/174691534037/why-did-the-neural-network-cross-the-road*> accessed 11 February 2021, it describes a programmer who made her machine learning algorithm attempt to tell jokes.

ware that posterior to their initial configuration continues to develop in response to the data it acquires without any further explicit programming.[129] Therefore, the AI can act in ways its original programmers may not have intended or even foreseen.[130] The entrepreneur Elon Musk has also voiced his concerns calling for the establishment of a regulatory authority that would oversee the development of AI - warning that this could be the most likely cause of World War III.[131]

Second, AI has the potential to act unexplainably. It may be possible to establish what an AI has done, but not how or why it came up with that course of action.[132] This is what is commonly known as the black box[133] as explained above.[134] This is a natural result derived from its machine learning functions given that it will have been exposed to billions of data,[135] rendering it impracticable to trace which specific data point led to a particular outcome. Relatedly, at a

---

[129] See, *e.g.*, Davide Castelvecchi, *Can We Open the Black Box of AI?,* NATURE (Oct. 5, 2016), available at: *<https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731> [hereinafter:* Castelvecchi, *Can we Open the Black Box]*.

[130] There has been a recent focus on biased decisions by machine learning algorithms — sometimes due to a programmer's implicit bias, sometimes due to biased training data; *see, e.g.,* Chris DeBrusk, *The Risk of Machine-Learning Bias (and How to Prevent It),* MIT Sloan Management Review (2018), available at: *<https://sloanreview.mit.edu/article/the-risk-of-machine-learning-bias-and-how-to-prevent-it/>*.

[131] *See for example:* Osborne, S., *Elon Musk Calls for Urgent Laws on Robot as They Will Soon Be Risk to Public*, *Express*, 28 November 2017, available at: *<https://www.express.co.uk/news/science/885344/elon-musk-artificial-intelligence-robotics- regulation>* Accessed 29 October 2020.

[132] *See, e.g.,* Castelvecchi, *Can we Open the Black Box, supra* note 129.

[133] *Id.*

[134] *See above:* Section I: Contextualization.

[135] *Id.*

tech conference held in Lisbon in 2017, the physicist Dr. Stephen Hawking cautioned about the risks of AI by asserting that AI can be the worst case of human intelligence.[136]

Third, AI may act autonomously. This stems from both the abovementioned features. In the context of this study, this would imply an AI causing harm without being directly operated by an individual. AI can receive sensory input, set targets, assess outcomes against criteria, make decisions and adjust behavior to increase its likelihood of success — all without being directed by human orders.[137] It may even be the case that the programmer who sets an AI in motion is not able to regain control of the AI — maybe even purposely so by design.[138] It is a widely known fact that some of the major militaries in the world, including the US Air Force (USAF), already use some *semi* and fully autonomous technologies and have invested very significant resources in order to continue increasing these systems' autonomy.[139]

[136] *See* for example: Murphy, M., *Stephen Hawking: AI Could Be Best – or Worst – Thing in Human History*, New York City: MarketWatch (2017) available at: <*https://www.marketwatch.com/story/stephen-hawking-ai-could-be-best-or-worst-thing-in-human-history-2017-11-06*> accessed 29 October 2020.

[137] Abbot & Sarch, *Punishing AI*, *supra* note 127, at 333.

[138] *Id.*, at 331; they mention that "The DAO" was the most famous attempt to create a decentralized autonomous organization, with the purpose to deploy an entity that could no longer be controlled by its creators, acting without further direction; it was supposed to operate through smart contracts, or pre-programmed rules according to publicly available, unalterable code on a distributed ledger to prevent mismanagement; however it failed shortly after launch due to programming flaws and hacker interference; *see also* Samuel Falkon, *The Story of the DAO—Its History and Consequences*, THE STARTUP (Dec. 24, 2017), <*https://medium.com/swlh/the-story-of-the-dao-its-history-and-consequences-71e6a8a551ee*>.

[139] Palmer, A., *Autonomous UAS: a partial solution to America's future airpower needs*. Air University in partial fulfillment of the graduation requirements (2010) available at: <*https://apps.dtic.mil/dtic/tr/fulltext/u2/1018416.pdf*> Accessed 29 November 2020.

Fourth, the interplay between AIs that are created to perform "narrow" or "specific" tasks[140] and those that might be developed with "general" AI that would be able to perform any task a par to human abilities,[141] or most likely even better, as was concluded by a study conducted by researchers from Oxford and Yale Universities.[142] However, it is uncertain when and how this technology will be fully developed and deployed.[143]

Granted, any conventional machine could also act unpredictably, unexplainably, or autonomously at a given moment. Yet, AI is far more likely to exhibit these characteristics and to a greater extent.[144] It can also be agreed that the interdisciplinary implications of these technologies render us increasingly less suited to understand, let alone regulate, their behavior, which in turn exacerbates their unpredictability.

In fact, machines have caused harm since ancient times, and robots have caused fatalities since at least the 1970s.[145] However, except for the cases in which the machines have been used in a way intended to inflict harm, the majority of these events have been ruled as accidents, or as the result of negligence or recklessness on the part of the operator. Yet, these

---

[140] Abbot & Sarch, *Punishing AI*, *supra* note 127, at 331-2; *see also* Ryan Abbott, *Everything is Obvious,* 66 UCLA L. Rev. at 25 (2019).

[141] *Id.*

[142] Grace, K. et al. (2018) *Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts*, 62 Journal of Artificial Intelligence Research, at 729-754.

[143] *See generally* Vincent C. Müller & Nick Bostrom, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion,* in Vincent C. Müller (ed.), *Fundamental Issues of Artificial Intelligence, at* 555 (2016). He describes a survey finding that experts think AI superintelligence will not be a reality for at least a few decades.

[144] Abbot & Sarch, *Punishing AI*, *supra* note127, at 332.

[145] *Id.*; *see also* Ryan Abbott, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability*, 86 The George Washington Law Review, at 8 (2018).

accidents implicate criminal law[146] which cannot be deployed against the harmful machines themselves.

The truth of the matter is that AI differs from conventional machines in the aforementioned essential ways has puzzled scholars and legal practitioners alike with regards to the application of traditional concepts of criminal law as we know them.

In this sense, it is key to consider the element of reducibility because if an AI engages in conduct that would be criminal for a person and the act is reducible, then there typically will be a human that could be held criminally liable for it. By contrast, if AI conduct is not effectively reducible, there may be no other party that is aptly punished, in which case criminal activity could occur with impunity.[147] As was stated by the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, if the nature of a weapon renders responsibility for its consequences impossible, its use should be considered unethical and unlawful as an abhorrent weapon.[148]

Therefore, it is imperative to maintain all AI conduct within the spectrum in which they are likely to be reducible. As we currently know AI technologies, even where it behaves autonomously, to the extent that a certain person uses AI as a tool to commit a crime, and the AI functions in a foreseeable fashion, the crime then is reducible to an identifiable individual causing the harm.[149] Even when AI causes unforeseeable harm, it may still be reducible — for

---

[146] *See* United States Department of Justice, *BP Exploration and Production Inc. Agrees to Plead Guilty to Felony Manslaughter, Environmental Crimes and Obstruction of Congress Surrounding Deepwater Horizon Incident*, U.S. DEP'T JUSTICE (15 November 2012) available at: <*https://www.justice.gov/opa/pr/bp-exploration-and-production-inc-agrees-plead-guilty-felony-manslaughter-environmental*> accessed 2 February 2021; (outlining BP's guilty plea to criminal offenses).

[147] Abbot & Sarch, *Punishing AI*, *supra* note 127, at 369-375.

[148] Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions,* U.N. Doc. A/HRC/23/47, at 80 (2013) [*hereinafter:* Heyns, *Report on Extrajudicial, Summary or Arbitrary Executions]*.

[149] Abbot & Sarch, *Punishing AI*, *supra* note 127, at 334.

example, if an individual creates an AI to steal financial information, but a programming error results in the AI shutting down an electrical grid that disrupts critical infrastructure.[150] This is a familiar problem in criminal law for which doctrinal tools have been developed by which liability could still be imposed.[151]

However, with the AIs we are knowing today it is becoming increasingly difficult to reduce AI crime to an individual due to the technology's autonomy, complexity and lack of explainability. Additionally, owing to the fact that a large number of individuals may contribute to the AI's development over a long period of time,[152] and as a result of their machine learning functions, it may be difficult to attribute responsibility to a specific individual for an AI output where the machine has gathered information on how to behave based on accessing billions of data points from heterogeneous sources.[153]

In our coexistence with intelligent agents, the forecast from the combination of these elements portrays a dire scenario which brings us to challenge the extent to which, as humans, we can still be in control of these technologies - especially when they have kinetic impacts on the physical world. Intuitively, this raises the uncomfortable questions that have been puzzling academics and researchers:[154]

---

[150] *Id.*

[151] *Id.*

[152] In 2017, for instance, more than 4,500 Microsoft employees contributed to open-source software hosted on GitHub, a development platform that host open-source code; *see* Matt Asay, *Who Really Contributes to Open Source*, INFOWORLD (7 February 2018) available at: <*https://www.infoworld.com/article/3253948/who-really-contributes-to-open-source.html*> accessed 2 February 2021; *see* Frederic Lardinois & Ingrid Lunden, *Microsoft Has Acquired GitHub for $7.5B in Stock*, Techcrunch, (June 4, 2018) available at: <*https://techcrunch.com/2018/06/04/microsoft-has-acquired-github-for-7-5b- in-microsoft-stock/*> accessed 2 February 2021.

[153] Lothar Determann & Bruce Perens, *Open Cars*, 32 Berkeley Tech. L.J. at 915, 988 (2017).

[154] *See* for example: Willick, M. *Artificial Intelligence: Some Legal Approaches and Implications*. 4 (2)AI Magazine, at 5 (1983) available at: <*https://aaai.org/ojs/index.php/ aimagazine/article/view/392*>; Curties E. A. Karnow, *Li-*

What happens if harm is brought upon by these technologies? Who will be held responsible? Are we in the face of a serious accountability gap?

Naturally, these queries become even more pressing in the face of AI technologies applied to AWS by which the lives of civilians are put at risk. Those who argue that there will indeed be an accountability gap[155] if civilians are unlawfully killed through the use of an AWS have advanced this proposition to justify either their prohibition or restriction.[156]

This disagreement is displayed in a very straightforward manner at the State level. The latest Reports of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (GGE)[157] records the fact that States are quite divided over the development and future use of AWS. On one side, the prohibitionist States contend

ability for Distributed Artificial Intelligences 11(1)Berkeley Technology Law Journal, at 188 (1996), available at: *<https://lawcat.berkeley.edu/record/1115611?ln=en>* both accessed on 18 November 2020.

155 Thompson Chengeta, *Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law*, 45 (1) Denver Journal of International Law and Policy at 2,4 (2020) *[hereinafter:* Chengeta, *Accountability Gap]*; Sparrow, *Killer Robots, supra* note 45, at 62; ICRC, *Autonomous Weapon Systems Implications of Increasing Autonomy*, at 44 (2016); Nathalie Weizmann et al., *Autonomous Weapon Systems Under International Law*, Geneva Academy of International Humanitarian Law, Academy Briefing no. 8, at 24 (2014) [*hereinafter:* Weizmann et al., *AWS under International Law]*.

156 Carrie McDougall, *Autonomous Weapon Systems and Accountability: Putting the Cart Before the Horse,* 20 Melbourne Journal of International Law, at 7, 13 (2019) [*hereinafter:* McDougall, *AWS and Accountability]*.

157 The Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (GGE) was established by the High Contracting Parties to the *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, (CCW) and is the primary international forum for discussions among states on the development and future use of autonomous weapon systems (AWS)*; see for instance:* GGE, *Working paper by the Bolivarian Republic of Venezuela on behalf of the Non-Aligned Movement (NAM) and Other States Parties to the Convention on Certain Conventional Weapons (CCW)*, CCW/GGE.1/2020/WP.5 (2020); *see also* GGE, *United Kingdom Expert paper: The human role in autonomous warfare,* CCW/GGE.1/WP.6 (2020).

that AWS should be banned outright or at least place the development of these weapons under a moratorium.[158] Another block is calling for negotiations on a regulatory treaty arguing that the potential of AWS should be constrained primarily on a requirement to ensure meaningful human control.[159] A third group argues that a political declaration would be sufficient, and yet others remain against any form of international regulation beyond the existing rules of international law.[160]

---

[158] Heyns, *Report on Extrajudicial, Summary or Arbitrary Executions, supra* note 148, at 22; Human Rights Watch (HRW) and International Human Rights Clinic, *Mind the Gap: The Lack of Accountability for Killer Robots,* Human Rights Watch, (2015) available at: *<https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots#> [hereinafter:* HRW, *Mind the Gap]*; Darren M. Stewart, *New Technology and the Law of Armed Conflict,* 87 International Law Studies, at 291-294 (2011) [*hereinafter:* Stweart, *New Technology and the Law of Armed Conflict]*; Mary Ellen O'Connell, *Banning Autonomous Killing: The Legal and Ethical Requirement That Humans Make Near-Time Lethal Decisions,* in Matthew Evangelista and Henry Shue (eds.), *The American Way of Bombing: Changing Ethical and Legal Norms, from Flying Fortresses to Drones,* Cornell University Press, at 224, 236 (2014).

[159] *See, e.g.,* Peter Margulies, *Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts*, in Jens David Ohlin (ed.), *Research Handbook on Remote Warfare*, Edward Elgar Press, at 19 (2016) [*hereinafter:* Margulies, *Making Autonomous Weapons Accountable]*; Chengeta, *Accountability Gap, supra* note 155, at 2,4; Swati Malik, *Autonomous Weapon Systems: The Possibility and Probability of Accountability,* 35(3) Wisconsin International Law Journal, at 621-25 (2018) [*hereinafter:* Malik, *AWS]*; Amos N Guiora, *Accountability and Decision Making in Autonomous Warfare: Who Is Responsible?,* (2017)(2) Utah Law Review, at 393-398 (2017) [*hereinafter:* Guiora, *Accountability and Decision Making]*. Guiora in fact suggests that 'accountability standards must be stricter' for AWS, at 418. Allyson Hauptman, *Autonomous Weapons and the Law of Armed Conflict* 218(Winter) Military Law Review at 170, 193 (2013); Jack M Beard, *Autonomous Weapons and Human Responsibilities*, 45(3) Georgetown Journal of International Law at 617, 674, 675, 681 (2014); Michael Aaronson, *Robots Don't Kill People, It's the Humans We Should Worry About*, The Conversation (31 May 2013) available at *<https://theconversation.com/robots-dont-kill-people-its-the-humans-we-should-worry-about-14779>* accessed 5 March 2021; ICRC, *Ethics and AWS, supra* note 122, at 1, 2, 11. Davison, *AWS under IHL, supra* note 44, at 5, 17-18; Weizmann et al., *AWS under International Law*, *supra* note 155, at 5.

[160] Group of Governmental Experts of the High Contracting Parties to the CCW, *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc CCW/GGE.1/2018/3 at 28-29 & annex III 46-48 (23 October 2018) [*hereinafter:* GGE, *Report of the 2018 Session]*.

In this respect, numerous concerns in relation to AWS have been raised by a wide variety of actors, including the International Committee of the Red Cross[161] and the Campaign to Stop Killer Robots, a coalition of 106 non-governmental organizations in 54 countries.[162] Scholars are equally divided, with views both for and against AWS firmly expressed.

While it is unknown how these technologies will continue to evolve and more importantly, in which circumstances they could be deployed, the questions regarding accountability for them are a pivotal matter that should be addressed at the forefront of these debates.

It has been noted that one of the leading arguments of those calling for the prohibition, moratorium or the regulation of AWS is that the use of such weapons will result in an accountability gap, meaning that there will be virtual impunity for any violation of the law resulting from their use.

Relatedly, it is noteworthy to mention that some scholars have contended that the criminalization for the use of AWS would be preferable to a prohibition of them on four grounds.

First, because placing the emphasis directly on individuals would brink other kinds of issues related to the signing and ratifying of international treaties, international enforcement and establishing state responsibility.[163] Second, given that this would, in their view, convey the

---

[161] *See generally* ICRC, *Autonomous Weapon Systems, Implications of the increasing autonomy in the critical functions of weapons,* Expert Meeting, Switzerland (15-16 March 2016); access to this and other publications on the topic can be found at: ICRC, *New Technologies and IHL,* available at <*https://www.icrc.org/en/war-and-law/weapons/ihl-and-new-technologies*> accessed on 11 March 2021.

[162] *See* the website for the *Campaign to Stop Killer Robots* available at <*https://www.stopkillerrobots.org*> accessed 11 March 2021.

[163] Hin-Yan Liu, *Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems,* in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge University Press, at 344 (2016) [*hereinafter:* Liu, *Refining Responsibility*].

message that any situation resulting in impunity arising from the use of AWS would be of a legal nature rather than a matter of technical inadequacy. Third, because a premature prohibition may unduly stifle the development of autonomous technologies which may also have other legitimate civilian applications. And fourth, because the criminalization approach could be readily rescinded[164] if the current concerns around the use of AWS were to be subsequently resolved.[165]

In this regard, as will be explained below,[166] it would be a significant leap to assume that individual criminal responsibility for the use of AWS could be ascertained in the absence of a previous prohibition or international agreement on the matter.

However, it is the opinion of the author that although the assessments about the legality or propriety of the use of this category of weapons which could lead to a potential regulation or ban and those regarding the accountability gap might be conceptually connected, these discussions have different objectives *in se* and thus should be analyzed separately. For the purposes of this Section, only the accountability gap debate will be explored.

What Kind of Accountability Can We Expect?

Some of the first probes on the challenges of legal responsibility for actions of intelligent machines came about over twenty years ago, acknowledging that where a machine attains a certain level of intelligence to the extent of making decisions by itself, difficulties arise

---

164 *See* McDougall, *AWS and Accountability, supra* note 156, at 25. She considers that the assertion that criminalization could be 'rescinded' that much more easily than a prohibition, or at least a moratorium, or that it would have a differentiated effect on civilian applications might be far-fetched.

165 *Id.*

166 *See below:* Accountability Via International Criminal Law.

in imputing responsibility.[167] Those early concerns remain just as valid today as we are faced with the same question marks since no matter how fast the machines' autonomy increases and how sophisticated they become, they still do not have moral agency.[168]

At this point, it is paramount to zoom out and make two important conceptual clarifications as caveats to bear in mind for the rest of this discussion.

The first is to acknowledge that AWS may have two facets, those that can be fully autonomous in which the human is "out of the loop" and those that are semi-autonomous as they operate automatedly in tandem with humans "inside the loop".[169] Moreover, given that until now the baseline in international discussions seems to be aiming at ensuring meaningful human control[170] as an attempt to ease objections against AWS, it is quite relevant to identify what the resulting dynamic would actually entail operationally, given that this term is far from being defined homogeneously.[171]

---

[167] *See* Perri 6, *Ethics, Regulation And the New Artificial Intelligence, Part II: Autonomy And Liability,* Information, Communication and Society, at 406-34, 414 (2001).

[168] Markus Wagner, *Taking Humans Out of the Loop: Implications for International Humanitarian Law,* 21 Journal of Information, Law & Science, at 5, 11 (2011) [*hereinafter:* Wagner, *Taking Humans Out of the Loop*]; Peter Asaro, *On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making,* 94 International Review of the Red Cross, at 693 (2012) [*hereinafter:* Asaro, *On banning AWS].*

[169] *See* Marta Bo, *The Human-Weapon Relationship in the Age of Autonomous Weapons and the Attribution of Criminal Responsibility for War Crimes*, Working Draft, at 1 (2019); *see also* William C Marra and Sonia K McNeil, *Understanding the Loop: Regulating the Next Generation of War Machines,* 36 Harv. J. L. & Pub. Pol'y at 1139, 1150 (2013). Marra and McNeil affirm that although the terms "automation" and "autonomy" are similar, automated systems are not self-directed, they also lack decision-making capability, they simply have the capacity to operate without [human intervention]. By contrast, autonomous entities are capable of being independent in the establishment and pursuit of their own goals.

[170] GGE, *Report of the 2018 Session, supra* note 160, at 22.

[171] For explorations of the meaning of 'meaningful human control', *see* Michael C. Horowitz and Paul Scharre, *Meaningful Human Control in Weapons Systems: A Primer*, Working Paper, Centre for New American Security

In any case, both of these scenarios need to be addressed in terms of accountability for their acts for they may equally result in impunity otherwise.

The second one is to recognize that given the particularities of AI and AWS, they have been widely subjected to anthropomorphisms. However, this is a cognitive error because it misleads to the correlative flawed expectation of AI being able to adhere to social norms or (human) preestablished behavioral patterns.[172] As mentioned before,[173] upon careful observation one realizes that these entities are neither weapons, conventional platforms, nor moral agents tantamount to humans for legal purposes. Yet they are often referred to as the first, on occasion as the second, and frequently treated as the third.[174]

The above is relevant in terms of this discussion because irrespective of how advanced the technology may become, the machine will never be a responsible moral agent[175]

(2015) available at: <*https://www.cnas.org/publications/reports/meaningful-human-control-in-weapon-systems-a-primer*>; Merel Ekelhof, *Autonomous Weapons: Operationalizing Meaningful Human Control*, Humanitarian Law and Policy (2018) available at: <*https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/*> accessed 15 March 2021.

172 Abbot & Sarch, *Punishing AI*, *supra* note 127, at 333; ("We will not attempt to articulate the non-functional differences between human and algorithmic reasoning, a subject which has fascinated and confounded computer scientists since the 1950s.").

173 *See above:* Autonomous Weapon Systems (AWS).

174 Shilo, *When Turing Met Grotius, supra* note 25, at 15.

175 This is asserted as an axiomatic fact by most authors writing on accountability and AWS; *see, e.g.*, Malik, *AWS, supra* note 158; *Heyns*, *Report on Extrajudicial, Summary or Arbitrary Executions, supra* note 148; Sparrow, *Killer Robots, supra* note 45, at 65-8, 71-3; Noel E Sharkey, *The Evitability of Autonomous Robot Warfare* 94(886) International Review of the Red Cross at 787, 790 (2012). For a contrary view, *see* John P Sullins, *When Is a Robot a Moral Agent?* 6 International Review of Information Ethics at 23 (2006); Jens David Ohlin, *The Combatant's Stance: Autonomous Weapons on the Battlefield*, 92 International Law Studies at 1, 2 (2016) [*hereinafter:* Ohlin, *The Combatant's Stance*].

and thus, accountability can never be transferred to it. States have reached a consensus at least on this much.[176]

Furthermore, as will be explained in detail below,[177] the analogy has been drawn between the relationship of a human commander *vis-á-vis* a human subordinate with that of a human commander *vis-á-vis* a *robot.*[178] The continued referral of a person deploying AWS as a *commander* gives a misleading impression that AWS are somewhat combatants or fighters.[179]

Having said that, another very important point that has reached consensus amongst States is that "accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control."[180]

In addition, States also agreed that "humans must at all times remain accountable in accordance with applicable international law for decisions on the use of force".[181]

---

[176] Amidst a scant list of 10 'possible guiding principles', the GGE agreed that '[h]uman responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system': GGE, *Report of the 2018 Session, supra* note 160, at 4.

[177] *See below:* Command Responsibility.

[178] Chengeta, *Accountability Gap, supra* note 155, at 3.

[179] Bonnie Docherty, *Losing Humanity: the Case Against Killer Robots,* 1 Human Rights Watch at 4, 33-34, 42-43 [*hereinafter:* Docherty, *Losing Humanity*].

[180] GGE, *Report of the 2018 Session, supra* note 160, at 4.

[181] *Id.,* at 5.

An operational view that also evinces the need for human accountability behind AI's decision-making is the Law of War Manual of the United States Department of Defense,[182] which includes Provision 6.5.9.3 "Law of War Obligations of Distinction and Proportionality Apply to Persons Rather Than the Weapons Themselves" and it stipulates that "the law of war does not require weapons to make legal determinations, even if the weapon (*e.g.*, through computers, software, and sensors) may be characterized as capable of making factual determinations, such as whether to fire the weapon or to select and engage a target". This has been regarded as a sign that robotic weapons are never responsible legal agents,[183] thereby raising the inevitable question – what prospects for accountability *actually* exist for AWS?

As a starting point, it is necessary to pause on the nomenclature "accountability" in order to envision what the scope of this concept could actually entail in the context of this debate.

It can be safely stated that accountability is a broad-spectrum concept, which has been used as an umbrella term to describe various forms of legal responsibility, including state responsibility, administrative and disciplinary proceedings undertaken in response to violations of IHL (even encompassing military justice), civil liability, and individual criminal responsibility.[184] All of these modalities are of a complementary nature to each other and by no means are they alternatives to the exclusion of the other.[185] Accountability is important in international law

---

[182] U.S. Department of Defense, *Law of War Manual,* (2016) available at: *<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>* accessed 10 August 2020.

[183] Bryson J.J., Diamantis M.E., Grant T.D., *Of, for, and by the people: the legal lacuna of synthetic persons*; 25 Artificial Intelligence and Law, at 273-291 (2017).

[184] McDougall, *AWS and Accountability, supra* note 156, at 7.

[185] Chengeta, *Accountability Gap*, *supra* note 155, at 3.

because where there is an accountability gap, the victims' right to a legal remedy is adversely affected.[186]

It is also noteworthy that the design, development, and/or use of AWS might implicate more general principles and rules found in various fields of international law such as *jus ad bellum*, IHL, international human rights law, international criminal law (ICL), and space law, among others.

Consequently, the legal recourses for accountability in each of these regimes may vary significantly in scope. For the purposes of this study, accountability is understood as the duty to account for the exercise of power over the design, development, or use (or a combination thereof) of AWS acknowledging that power may be exercised by a wide variety of actors.[187]

As a bottom-up approach, the first and most intuitive approximation for the author would be individual criminal responsibility under international law for the commission of international crimes, such as war crimes, involving the use of AWS. In order to determine this criminal liability it is necessary to ascertain the commission of a defined crime under international law, establish the competent jurisdiction over that crime, delimit which mode of responsibility is fulfilled by the conduct of a particular individual, demonstrate that the material (*actus reus*) and subjective (*mens rea*) elements of the crime in question are met, assess the existence of a legal justification if applicable, and if there is a conviction, impose a sentence and if applicable, reparations for victims.[188]

---

[186] Megan Burke & Loren Persi-Vicentic, *Remedies and Reparations,* in S. Casey-Malsen (ed.), *Weapons Under International Human Rights Law*, at 542-89 (2014) [*hereinafter:* Burke & Persi, *Remedies and Reparations]*; Luke Moffett, *Justice for Victims before the International Criminal Court,* at 146 (2014).

[187] Lewis et al., *War-Algorithm*, *supra* note 32, at 11.

[188] *Id.*, at 12, 77.

Conversely, from a top-down approach, the second avenue is to invoke state responsibility derived from acts or omissions involving the use of AWS where those acts or omissions entail a breach of an existing rule of international law.[189] In order to allocate this State responsibility, it is required to establish the existence of a rule enshrined in treaty or customary law, discern the legal obligation derived from the rule in question, identify a breach to said rule, and most importantly, *attributing* that breach to the State, determine if there are any applicable legal excuses for such act or omission and if the State is deemed responsible, impose reparations for the victims.[190]

On the other hand, an ideally parallel approach to the two mentioned above would be to apply scrutiny governance. Albeit the consequences derived from this option might appear to be of a laxer nature, this channel must be included considering that in the current geopolitical context it might become *in practice* the most available recourse. This route contemplates the extent to which a person, *or entity,* is and should be subject to, or should exercise, forms of internal or external scrutiny, monitoring, or regulation (or a combination thereof) concerning the design, development, or use of an AWS. Some examples of scrutiny governance include independent monitoring, regulatory development, adopting non-binding resolutions and codes of conduct, normative design of technical architectures (including maximizing the auditability of algorithms) and community self-regulation.[191]

The author will continue to explore the most prevalent of these legal alternatives throughout the series of this broader study, however for the purposes of this Section, 'accountability' is related only to individual criminal responsibility.

---

[189] *Id.,* at 83-84.

[190] *Id.,* at 54, 84.

[191] *Id.*, at 91.

Given that this is a global issue incumbent upon all of humanity, let us begin by exploring the question of accountability from the perspective of international criminal law.

Accountability Via International Criminal Law

At this stage, it is important to recall the *raison d'être* and salience of this discussion. Accountability mechanisms are essential to bring about deterrence[192] by complying with IHL obligations to prosecute grave breaches and war crimes,[193] procuring prevention which is pivotal for the protection of civilians,[194] and which are also fundamental to ensure the victims' rights to reparations.[195] In short, accountability has been called "the crux of international law".[196]

Therefore, as a matter of pragmatism, international law cannot be understood as limited to setting standards for governments, non-state actors and their agents, but rather as en-

---

[192] *See, e.g.,* Guiora, *Accountability and Decision Making, supra* note 159, at 398: '"[k]ill/not kill" decisions authorized by the nation-state where standards of accountability are neither inherent nor integral is akin to authorizing the new Wild West'; Stewart, *New Technology and the Law of Armed Conflict, supra* note 158, at 292; he refers to 'the broader public policy issues associated with the possibility of military operations being conducted in a "blameless environment"'.

[193] GC IV, Art 146; it requires grave breaches to be criminalized and prosecuted. There is also an obligation to prosecute a broader range of war crimes under customary international law.

[194] *See, e.g.,* Heyns, *Report on Extrajudicial, Summary or Arbitrary Executions, supra* note 148, at 75.

[195] *See, e.g.,* Chengeta, *Accountability Gap, supra* note 155, at 5.

[196] *Id.,* at 49; *see also* Malik, *AWS, supra* note 159 at 620; Guiora, *Accountability and Decision Making, supra* note 159, at 398; HRW, *Mind the Gap, supra* note 158; Heyns, *Report on Extrajudicial, Summary or Arbitrary Executions, supra* note 148, at 75.

compassing the prescription of consequences for failures in compliance with them.[197] Furthermore, IHL norms —some of which are *jus cogens*— lack value without accountability recourses for infringing them.[198] It can certainly be argued that an accountability void for international law violations effectively poses a global threat to the general maintenance of peace and security.[199]

The author concurs with the statement that, "after all, without accountability, international law is nothing but the proverbial *brutumfulmen* - a harmless thunderbolt."[200]

Additionally, as mentioned above, accountability is also fundamental because it is inherently connected to the right to remedy for both civilian and military victims in cases of unlawful use of a weapon that has been outlawed or that is indiscriminate as a method of warfare in an armed conflict, or the use of force that is disproportionate or excessive during law enforcement. The afore also extends to willful or negligent failure to protect victims from harmful weapons insofar as these have been recognized as unlawful conduct tantamount to human rights violations.[201] It is the author's opinion that AWS classify squarely as a harmful weapon, from which all persons are entitled to protection by international standards,[202] therefore the accountability challenges that are posed by their use must be taken very seriously.

---

[197] Steven Ratner et al., *Accountability for Human Rights Atrocities in International Law*: *Beyond the Nuremberg Legacy,* Oxford 3rd ed. (2009) [*hereinafter:* Ratner et al., *Accountability for Human Rights Atrocities].*

[198] Chengeta, *Accountability Gap, supra* note 155, at 5; Anja Seibert-Fohr, *Prosecuting Serious Human Rights Violations,* Oxford at 292-93 (2009).

[199] *See* John R.W.D. Jones & Steven Powles*, International Criminal Practice,* 3rd ed 2 (2003), [*hereinafter:* Jones & Powles, *International Criminal Practice]*.

[200] Chengeta, *Accountability Gap, supra* note 155, at 5.

[201] Burke & Persi, *Remedies and Reparations*, *supra* note 186, at *554.*

[202] *See above:* Applicability of the Martens Clause. Sharing this view*, see* Chengeta, *Accountability Gap*, *supra* note 155, at 5.

The right to remedy, understood as a process which is meant to provide victims with justice, remove or redress to the extent possible the damage done by the unlawful acts through prevention and deterrence,[203] is relevant in the context of this Section because it is the duty of States to give effect to victims' rights by investigating human rights violations and bringing perpetrators to justice through prosecution.[204]

In this regard, individual accountability can be characterized as a "complex amalgam of law and a wide spectrum of sanctioning processes that transcends the orthodox divisions of subjects of international law."[205]

As a result of the above, individual criminal responsibility has become a part of customary international law[206] in order to deter and/or punish unlawful acts committed in international armed conflicts, non-international armed conflicts[207] and also during peacetime.[208]

---

[203] *See* Roman David & Susanne Choi Yuk-ping, *Victims on Transitional Justice: Lessons from the Reparation of Human Rights Abuses in the Czech Republic,* 27 Human Rights Quarterly at 392-393 (2005); Riccardo Pisillo Mazzeschi, *Reparation Claims by Individuals for State Breaches of Humanitarian Law and Human Rights: An Overview,* I Journal of International Criminal Justice, at 339, 344 (2003).

[204] United Nations Human Rights Committee (HRC), *General comment no. 31 [80], The nature of the general legal obligation imposed on States Parties to the Covenant,* CCPR/C/21/Rev.1/Add.13, at 2-3 (2004); *see* European Court of Human Rights, ECtHR, *Aksoy v. Turkey*, Judgment, European Court of Human Rights (ECtHR) (1996); *see* African Commission on Human and People's Rights, *Social and Economic Rights Action Centre And Centre for Economic and Social Rights v. Nigeria*, Communication No. 155/96, at 44-48 (2001).

[205] Ratner et al., *Accountability for Human Rights Atrocities, supra* note 197, at 3.

[206] Bert Swart, *Modes of International Criminal Liability, in* The Oxford Companion to International Criminal Justice at 82, 91 (Antonio Cassese ed., 2009) [*hereinafter:* Swart*, Modes of International Criminal Liability]*.

[207] International Criminal Tribunal for the Former Yugoslavia (ICTY), *Prosecutor v. Tadic,* Decision on the Defense Motion for Interlocutory Appeal on Jurisdiction, at 129, (2 October 1995).

[208] Crimes Against Humanity and Genocide do not require the existence of an armed conflict; *see for instance* Efrat Bouganim-Shaag and Yael Naggan, *Emerging Voices: Peace-Time Crimes Against Humanity and the ICC*, Opinio

The Crimes

International Criminal Law (ICL) is the branch of public international law that establishes individual criminal responsibility for international crimes, *i.e.*, war crimes, crimes against humanity, genocide, and aggression. Its purpose within the international legal order is of a multifaceted nature, as it aims at contributing to general and specific deterrence, incapacitation, rehabilitation, reconciliation, justice to victims, retribution, truth-telling, promotion of the rules-based international order, establishing and maintaining an inclusive and sustainable peace, etc.[209]

Therefore, for international criminal law to be applicable, we need to be *vis-á-vis* an international crime.

In this sense, the following quote from one of the fathers of modern international law, the late Cherif Bassiouni comes to mind: "international crimes have developed to date, without… an agreed-upon definition of what constitutes an international crime, what are the criteria for international criminalization, and how international crimes are distinguished".[210]

Yet, one could conclude from the jurisprudence of the international criminal courts and tribunals that individual criminal responsibility has only been attributed for conducts that are already prohibited by custom or treaty under another branch of public international law such as IHL, international human rights law or *jus ad bellum*.[211]

Juris, (2013) available at <*http://opiniojuris.org/2013/08/30/emerging-voices-peace-time-crimes-humanity-icc/*> accessed 20 February 2021; *see also Convention on the Prevention and Punishment of the Crime of Genocide,* article 1 (1948).

[209] McDougall, *AWS and Accountability, supra* note 156, at 28.

[210] M Cherif Bassiouni, *Introduction to International Criminal Law,* Transnational Publishers at 111 (2003).

[211] Yoram Dinstein, *International Criminal Law,* 20(2-3) Israel Law Review, at 206, 221 (1985); Ilias Bantekas and Susan Nash, *International Criminal Law,* Cavendish Publishing, 2nd ed, 5 (2003); Antonio Cassese, *International Criminal Law,* Oxford University Press, 2nd ed, at 11-12 (2008).

The Rome Statute was adopted in 1998 in order to establish a permanent and global institution, *i.e.*, the International Criminal Court (ICC), that would prosecute and judge the most abhorrent crimes known to man. As such, the Statute was meant to provide normative guidance for what those crimes could amount to in the future as well albeit with the limitation of the temporal references of the conflicts of the 20th century - which clearly did not yet involve the technological sophistication of those that we are starting to know today.

Evidence of the latter is the fact that the majority of discussions circled around the issue of custom, requiring evidence of a widely accepted prohibition of the conduct in question, or even of a pre-existing *crime* under customary international law.[212] Therefore, this meant that the crime definitions were adopted looking backwards in time - not forwards.[213]

The drafters did provide for the establishment of a Working Group on Amendments, an ICC's Assembly of States Parties that considers amendments to the Rome Statute and the Court's Rules of Procedure and Evidence.[214]

---

[212] United Nations Diplomatic Conference of Plenipotentiaries on the Establishment of an International Criminal Court, *Summary Records of the Plenary Meetings and of the Meetings of the Committee of the Whole*, UN Doc A/CONF.183/13 (Vol II) (15 June-17 July 1998); *see in particular:* 150 (UK), 151 (Slovenia), 154–5 (Canada), 155 (Israel), 158 (Syria), 160 (New Zealand), 160 (Greece), 162 (Belgium), 164 (France), 187 (Jordan), 270 (China), 277 (Switzerland), 277 (Brazil), 278 (Korea), 285 (Bosnia and Herzegovina), 287 (Indonesia), 289 (Russia).

[213] Herman von Hebel and Darryl Robinson, *Crimes within the Jurisdiction of the Court* in Roy S Lee (ed), *The International Criminal Court: The Making of the Rome Statute*, Kluwer International, 79, 104, 122-3 (1999); William A Schabas, *An Introduction to the International Criminal Court,* Cambridge University Press, at 23 (2001). For a comparison of the *Rome Statute*'s definitions of crimes with customary international law, *see* Antonio Cassese, *Genocide* in Antonio Cassese, Paola Gaeta and John RWD Jones (eds.), *The Rome Statute of the International Criminal Court: A Commentary,* Oxford University Press vol 1, at 335 (2002); Antonio Cassese, *Crimes against Humanity,* in Antonio Cassese et al. (eds.), *The Rome Statute of the International Criminal Court: A Commentary*, Oxford University Press vol 1 at 353 (2002); Michael Bothe, *War Crimes,* in Antonio Cassese et al. (eds.), *The Rome Statute of the International Criminal Court: A Commentary*, Oxford University Press vol 1 at 379 (2002)

[214] International Criminal Court, *Rules of Procedure and Evidence*, Doc No ICC-ASP/1/3 (adopted 9 September 2002).

Moreover, their Terms of Reference provide that in order to propose a new crime they must consider whether it can be characterized as one of the most serious crimes of concern to the international community as a whole and, again, if the crime is based on an existing prohibition under international law.[215]

Here we must zoom out and recall that the whole purpose of this study is to point out that there are significant legal lacunas when it comes to AWS because ICL and criminal law, in general, have been envisioned by humans for humans, and that these must urgently be addressed in order for the legal void not to be understood as a lack of proscription.

Although one could effectively argue that the core legal values inherently endangered by the nature of AWS when deployed are already deeply entrenched in and safeguarded by the international legal order. For instance, one can allude to the afore developed Martens Clause,[216] the right to life, integrity, protection, legal certainty, need for accountability and to reparations as well as the prohibition to conduct indiscriminate and disproportionate attacks, amongst others, and as a consequence, conclude that there is an existing prohibition from violating any of these core legal values.

It can also be effectively argued that what is of concern is the enforcement of the Geneva Conventions, although these are in principle technologically neutral inasmuch as they prohibit a result (*e.g.*, unlawful killing of civilians) and the means used are immaterial.[217]

---

[215] *Strengthening the International Criminal Court and the Assembly of States Parties*, Doc No ICC-ASP/11/20, Annex II, *Terms of Reference of the Working Group on Amendments*, at 9 (adopted 21 November 2012).

[216] *See above* in: Applicability of the Martens Clause; *see also* Davison, *AWS under IHL, supra* note 44, at 8.

[217] McDougall, *AWS and Accountability, supra* note 156, at 27.

In other words, the argument here is that the unique nature of AWS – *i.e.,* their unpredictability - requires criminalization, not because AWS are unlawful *per se*, but because the deployment of AWS risks an increased non-compliance with the IHL rules aimed at protecting non-combatants due to the enforcement problem.[218] Basically what this contention aims at is to prohibit all conduct that can lead to an unwanted (and criminal) result for which commission no criminally responsible individual could be identified, thus rendering it, *a priori*, virtually impune.

There is a view that if there are recognizable war crimes, there must be recognizable criminals.[219] However, when it comes to international criminal law as currently applied by the international courts and tribunals, these core values have often been required to be explicit in order to attach individual criminal responsibility to a certain individual.

Given the stringent nature of criminal law in general, abiding by the maxim *nullum crimen, nulla poena sine praevia lege,*[220] AWS related conducts should be expressly prohibited before they are criminalized. Therefore, this is ultimately not entirely a legal issue but also a matter of policy.

The Rome Statute Regime

Taking a look at the international crimes as defined in the Rome Statute, we could anticipate that AWS could end up involved in the commission of at least the following: Article 6 genocide, Article 7 crimes against humanity, Article 8 war crimes, Article 8 *bis* aggression.[221]

---

[218] *Id.,* at 28.

[219] *See generally* Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations* (2015).

[220] *See generally* Claus Kreß, *Nulla poena nullum crimen sine lege,* Max Planck Encyclopedia of Public International Law (2010) available at: <*https://www.legal-tools.org/doc/f9b453/pdf/*> accessed 10 February 2021.

[221] Rome Statute of the International Criminal Court, Arts. 5-8 *bis,* (1998) *[hereinafter:* Rome Statute].

It could seem that the most obvious crime could be war crimes, but it is the view of the author that special focus should be placed on the crime of aggression, as will be explored in a subsequent part of this broader study. In any case, a full-fetched deployment could very well be used for genocide and crimes against humanity, especially due to the targeting functions of AWS.

Nonetheless, this can only be a hypothetical exercise due to the unpredictable nature of AWS and the fact that the full capabilities of these technologies are yet to be known.

The basis for individual criminal responsibility hinges on two basic factors, a guilty criminal state of mind *(mens rea)* coupled with wrongful action *(actus reus)* of the perpetrator.[222] The latter comprises the objective elements of the crime - such as the illegal conduct of the perpetrator (be it an act or an omission), its consequences, the causation link between the conduct and the consequence, and sometimes, specific circumstances related to the context, subject or object of the crime or its modalities.[223]

In respect to the former, the idea of punishing only those with a guilty mind is derived from notions of natural justice and human rights[224] dating back to two centuries ago. In 1819,

---

[222] *See* Jones & Powles, *International Criminal Practice, supra* note 199, at 414-24; Mohamed Badar, *The Concept of Mens Rea in International Criminal Law: The Case for a Unified Approach,* at 234-52 (2013); Andri Klip Goran Sluiter, *Annotated Leading Cases of International Criminal Tribunals: The International Criminal Tribunal For The Former Yugoslavia,* 321 (2001); Jose Doria et al., *The Legal Regime of The International Criminal Court: Essays in Honour of Professor Igor Blishchenko, at* 144 (2009); Iryna Marchuk, *The Fundamental Concept of Crime in International Criminal Law: A Comparative Law Analysis, at* 134 (2013); Beatrice Bonaft, *The Relationship Between State and Individual Responsibility for International Crimes,* 247 (2009); Trial of Bruno Tesch et al., (Zyklon B Case), UNWCC, Case Number 9, British Military Court (1946), *In* Law Reports Of Trials Of War Criminals 93-104 (1949).

[223] Carsten Stahn, *A Critical Introduction to International Criminal Law,* Cambridge University Press, at 22 (2019).

[224] *See* William Cobbett, *Cobbett's Parliamentary History of England: From The Norman Conquest, In 1066 To The Year 1803*, at 1079 (1819).

Bagshaw stated that the conception that "no man ought to be punished, except for his own fault" is a clear maxim of natural justice.[225]

Accordingly, one must look at the *actus reus* and *mens rea*, as well as the specific contextual elements of each ICC core crime in order to ascertain whether or not they could be compatible with potential crimes committed by or with the use of AWS. It is noted that these crimes can be committed by various conducts or modalities, however, these are some general elements required for each of the crimes to be materialized:[226]

1. Genocide

Actus reus:

– The specific conduct must be directed against one or more persons belonging to a particular national, ethical, racial or religious group (protected group).

Mens rea:

– The perpetrator intends to commit the act, cause its effects or is aware that it will occur in the ordinary course of events.
– The perpetrator is aware that his conduct took place in the context of a manifest pattern of similar conduct directed against that protected group or was conducted that could itself effect such destruction.

Dolus specialis:

– More specifically the perpetrator intended to destroy, in whole or in part, that national, ethnic, racial or religious group, as such.

[225] *Id.*

[226] International Criminal Court (ICC), *Elements of Crimes,* Doc. No. ICC-ASP/1/3 and Corr. 1 [*hereinafter:* ICC, *Elements of Crimes]*; *see also* ICC, *Official Records of the Review Conference of the Rome Statute of the International Criminal Court, Kampala,* RC/11 (2010).

2. Crimes Against Humanity

Actus reus:

– The conduct (attack) must be directed against one or more persons belonging to a civilian population.

Mens rea:

– The perpetrator intends to commit the act, cause its effects or is aware that it will occur in the ordinary course of events.
– The perpetrator knew that the conduct was part of or intended the conduct to be part of a widespread or systematic attack against a civilian population.

Contextual elements:

– The conduct was committed as part of a widespread or systematic attack directed against a civilian population.

3. War Crimes

Actus reus:

– Article 8(2)(a): Grave breaches of the Geneva Conventions of 1949.
– Article 8(2)(b): Other serious violations of law and customs applicable in international armed conflict.
– Article 8(2)(c): Serious violations to common article 3, *i.e.*, the specific acts against persons taking no active part in the hostilities.
– Article 8(2)(e): Violations of the law and customs applicable in non-international armed conflicts.

Mens rea:

– The perpetrator intends to commit the act, cause its effects or is aware that it will occur in the ordinary course of events.

– The perpetrator is aware of factual circumstances that established the existence of an armed conflict.[227]

Contextual elements:

– For Crimes under article 8(2)(a)&(b): The conduct took place in the context of and was associated with an international armed conflict (IAC).

– For Crimes under article 8(2)(c)&(e): The conduct took place in the context of and was associated with a non-international armed conflict (NIAC).

4. Crime of Aggression

Actus reus:

– The act of aggression is the use of armed force by a State against the sovereignty, territorial integrity or political independence of another State, or in any other manner inconsistent with the Charter of the United Nations.[228]

Mens rea:

– The perpetrator intends to commit the act, cause its effects or is aware that it will occur in the ordinary course of events.

[227] There is no requirement for a legal evaluation by the perpetrator as to the existence of an armed conflict or its character as international or non-international, it only requires awareness of the factual circumstances that established the existence of an armed conflict.

[228] The crime of aggression has no different modalities for its commission.

– The perpetrator was aware of the factual circumstances that established that such a use of armed force was inconsistent with the UN Charter, as well as a manifest violation of said Charter.

Moreover, in terms of the *mens rea*, the Statute allocated in Article 30 a blanket provision applicable in addition to each crime's specific mental element requirements. It states that a person can be held criminally responsible and liable for punishment only if the material elements of the crime were committed with *intent* and *knowledge*. On the one hand, it defines "intent" whereby the person means to engage in a conduct or cause a consequence or is aware that it will occur in the ordinary course of events. On the other hand, it construes "knowledge" as awareness that a circumstance exists, or a consequence will occur in the ordinary course of events.[229]

The fact that the above blanket provision effectively constitutes an additional requirement[230] has been widely questioned as it has been reflected as setting a higher threshold and thus –unfairly to some– limiting the scope of persons that may be held accountable under the Statute.

Modes of Responsibility

After looking into the *ratione materiae* elements of possible crimes committed by or with the use of AWS, we must look at the *ratione personae* factors necessary in order to establish individual criminal responsibility as we know it.

---

[229] Rome Statute, art. 30; *see* International Criminal Court (ICC*), Prosecutor v. Thomas Lubanga Dyilo,* Decision on the Confirmation of Charges, at 350-352, (7 February 2007).

[230] ICC, *Prosecutor v. Thomas Lubanga Dyilo,* Judgment pursuant to Article 74 of the Statute, at 1014-1018 (5 April 2012).

The person who commits the crime is the perpetrator[231] and there can be different (co)perpetrators of the same crime provided that the actions of each person satisfy the requisite elements of the crime in question.[232]

Thus, we must now look into the *status quo* applicable provisions of the Rome Statute, *i.e.*, Articles 25 on modes of responsibility and 28 on command responsibility.

Let us begin by dissecting Article 25 on individual criminal responsibility.[233] First of all, one must note it delimits the jurisdiction of the Court over natural persons, thus excluding *ab initio* the possibility of attributing responsibility to the technology itself.

Secondly, the person shall be individually responsible and liable for punishment if that person:

- Commits such a crime, whether as an individual, jointly with another or through another person, regardless of whether that other person is criminally responsible;
- Orders, solicits or induces the commission of such a crime which in fact occurs or is attempted;
- For the purpose of facilitating the commission of such a crime, aids, abets or otherwise assists in its commission or its attempted commission, including providing the means for its commission;

---

231 'Perpetrator, n', *Oxford English Learner Dictionaries* available at: *<https://www.oxfordlearnersdictionaries.com/definition/american_english/perpetrator#:~:text=a%20person%20who%20commits%20a,bring%20the%20perpetrators%20to%20justice.>* accessed 3 February 2021.

232 *See* ICTY, *Prosecutor v. Kunarac* et al., Trial Judgement, at 390 (22 February 2001); International Criminal Tribunal for Rwanda (ICTR), *Prosecutor v. Kayishema & Ruzindana*, Appeal Judgement, 187 and 192 (1 June 2001); ICTY *Prosecutor v. Krstic*, Trial Judgement, at 601 (1 August 2001).

233 Rome Statute, art. 25; *see generally* Kai Ambos, *Article 25, Individual Criminal Responsibility,* in Otto Triffterer (ed.), *Commentary on the Rome Statute of the International Criminal Court: Observers' Notes, Article by Article*, 2nd Edition (2008).

– In any other way contributes to the commission or attempted commission of such a crime by a group of persons acting with a common purpose. This contribution must be intentional and shall either be made with the aim of furthering the criminal activity or purpose of the group, or be in the knowledge of the intention of the group to commit the crime;
– In respect of the crime of genocide, directly and publicly incites others to commit genocide;
– Attempts to commit such a crime by taking action that commences its execution by means of a substantial step, but the crime does not occur because of circumstances independent of the person's intentions.

Further, the crime of aggression can only be committed by a person who is effectively in a position to exercise control over or to direct the political or military action of the State which committed the act of aggression.

Moreover, the Statute clearly states that no provision relating to individual criminal responsibility shall affect the responsibility of States under international law.

Additional to those of Article 25, the Statute provides for an important different mode of liability in Article 28, which is known as command responsibility and will be addressed in detail below.[234] This modality is applicable to military commanders or persons effectively acting as a military commander for crimes committed by forces under their effective command and control, or effective authority and control as the case may be, as a result of their failure to exercise control properly over such forces.

The distinction between the various modes of liability is of paramount importance when it comes to sentencing.[235]

---

[234] *See* Command Responsibility.

[235] Jones & Powles, *International Criminal Practice, supra* note 199, at 414-415.

In Concreto: *Who Could Be Responsible for Doing What?*

It is a fact that the unique nature of AWS necessitates the involvement of diverse actors in the different stages of their development, evaluation and throughout their use until final deployment.[236]

In this sense, the author argues that these actors could be broadly grouped into three categories:

- Creators (manufacturers, developers, roboticists, programmers);
- Users (commanders, soldiers or civilian operators);
- Authorizers (civilian and military leaders).

First of all, it is paramount to stress that in congruence with the principles of accountability observed by international law, the responsibility of one person does not affect the responsibility of another.[237]

Accordingly, the fact that a manufacturer can bear a certain criminal responsibility does not exclude the end users from bearing a different type of criminal responsibility[238] however they are not necessarily dependent on each other.

---

[236] Heyns, *Report on Extrajudicial, Summary or Arbitrary Executions, supra* note 148, at 79.

[237] Rome Statute, art. 25(4); ICTY, *Prosecutor v. Tadic*, Appeals Judgment, at 227-29 (15 July 1999), the Chamber describes the elements that need to be satisfied for aiding and abetting.

[238] ICTY, *Prosecutor v. Delalić*, Appeals Judgment, 182 (20 February 2001); *see also* Grodzinsky, Frances et al., *Moral Responsibility for Computing Artifacts, "the Rules" and Issues of Trust*, Computer Science & Information Technology at 16 (2012) [*hereinafter:* Grodzinsky et al., *Moral Responsibility for Computing Artifacts*]; rule 2 provides "[t]he shared responsibility of computing artefacts is not a zero-sum game. The responsibility of an individual is not reduced simply because more people become involved in designing, developing, deploying or using the artifact".

Traditionally, for the purposes of holding a combatant responsible for war crimes, IHL and ICL are not concerned about the manufacturer of the weapon they used.[239] It is concerned about the bearer of the weapon, the one who chose to use that particular weapon or who ordered or authorized its deployment.[240] The reasoning behind this is that the combatant is the individual who is effectively in control of the weapon and also who makes the choices regarding which weapon to use.

For the end user (the person deploying the weapon) the golden and most basic rule is that they must never use a weapon which effects they cannot control.[241] The combatant or fighter must only use those weapons that do not obfuscate their responsibilities under international law.[242] These same constraints also apply to leaders who are responsible for making the decisions of which weapons to authorize for deployment by their armed forces.

Of course, in an additional way, manufacturers can certainly be co-perpetrators, aiders or abettors of the crime if the requisite conditions are fulfilled.

Also, those in leadership positions who authorize their development, use and/or deployment must also bear another form of criminal responsibility for their catalyst participation. It is clear that in order to be just, each one of these actors must be responsible in their own capacity.

---

[239] *See* Chengeta, *Accountability Gap, supra* note 155, at 35.

[240] *See* API, art. 75(4)(b); APII, art. 6(2)(b); GC IV, art. 33; Hague Convention (IV) of 1907 Respecting the Laws and Customs of War on Land and its Regulations, art. 50; Henckaerts & Doswald, *Customary IHL,* supra *note 125,* Rule 102; Weizmann et al., *AWS under International Law, supra* note 155, at 25; note the critic in Heyn's approach "for violating the fundamental principle that no penalty may be inflicted on a person for an act for which he or she is not responsible".

[241] API, art. 51(4).

[242] *Id.*

Users: Command Responsibility

In the view of some scholars, Article 28 of the Rome Statute on command responsibility is the best suited to deal with operators of AWS since commanders are the closest actors to exercise "effective command and control" as required by this liability mode.[243]

However, it is the firm opinion of the author that this is neither an adequate nor desirable solution.

Command responsibility is a criminal modality that has become part of customary international law[244] and is widely considered as a tool to reinforce deterrence and prevent impunity.[245] Command responsibility allows commanders to be held accountable for the actions of their subordinates for the failure to prevent or punish the commission of crimes by such subordinates.[246]

In IHL and ICL alike, a commander has been understood to be a natural person exercising authority over natural persons in a military operation.[247] Likewise, the fact that Article 28 of the Rome Statute uses terms such as "forces" and "subordinates" who are capable of

---

[243] *See* generally Margulies, *Making Autonomous Weapons Accountable, supra* note 159.

[244] ICTY, *Prosecutor v. Delalić*, Trial Judgment, at 330-343 (16 November 1998) [*hereinafter:* ICTY, *Delalić's Judgment*]; Jones & Powles, *International Criminal Practice*, *supra* note 199, at 432-3.

[245] T. Markus Funk, *Victim's Rights and Advocacy in the International Criminal Court,* at 16 (2010).

[246] Swart, *Modes of International Criminal Liability, supra* note 206, at 88; *see* International Criminal Law Services, *Modes of liability: Superior Responsibility*. Module 10 of training materials, at 3-7, (2018) available at: <*https://iici.global/0.5.1/wp-content/uploads/2018/03/icls-training-materials-sec-10-superior-responsibility.pdf*>.

[247] Jones & Powles, *International Criminal Practice*, *supra* note 199, at 424; Michael L. Smidt, *Yamashita, Medina, and Beyond; Command Responsibility in Contemporary Military Operations,* 164 Military Law Review, at 168-69, 176 (2000) [*hereinafter:* Smidt, *Yamashita, Medina, and Beyond*].

being subjected to prosecution and punishment clearly refers to *humans* on the battlefield,[248] not machines.[249]

Consequently, command responsibility is inapplicable to those individuals deploying AWS given that no analogy may be drawn between the relationship of human superiors and their subordinates and the interactions of humans operating machines.[250] The continued referral of a person deploying AWS as a *commander* fuels the misleading impression that AWS are somewhat combatants or fighters,[251] thus adding to the anthropomorphic conceptual distortion explained earlier.[252]

Moreover, under the Rome Statute regime, in order for a commander to be held responsible for the actions of their subordinates, there are six basic elements that must be satisfied:[253]

1. Crimes under the jurisdiction of the Court were committed by armed forces;

2. The accused is a *de jure* or *de facto* military commander;

---

[248] *See* Chengeta, *Accountability Gap, supra* note 155, at 32. He compares the definition of Commander within the Rome Statute's article 28 and the definition in Customary International Law; *see cf.* Smidt, *Yamashita, Medina, and Beyond, supra* note 247, at 176.

[249] *Id; see also* Guénaël Mettraux, *The Law of Command Responsibility,* at 5-11 (2009), *[hereinafter:* Mettraux, *The Law of Command Responsibility]*.

[250] *Id.,* at 50.

[251] Docherty, *Losing Humanity, supra* note 179, at 4, 33-34, 42-43.

[252] *See above:* What Kind of Accountability Can We Expect?

[253] Rome Statute, art. 28; *see also* API, arts. 86(2), 87; ICC, *Prosecutor v. Jean Pierre Bemba Gombo*, Judgment Pursuant to Article 74, at 170 (21 March 2016) [*hereinafter:* ICC, *Bemba's Judgment]*.

3. The accused had effective control over the forces that committed the crimes;
4. The accused knew or owing to the circumstances, should have known, that the forces were committing or about to commit such crimes;
5. The accused failed to take all necessary and reasonable measures within his power to prevent or repress the commission of such crimes or submit the matter to competent authorities for investigation and prosecution; and
6. The crimes committed by the forces must have been a result of the failure to exercise control properly.

The above elements are a result of carefully refined jurisprudence of various international criminal courts and tribunals.[254] However, it is noteworthy that the caselaw on command responsibility took an unexpected turn after the controversial ICC Appeals Chamber (AC) decision in the *Prosecutor v Bemba Gombo* case[255] in which the accused was acquit-

[254] *See* ICC, *Bemba's Judgment*, *supra* note 253, at 170-213; ICTY, *Delalić's Judgment*, *supra* note 244, at 338-340; ICTY *Prosecutor v. Galik*, Trial Judgment, at 173 (5 December 2003); *see* Swart, *Modes of International Criminal Liability, supra* note 206, at 88-89.

[255] ICC, *Prosecutor v Bemba Gombo,* Judgment on the Appeal against Trial Chamber III's Judgment Pursuant to Article 74 of the Statute, at 167-171 (8 June 2018). Bemba was acquitted on appeal from a conviction for crimes against humanity and war crimes on the basis of command responsibility. The majority of the Appeals Chamber held, inter alia, that the scope of the duty to take 'all necessary and reasonable measures' is intrinsically connected to the extent of a commander's material ability to prevent or repress the commission of crimes or to submit the matter to the competent authorities for investigation and prosecution considering that '[a]n assessment of whether a commander took all "necessary and reasonable measures" must be based on considerations of what crimes the commander knew or should have known about and at what point in time'; furthermore, the majority held that 'it is not the case that a commander is required to employ every single conceivable measure within his or her arsenal, irrespective of considerations of proportionality and feasibility' since it was necessary 'to consider other parameters, such as the operational realities on the ground at the time faced by the commander' emphasizing that [t]here is a very real risk, to be avoided in adjudication, of evaluating what a commander should have done with the benefit of hindsight. Simply juxtaposing the fact that certain crimes were committed by his subordinates

ted, at least in part, due to the fact that he was a remote commander operating in a foreign country.[256]

Particularly, on the appeal judgment, the AC focused on the fifth of the above elements, whether Bemba "took all necessary and reasonable measures" within his power to prevent or repress the commission of such crimes or submit the matter to competent authorities for investigation and prosecution, and unlike the Trial Chamber, reaching the conclusion that he did.[257]

It follows that the "effective control" threshold is set to require that the commander has the material ability to prevent or repress the commission of crimes or submit the matter to the competent authorities.[258] However, as has been stated above, machines have no moral agency[259] and thus cannot be punished.[260]

In this regard, the nature of command responsibility does not allow commanders to abdicate their moral and legal obligations to determine if the use of force is appropriate in a

---

with a list of measures which the commander could hypothetically have taken does not, in and of itself, show that the commander acted unreasonably. The trial chamber must specifically identify what a commander should have done *in concreto*.'

256 *Id.,* at 170-171. On the facts, the majority held, inter alia, that 'the Trial Chamber paid insufficient attention to the fact that the MLC troops were operating in a foreign country with the attendant difficulties on Mr Bemba's ability, as a remote commander, to take measures.'

257 *Id.,* at 120-136, 184-194.

258 *Id.,* at 167*; see also* Bemba Judgment, *supra* note 253, at 183-184.

259 *See above in:* What Kind of Accountability Can We Expect?

260 Chengeta, *Accountability Gap, supra* note 155, at 11; *see* also Wagner, *Taking Humans out of the Loop, supra* note 168, at 5, 11; Asaro, *On banning AWS, supra* note 168, at 693; Kenneth Einar Himma, *Artificial Agency, Conciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to be a Moral Agent?,* 11 Ethics & Information Technologies, at 19-29 (2009).

given situation.[261] When they delegate obligations to a subordinate, they still retain the duty to oversee the conduct of that responsible human agent. Consequently, insofar as AWS are not responsible human agents, commanders cannot delegate any authority to them.[262]

Another requirement for this form of liability is evidence that the commander should have known, owing to the circumstances at the time, that crimes were about to be, or were, committed.[263] While this actually lowers the mental element requirement comparably to a negligence threshold, given the unpredictable nature of AWS, some authors have argued in favor of the potential operators putting forth the defense that to the best of their knowledge, the AWS would comply with IHL targeting norms.

However, the author counterargues that users should have —*at a minimum*— a good grasp of the AI capabilities, and therefore of the residual risk of unpredictability, well before deploying it.

Contrary to those defense arguments, the result of the unpredictability of an AWS with full or high levels of autonomy functioning in unstructured environments[264] is in fact that once it is deployed all of its eventual actions are by default attributable either to the programmer or the individual deploying it.[265] In this regard, the mere deployment of an AWS is already an exercise of sufficient control by the user.[266]

---

[261] Asaro, *On banning AWS supra* note 168, at 701.

[262] *Id.*

[263] *See* Rome Statute, art. 28*;* ICC, *Bemba's Judgment*, *supra* note 253, at 50-53, 170, 196.

[264] Chengeta, *Accountability Gap*, *supra* note 155, at 34. Chengeta notes that Michael Schmitt while defending AWS ignores the problem of unpredictability; *see generally* Schmitt, *AWS and IHL, supra* note 29.

[265] *See* Schmitt, *AWS and IHL, supra* note 29, *at* 16-17, 33.

[266] *See* Chengeta, *Accountability Gap*, *supra* note 155, at 34.

The view of the author is that if there is room for any unforeseeability in relation to the deployment of a weapon such as AWS, then it is reasonably foreseeable to expect the worst-case scenario, thereby attracting the corresponding criminal responsibility for any crimes committed.[267] In other words, whenever a crime is committed as a result of the use of AWS, it is the individual who deployed it who is criminally liable.[268]

Therefore, when they are developed, they must not be given autonomy or functions that make them cease being weapons but *de facto* robot combatants. Rather, AWS must always be developed in a manner that they remain weapons in the hands of a fighter who is liable on the basis of individual responsibility in cases where crimes are committed.[269]

Understandably, concepts of law can sometimes be adjusted to address new situations, but regarding this attempted fictional equation of AWS to combatants, the concept of command responsibility cannot be stretched so far without inherently losing its essence.

In the author's view, the only instance where the issue of command responsibility is relevant is when the commander or civilian who supervises the individual programming or deploying an AWS knew or should have known that their subordinate was programming or using an AWS in an unlawful manner and did nothing to prevent or stop it or punish them after the fact.[270] This is just the same line of reasoning as in relation to other weapons.

Consequently, command responsibility cannot be applicable to a *human-machine* relationship because there is no legal justification to allocate combatant status to AWS - they are weapons and those who deploy them are the combatants. Conclusively, from a legal per-

[267] Schmitt, *AWS and IHL, supra* note 29, at 16-18, 33.

[268] Weizmann et al., *AWS under International Law, supra* note 155, at 24-25.

[269] Sassóli, *Autonomous Weapons and IHL, supra* note 52, at 308, 324.

[270] Mettraux, *The Law of Command Responsibility, supra* note 249, at 55; Schmitt, *AWS and IHL, supra* note 29, at 33-34.

spective, AWS cannot and should not commit crimes. As Seneca observed, "a sword is never a killer, it is a tool in the killer's hands".[271]

Nonetheless, in light of the hurdles explained above, a different group of scholars has suggested a revision of the doctrine of command responsibility in order to facilitate pinning down an actual culprit. This would require lowering the mental element standard.

In this line of thought, it is argued that with 'a modest revision of the doctrine' that extends its application to the supervision of machines[272] command responsibility would apply if there was a requirement of "dynamic diligence" on the part of commanders.[273]

This approach would necessitate at least the following: a dedicated command structure, technical expertise, real-time human monitoring (including an AWS capability to request a review), periodic and frequent review of outputs, the input of dynamic parameters governing AWS use in relation to time, distance and maximum expected collateral damage, and that target selection decisions be transparent and interpretable to humans.[274]

This alternative intends to offer a more apt solution by normatively characterizing the commander as a direct perpetrator, without needing to rely on a contorted doctrine of command responsibility that equates an autonomous machine with a subordinate.[275]

---

[271] Letters to Lucilius, 1st c., cited in Michael C. Thomsett & Jean F. Thomsett (eds.), *War and Conflict Quotes,* at 158 (1997).

[272] Margulies, *Making Autonomous Weapons Accountable, supra* note 159, at 441.

[273] *Id.,* at 406.

[274] *Id.,* at 431-40; *see also* Allyson Hauptman, 'Autonomous Weapons and the Law of Armed Conflict' 218 (Winter) *Military Law Review* 194-5 (2013); ICRC, *Ethics and AWS*, *supra* note 122, at 1, 5; Stewart, *New Technology and the Law of Armed Conflict, supra* note 158, at 291-292.

[275] Similar conclusions are reached by HRW, *Mind the Gap, supra* note 158; Chengeta, *Accountability Gap*, *supra* note 155, at 31-4; Heather M Roff, *Killing in War: Responsibility, Liability, and Lethal Autonomous Robots* in

In any case, considering all of the above, despite command responsibility might be viewed by some as an attractive fix given that it would recognize the autonomy of an AWS, it is contended in this study that it is an inadequate avenue to bridge the accountability gap given the current legal challenges to meet the doctrinal criteria.

Creators: Aiding and Abetting

On the other hand, some scholars have looked into the responsibility of the AWS manufacturer, developer or programmer, arguing that they will exert greater control over not only the range of actions the weapons system is capable of performing, but over the specific actions that it, in fact, performs after being deployed.[276]

The author notes that it is more likely that the actions of these types of actors concern the domain of national courts, unless their conduct satisfies all the constitutive elements of a crime within the jurisdiction of the ICC.[277]

In that event, it is argued that Article 25(3)(c) of the Rome Statute, referring to aiding and abetting,[278] would be the best suited to deal with designers and manufacturers of AWS.[279]

Fritz Allhoff, Nicholas G Evans and Adam Henschke (eds.), *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century* Routledge, at 352, 358 (2013); Daniele Amoroso & Guglielmo Tamburrini, *Autonomous Weapon Systems and Meaningful Human Control: Ethical and Legal Issues,* Curr Robot Rep at 19 (2020) [*hereinafter:* Amoroso & Tamburrini, *AWS and Meaningful Human Control]*; Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons,* University of Pennsylvania Law Review, at 1379-81 (2016).

276 Tim McFarland and Tim McCormack, *Mind the Gap: Can Developers of Autonomous Weapon Systems be Liable for War Crimes?* 90 International Law Studies, at 366 (2014) [*hereinafter:* McFarland & McCormack, *Mind the Gap]*.

277 Rome Statute, art. 5; *see also* ICC, *Elements of Crimes*, *supra* note 226.

278 "For the purpose of facilitating the commission of such a crime, aids, abets or otherwise assists in its commission or its attempted commission, including providing the means for its commission".

279 McFarland & McCormack, *Mind the Gap*, *supra* note 276, at 376-378.

This provision establishes a form of accessory liability where intent is always required,[280] *i.e.*, with the *purpose* to facilitate the crime, as mere knowledge is not enough for responsibility under this article.[281] According to the Court, what is required is that the person provides assistance to the commission of a crime and that, in engaging in this conduct, they intend to facilitate the commission of the crime.[282] Such assistance does not need to be "substantial"[283] since the liability of accessories requires a lesser contribution than those incurring on principal liability.[284]

Concretely, "aiding" implies the provision of practical or material assistance in the form of providing the means for the commission of a crime whereas "abetting" describes the moral or psychological assistance of the accessory to the principal perpetrator, taking the form of

---

[280] In some cases, 'the intent' required for article 25(3)(c) has been established by the Court analyzing article 25(3)(d), which merely requires 'knowledge' in opposition to *inter alia* article 25(3)(c); *see for instance*: ICC, *Prosecutor v. Lubanga Dyilo*, Decision on the confirmation of charges, at 337 (29 January 2007) [*hereinafter*: ICC, *Lubanga's Confirmation of Charges]*; ICC, *Prosecutor v. Mbarushimana*, Decision on the Prosecutor's Application for a Warrant of Arrest against Callixte Mbarushimana, at 38-39 (28 September 2010); ICC, *Prosecutor v. Mbarushimana*, Decision on the confirmation of charges, at 289 (16 December 2011) [*hereinafter:* ICC, *Mbarushimana's Confirmation of charges]*.

[281] ICC, *Mbarushimana's Confirmation of charges, supra* note 280, at 274; *see also* ICC, *Prosecutor v. Ngudjolo Chui,* Judgment pursuant to Article 74 of the Statute, Concurring Opinion of Judge Christine Van den Wyngaert, at 25 (18 December 2012); she addresses the fact that the Rome Statute adds a stricter mental element for aiding and abetting (*i.e.,* the intent or purpose) than that under Article 7(1) of the ICTY's statute, which only required knowledge.

[282] ICC, *Prosecutor v. Blé Goudé*, Decision on the confirmation of charges against Charles Blé Goudé, at 167 (11 December 2014) [*hereinafter:* ICC, *Blé Goudé's Confirmation of Charges]*.

[283] ICC, *Prosecutor v. Dominic Ongwen*, Decision on the Confirmation of charges against Dominic Ongwen, at 43 *(*23 March 2016); ICC*, Prosecutor v. Al Mahdi*, Public redacted Decision on the confirmation of charges against Ahmad Al Faqi Al Mahdi, at 26 (24 March 2016).

[284] ICC, *Prosecutor v. Lubanga Dyilo*, Judgment pursuant to Article 74 of the Statute, at 997-998 (14 March 2012).

encouragement of or even sympathy for the commission of the particular offense, which does not need to be explicit.[285]

The Court has stated that this kind of assistance must have an effect on the commission of the crime, yet the contribution is not held to a specific threshold and the participation of the accessory need not be condition *sine qua non* to the commission of the principal crime. The only requirement is that the individual furthered, advanced or facilitated the commission of such crime, before, during or after the fact,[286] with the purpose of doing so.[287]

It is important to recall that this liability mode is accessorial, derivative of the main conduct of a principal perpetrator.[288] This means it is dependent on the commission, or at least attempted commission, of an offense by the principal perpetrator - albeit it is not required that the latter is identified, charged or convicted.[289]

Furthermore, it is argued that it wouldn't even be necessary to prove that there was a common plan between the manufacturer and the individual who deploys the AWS. According to earlier jurisprudence, since an aider or abettor is always an accessory to a crime perpetrated by another person,[290] no proof is required of the existence of a common concerted plan, let alone of the preexistence of such a plan.[291] The person deploying the AWS who

---

[285] ICC, *Bemba's Judgment, supra* note 253, at 88-89.

[286] ICC, *Prosecutor v. Jean Pierre Bemba* et al., Judgment pursuant to Article 74 of the Statute, at 96 (19 October 2016) [*hereinafter:* ICC, *Bemba* et al., *Judgment*].

[287] *Id.,* at 90-97.

[288] *See* ICC, *Lubanga's Confirmation of Charges, supra* note 280, at 337.

[289] ICC, *Bemba* et al., *judgment, supra* note 286, at 83-85.

[290] ICTY, *Prosecutor v. Kordić & Erkez*, Trial Judgement, at 399 (26 February 2001).

[291] *See* ICTY, *Prosecutor v. Tadic*, Appeals Chamber Judgment, at 227-9 (15 July 1999).

is the principal may not even know about the accomplice's (manufacturer or programmer's) contribution.[292]

However, the author identifies two main hurdles for prosecutions against creators as accessories.

The first is the *mens rea* requirement, for it must be proven that:[293]

a) They act with awareness of the eventual physical perpetrator's intention to commit the crime;

b) They act with the knowledge that their conduct would assist in the perpetration of the offense; and

c) They act for the purpose of facilitating said crime.

The second is that if attempting to make a war crimes charges, in most cases it will be difficult to establish the required contextual element that the creator's conduct took place in the context of, and was associated with, an armed conflict because generally their kind of contributions will be completed in the weapon's development phase, which could likely occur prior to the commencement of the relevant armed conflict.[294]

Therefore, there must be a revision on the contextual element to either explicitly include, or be interpreted as to implicitly include, acts of preparation prior to the commencement of the armed conflict provided that the completion of the crime occurred in the relevant context.[295]

---

[292] Chengeta, *Accountability Gap*, *supra* note 155, at 22.

[293] McFarland & McCormack, *Mind the Gap*, *supra* note 276, at 380; *see also* ICC, *Blé Goudé's Confirmation of Charges, supra* note 282.

[294] *Id.,* 372-4.

[295] *Id.,* 384.

Hence, accessorial liability, as it currently stands may equally not provide an appropriate framework for a satisfactory solution and therefore must also be subject to adaptations because a human must always decide how to program the system, and clearly, that individual must be held accountable for programming it to engage in actions that amounted to war crimes.[296] It has been noted that it is a creator's duty to ensure that AWS are as safe as possible to both combatant and noncombatant alike.[297]

Considerations for Revisions

As can be observed above, there is a lack of an appropriate parallel in the Rome Statute. Therefore, in order to properly bridge the impunity gap the international community must make room for legal adaptations that can encompass the operational realities, be it by incorporating an AWS specific crime or accepting an "AWS-friendly" mode of responsibility.

This can happen in two ways, either by the stretching out the existing statutory framework *via* jurisdictional interpretations or by normatively incorporating new elements.

Since the former can only happen *ex post facto*, it is imperative to mobilize the efforts necessary to materialize the latter.

In order to do so, Article 21(1)(c) of the Statute recognizes that in exceptional cases general principles of law derived from national laws of legal systems of the world could be applicable law for the Court.[298] It is the author's contention that the legal lacuna surrounding AWS

---

[296] Schmitt, *AWS and IHL, supra* note 29, at 33.

[297] Ronald C. Arkin, *Governing Lethal Behavior: embedding Ethics in a Hybrid Deliberative/ Reactive Robot Architecture*, at 9 (2011).

[298] While the Court has not often made inquiries regarding this article, it is able to do so as shown in: ICC, *Prosecutor v. Katanga*, Public Redacted Judgment on the Appeals against the Order of Trial Chamber II of 24 March 2017

would squarely merit invoking this provision.

Therefore, we must take a look at criminal responsibility models prevalent in most national legal systems in order to find the most hospitable scheme for AWS.

It is important to clarify that the two models explained below are not alternative to each other, they could be applied coordinately and simultaneously in order to create a full image of criminal liability in the specific context of AI system involvement.[299] As a result, when AWS and humans are involved directly or indirectly in the perpetration of a specific crime, it would be much more difficult to evade criminal liability.

Modes of Responsibility

Perpetration-by-Another

In most legal systems,[300] when a crime is committed by an innocent agent, *i.e.*, where a person causes a child,[301] a mentally incompetent,[302] or a person who lacks a criminal

Entitled "Order for Reparations pursuant to Article 75 of the Statute", at 148 (8 March 2018); ICC, *Prosecutor v. Lubanga*, Decision Regarding the Practices Used to Prepare and Familiarise Witnesses for Giving Testimony at Trial, at 40-1 (1 December 2007).

299 Gabriel Hallevy, *The Criminal Liability of Artificial Intelligence Entities- From Science Fiction to Legal Social Control,* Akron Law Jorunals, at 196 (2016) [*hereinafter:* Hallevy, *The Criminal Liability of AI]*.

300 *See* Fletcher, G.P. *Rethinking Criminal Law,* New York, Oxford University Press, at 639 (2000); *see also* ICC, *Prosecutor v Germain Katanga and Mathieu Ngudjolo Chui*, Decision on the Confirmation of Charges, at 495 (30 September 2008) [*hereinafter:* ICC, *Katanga's Confirmation of Charges]*.

301 Maxey v. United States, 30 App. D.C. 63 (App.D.C.1907); Commonwealth v. Hill, 11 Mass. 136 (1814); R v Michael, (1840) 2 Mod. 120, 169 E.R. 48.

302 Johnson v. State, 142 Ala. 70, 38 So. 182 (1904); People v. Monks, 133 Cal. App. 440, 24 P.2d 508 (Cal. App.4Dist.1933).

state of mind to engage the conduct,[303] that person is criminally liable as a perpetrator-by-another.[304] In such cases, the intermediary is regarded as a mere instrument and the originating actor (the perpetrator-by-another) is the real perpetrator.[305] That perpetrator-by-another is liable for the conduct of the innocent agent, and the liability is determined on the basis of the conduct[306] and their mental state.[307]

Quite a controversial approach has emerged around this modality, drawing an analogy between the bellicose use of AWS and the recruitment and use of child soldiers[308] for they are both not "conscious" agents committing a crime and thus absolved of responsibility for their participation in, or perpetration of, international crimes.[309] The parallel is drawn by the fact that although child soldiers are autonomous —perhaps even much more than AWS— they "lack full moral autonomy".[310] This vitiates their understanding of the full moral dimensions of what they do, thereby rendering child soldiers as ill-suited objects of punishment,[311] and thus ineligible for a combatant role, just as AWS.[312]

---

[303] United States v. Bryan, 483 F.2d 88 (3rd Cir.1973); Boushea v. United States, 173 F.2d 131 (8th Cir.1949).

[304] Morrisey v. State, 620 A.2d 207 (Del.1993); Conyers v. State, 367 Md. 571, 790 A.2d 15 (2002); ICC, *Katanga's Confirmation of Charges*, *supra* note 300, at 495.

[305] Hallevy, *The Criminal Liability of AI, supra* note 299, at 179.

[306] Dusenbery v. Commonwealth, 220 Va. 770, 263 S.E.2d 392 (1980).

[307] United States v. Tobon-Builes, 706 F.2d 1092 (11th Cir.1983); United States v. Ruffin, 613 F.2d 408 (2nd Cir.1979).

[308] Sparrow, *Killer Robots, supra* note 45, at 73-4; *see* Henckaerts & Doswald, *Customary IHL, supra* note 125, at 482-85 (2005), Rule 136 deals with the recruitment of child soldiers.

[309] Liu, *Refining Responsibility*, *supra* note 163, at 343-4; Davison, *AWS under IHL, supra* note 44, at 17.

[310] Sparrow, *Killer Robots, supra* note 45, at 73.

[311] *Id.,* at 73.

[312] Henckaerts & Doswald, *Customary IHL, supra* note 125, at 482-85, Rule 136.

In this respect, Article 26 of the Rome Statute indeed states that the Court shall have no jurisdiction over any person who was under the age of 18 at the time of the alleged commission of a crime. Albeit in the author's view it is a rather long shot to equate human children to AWS, at a minimum because the criminalization of the recruitment and use of child soldiers is aimed at the protection of those children,[313] not those who they might in turn harm. In this regard, it would be extremely inappropriate to grant AWS the same "innocent agent" consideration afforded to children.

Moreover, scholars advancing this notion claim that the potential void that the aforesaid creates in terms of individual responsibility is avoided by the clear prohibition of introducing child soldiers into armed conflict in the first place.[314] They further state that an individual would not be responsible for the crimes committed by the child soldiers but, rather, for having introduced them as irresponsible entities into the battlefield.[315] This line of reasoning could serve to promote the AWS specific crime proposition detailed below,[316] or at least an *ab initio* AWS prohibition posture.

In any case, the real question is who the perpetrator-by-another is. As stated above, the author considers three possible groups of persons, the creators, the users and the authorizers.

Regarding those three possibilities, the *actus reus* of the crime has been carried out by the AI system. The perpetration-by-another liability model considers the conduct committed

---

313 Liu, *Refining Responsibility, supra* note 163, at 343.

314 *Id.*

315 *Id.,* at 343-4.

316 *See below,* AWS Specific Crime.

by the AI system as if it is the programmer's, the user's or the authorizer's on grounds of its instrumental usage as an innocent agent, [317] as it is legally merely a machine.

However, it could be argued that this model finds more complexity when the AI system has not been specifically designed to commit the crime in question and committed it pursuant to its deep learning capabilities, *i.e.*, the experience or knowledge it has gained by itself.

In this sense, some scholars would argue that given that this model requires the intention of the programmers or the users to commit an offense through the AWS using some of its capabilities instrumentally, if there is room for the AI system to be considered a "semi-innocent agent" due to some degree of autonomy, this model could be challenged.[318]

The Natural Probable Consequence Liability Model

For cases in which the previous model cannot provide a suitable solution, the natural probable consequence liability model could come into play. This model could also be applicable as a response to the current state of debates, in which the "unpredictability" argument is used as an excuse for liability.

---

[317] The AI system is used as an instrument and not as a participant, although it uses its features of processing information; *see, e.g.,* George R. Cross and Cary G. Debessonet, *An Artificial Intelligence Application in the Law: CCLIPS, A Computer Program that Processes Legal Information*, 1 HIGH TECH. Law Journal, at 329 (1986).

[318] Nicola Lacey and Celia Wells, *Reconstructing Criminal Law – Critical Perspectives on Crime and the Criminal Process,* 2nd ed. at 53 (1998).

This model of criminal liability considers the scenario in which there is indeed clear involvement of the creators and/or users in the AI's functioning, and/or of the authorizers in its general use and deployment, but they did not intend to commit any offense through it, they do not know about the commission of the crime until it has already happened, nor did they plan or participate in any part of it.

In concrete terms, what this model is based upon is the ability of the creators, users and/or authorizers to foresee the forthcoming commission of the crime, holding them accountable insofar as that offense is a natural and probable consequence of that person's conduct, *i.e.,* the creation, use or authorization of the AWS in the first place. Broadly speaking, this approach entails lowering the mental element to something similar to recklessness or negligence.[319]

Traditionally, the natural probable consequence liability is used to impose criminal liability upon accomplices or negligent perpetrators.[320] With regards to the former, the established rule stated by courts and commentators is that accomplice liability extends to acts of the perpetrator which were a natural and probable consequence[321] of a criminal scheme the accomplice

[319] *See, e.g.*, Ohlin, *The Combatant's Stance, supra* note 175, at 21-23, *see also* Mc Dougall, *AWS and Accountability, supra* note 156, at 22. Mc Dougall discusses Ohlin's ideas mentioning that commanders should be prosecuted on the basis of the doctrine of indirect perpetration. Under article 25(3)(a) of the *Rome Statute*, individuals are held criminally responsible on the basis that they acted through a person, organization or organization-like entity that they controlled, such that the perpetrator's orders, which resulted in the ultimate criminal act, were carried out by the organization as a matter of course. Thus, Ohlin considers that the doctrine could be re-oriented to "shift the metaphorical language of machine to a literal case of machine liability". Furthermore, she notes that Ohlin adds this complication, equating an AWS with a subordinate solider, to address the possibility that an AWS is properly viewed as a culpable agent.

[320] Hallevy, *The Criminal Liability of AI, supra* note 299, at 184 (2016).

[321] United States v. Powell, 929 F.2d 724 (D.C.Cir.1991).

encouraged or aided.[322] This has been widely accepted in accomplice liability statutes and recodifications,[323] including the Statutes of the ICC and *ad hoc* tribunals.[324]

In relation to the latter, the natural probable consequence liability model requires the perpetrator to be in a mental state of negligence, not more.[325] The creators, users and/or authorizers are not required to know about any forthcoming crime commission as a result of their activity, inasmuch as such a commission is a natural probable consequence of their acts.[326]

A negligent person, in a criminal context, is a person who does not want or know about the crime but in a situation where a reasonable person could have known about it since the specific crime is a natural probable consequence of that person's conduct.[327] Negligence is in fact an awareness omission or a knowledge omission, not of acts.

---

[322] William M. Clark and William L. Marshall, *Law of Crimes,* 7th ed. At 529 (1967); Francis Bowes Sayre, *Criminal Responsibility for the Acts of Another*, 43 HARV. L. REV. at 689 (1930); People v. Prettyman, 14 Cal.4th 248, 58 Cal.Rptr.2d 827, 926 P.2d 1013 (1996); Chance v. State, 685 A.2d 351 (Del.1996).

[323] Hallevy, *The Criminal Liability of AI, supra* note 299, at 241-247; *see* State v. Kaiser, 260 Kan. 235, 918 P.2d 629 (1996); United States v. Andrews, 75 F.3d 552 (9th Cir.1996).

[324] Morten Bergsmo and Carten Stahn (eds.), *Quality Control in Preliminary Examination: Volume 2*, TOAEP, at 199 (2018) "[aiding and abetting] has taken on normative acceptance in international criminal law and has been included in the Statutes of all the post-Cold War international criminal courts and tribunals… In effect, the mode of aiding, abetting and accessorizing also criminalized the conduct of waging war by proxy (where proxy forces commit crimes)".

[325] Hallevy, *The Criminal Liability of AI, supra* note 299, at 183.

[326] *Id.*

[327] Robert P. Fine and Gary M. Cohen, *Is Criminal Negligence a Defensible Basis for Criminal Liability?*, 16 BUFF. L. REV. 749 (1966); Herbert L.A. Hart, *Negligence, Mens Rea and Criminal Responsibility*, OXFORD ESSAYS IN JURISPRUDENCE 29 (1961); Donald Stuart, *Mens Rea, Negligence and Attempts*, [1968] CRIM. L.R. 647 (1968).

The natural probable consequence liability model would offer an alternative mode of liability for AWS crimes predicated upon negligence when the elements of the underlying crimes require a different *mens rea*.[328] The logic behind this is that a reasonable creator, user or authorizer could have foreseen the commission of the crime, and therefore had the opportunity to prevent it at the origin stage.

When the AWS carries out the *actus reus* of the crime, the individual in question might be considered to be negligent if no crime was deliberately planned, or they might be considered fully liable for that specific crime if it derived from another crime that was deliberately planned even though it wasn't part of the original criminal scheme.[329]

This approach must however be carefully modulated as it can come to resemble what is known as strict liability. In some criminal law systems, strict liability exists when a person is liable for committing an action, regardless of what their intent was when committing the action.[330] The logic behind this is that the perpetrator's awareness of what they are doing does not negate the fact that they nevertheless carried out the conduct in question.[331]

This concept does bring about some controversy, some scholars oppose it for reasons related to the unfairness of a person being held responsible for acts beyond their intentions - or lack thereof.[332] In any case, strict liability typically results in more lenient pun-

---

328 American Law Institute, *The Model Penal Code- Official draft and explanatory notes,* at 312 (1985); State v. Linscott, 520 A.2d 1067 (Me.1987).

329 Hallevy, *The Criminal Liability of AI,* supra note 299, at 184-5.

330 *See for instance* Legal Information Institute, *Strict liability*, Cornell Law School, available at: <*https://www.law.cornell.edu/wex/strict_liability*> accessed 14 March 2021.

331 In criminal law, possession crimes and statutory rape are both examples of strict liability offenses.

332 Richard G. Singer, *The Resurgence of Mens Rea: The Rise and Fall of Strict Criminal Liability,* 30(2) Boston College Law Review, at 406-408 (1989); R. A. Duff, *The Realm of Criminal Law,* at 19 (2018).

ishments than other *mentes reae* which might mitigate the arguments of perceived unfairness.[333]

AWS Specific Crime

Given the challenges associated with the aforementioned revisions to command responsibility and accessorial liability, the author joins other scholars in contemplation of a tailor-made AWS-specific crime, that also focuses on a third group of perpetrators – leaders and decision makers. This avenue would of course be complementary to all of the above.

As briefly mentioned above,[334] a different kind of solution is to criminalize the introduction of AWS onto the battlespace on the grounds that they are irresponsible agents.[335] In this way, it is proposed that the accountability gap could be closed through the construction of a crime for the procurement or authorization of the use of AWS.[336]

Although it would seem that this line of reasoning would be more aimed towards strengthening the arguments to prohibit their use rather than finding compatibility with the current ICL framework, it is the author's view that it could provide a workable alternative. This would nonetheless be geared more towards a leadership crime, insofar as the perpetrators would need to be in such a position to have the authority to make those decisions.

---

[333] *Id.,* at 383; Laurie L. Levenson, *Good Faith Defenses: Reshaping Liability Crimes,* 78 Cornell Law Review, at 404, 433-4.

[334] *See above,* Perpetration-by-Another.

[335] Liu, *Refining Responsibility, supra* note 163, at 344.

[336] *Id.*

It is the author's view that there are two consecutive levels of responsibility derived from this approach.

The first level involves the military and/or civilian leaders responsible for procuring and fielding these weapons systems.[337] These envisioned perpetrators would be accountable in the first moment in time, for authorizing the use of AWS regardless of objectively foreseeable failures, this includes the weapon review and compliance validation process.[338]

The second would rather hold accountable the commanders who, after the first group of leaders have authorized their use, evaluate the ability of the AWS to perform the tasks assigned to it in compliance with IHL and moreover gives the authorization to deploy it for a certain operation.[339]

The above amounts to identifying an "authorizer", an individual who would be criminally responsible if they should have been aware of a substantial and unjustifiable risk of harm resulting from AWS conduct, and this would be established if, given their circumstances and knowledge, their failure to counter this risk constituted a considerable deviation from the standard of care expected to observe by a reasonable person in the same situation.[340]

---

[337] Geoffrey S Corn, *Autonomous Weapons Systems: Managing the Inevitability of "Taking the Man out of the Loop"* in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* Cambridge University Press, (2016) [*hereinafter:* Corn, *AWS]*. Geoffrey Corn refers to it as procurement responsibility, where parallels may be drawn between procurement commanders and command responsibility. McDougal, *AWS and Accountability supra* note 156, at 22.

[338] *Id.,* at 235.

[339] Neha Jain, *Autonomous Weapons Systems: New Frameworks for Individual Responsibility* in Nehal Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* Cambridge University Press, at 314 (2016) [*hereinafter:* Jain, *AWS]; see also* Ohlin, *The Combatant's Stance, supra* note 175, at 28-9. Ohlin argues that defining this crime would need to "make clear that the crime is less culpable than the other core crimes of international law".

[340] *Id.,* at 318.

These approaches can be derived from the Natural Probable Consequence Liability Model described above and rely on the argument that the person who approves the overall and specific bellicose use of the AWS, would provide a Court more elements to determine responsibility for civilian deaths caused by an unpredictable AWS than the person who was simply ordered by them to activate the AWS or the programmer who worked on its development phase.[341]

A different set of scholars have counterargued that both these approaches condition criminal responsibility on a lack of proper care in the decision to procure or deploy an AWS, and thus would not cover the feasible scenario in which all of the relevant assessments about IHL compliance are properly made, but given the unpredictable black box features of AWS, civilians still die unlawfully.

In this regard, the author notes that according to Article 49 of API, the decisive instant for the establishment of criminal accountability is the launching of the attack.[342] Therefore, in an AWS context, the decisive moment is transposed to the point in time when the human delegates potentially lethal decisions to the AWS.[343] This conclusion also rises from the mandate that belligerents may only choose weapons whose effects they can control.[344]

If there is a possibility that AWS, on account of various levels of autonomy, will act in an unpredictable way, and that unpredictability might result in the commission of crimes, then

---

341 Hallevy, *The Criminal Liability of AI, supra* note 299, at 181-185.

342 API, art. 19(1); *see also* ICRC, *Commentary on the Additional Protocol I to the Geneva Conventions*, at 1879-1882 (1987).

343 ICRC, *Ethics and AWS, supra* note 122, at 21.

344 API, art. 51(4) (c).

it is clear that, upon deployment, the combatant has no meaningful control over the weapon since they cannot limit its effects.[345]

Some scholars purport that as a result of the known unpredictability feature of AWS, it is difficult if not impossible to establish *mens rea*, therefore, diminishing the culpability of the individual deploying it.[346] Moreover, a notable objection is that this would open the responsibility window far too wide thus leading to circumstances in which the operator, commander or programmer might not be truly culpable.[347]

However, it is the author's opinion that this view is rather convenient for those who intend to deploy it, and that quite on the contrary, responsibility arises from this very fact as there can be no lawful justification to use unpredictable and/or uncontrollable weapons in the first place.

If it is to become a popular opinion that the "unpredictability problem" is an irresolvable one, this would then lead to only one possible conclusion: that AWS must be prohibited or restricted until it is possible to ensure meaningful human control.[348]

---

[345] However, Schmitt argues that "autonomous weapon systems are not unlawful *per se*. Their autonomy has no direct bearing on the probability they would cause unnecessary suffering or superfluous injury, does not preclude them from being directed at combatants and military objectives, and need not result in their having effects that an attacker cannot control."; *see* Schmitt, *AWS and IHL, supra* note 29, at 35.

[346] *See* United Kingdom Ministry of Defence, Development, Concepts, and Doctrine Centre, *The UK Approach to Unmanned Aircraft Systems,* JDN 2-11, at 510 (2011).

[347] Chengeta, *Accountability Gap*, *supra* note 155, at 15; Malik, *AWS, supra* note 159, at 634; Liu, *Refining Responsibility supra* note 163, at 326-7; Jain *supra* note 339, at 303, 320-2; Roff, *supra* note 44, at 355; Heyns, *Report on Extrajudicial, Summary or Arbitrary Executions, supra* note 148, at 80; Amoroso & Tamburrini, *AWS and Meaningful Human Control, supra* note 275, at 20-21.

[348] Chengeta, *Accountability Gap*, *supra* note 155, at 23-7, 50; Beard *supra* note 159, at 681.

In the 2014 Convention on Conventional Weapons expert meeting on AWS, the U.S. delegation suggested that "Meaningful Human Control" starts from the manufacturing of different components of AWS, continues during the programming of software and extends up to the final deployment of autonomous weapon systems.[349] Thus, there was a suggestion that in considering what "Meaningful Human Control" of AWS means, there should be a "capture [of] the full range of human activity that takes place in weapon systems development, acquisition, fielding and use; including a commander's or an operator's judgment to employ a particular weapon to achieve a particular effect on a particular battlefield."[350]

The notion of control over the weapon is central to the responsibility of the person using it and deploying it.[351] For there to be meaningful control, programming alone is not sufficient. There is a need for some form of supervision after activation. Such supervision must be in real-time. The actions of an AWS must be well within the control of a human combatant who approves targets, prevent or abort missions whenever the situation requires.[352]

In any case, the author can agree with the supporters of these approaches arguing that it would —*at a minimum*— encourage active due diligence[353] and in this way address at least some of the policy concerns related to the accountability gap, irrespective of how many prosecutions can actually be materialized.

---

[349] Closing Statement U.S., *The Convention on Certain Conventional Weapons (CCW), Informal Meeting of Experts on Lethal Autonomous Weapons Systems (2014),* U.S. Delegate closing statement, available at: *<http://www.unog.ch/80256EDD006B8954/%28httpAssets%29/6D6B35C716AD388CC1257CEE004871E3/$file/1019.MP3>* accessed 10 October 2020.

[350] *Id.*

[351] *See, e.g.,* Sassóli, *Autonomous Weapons and IHL, supra* note 52, at 324-25; ICRC, *Ethics and AWS, supra* note 122, at 11-13.

[352] Sassóli, *Autonomous Weapons and IHL, supra* note 52, at 323-25.

[353] Jain, *AWS, supra* note 339, at 319; Corn, *AWS, supra* note 337, at 235, 241. Indeed, Corn suggests that this might be 'the operational Achilles heel that results in the hesitation to pursue [AWS]'.

Is There Actually an Accountability Gap?

It is noteworthy that there are some scholars who deny altogether the existence of such a gap in the *status quo*, primarily due to a conservative (or arguably, in fact limited) understanding on the reaches of AI technology in AWS. They contend that there will always be a straightforward connection between any harm done and a human[354] according to two assumptions.

The first one being, that it is implausible that AWS could ever become truly independent from humans,[355] and the second, that even if this would be possible, they could only be deployed if their use can meet the legality criteria[356] and thus ensuring that their actions be attributed to a human at all times.[357]

---

[354] *See, e.g.,* Kelly Cass, *Autonomous Weapons and Accountability: Seeking Solutions in the Law of War,* 28(3) Loyola of Los Angeles Law Review, 1017, 1049-53 (2015).

[355] Michael Schmitt has rejected the very possibility of an unpredictable AWS, stating that robots will not 'go rogue.' Of their own is an invention of Hollywood'. Schmitt, *AWS and IHL*, *supra* note 29, at 7; *see also* Schmitt & Thurnher, *Out of the Loop, supra* note 83.

[356] McFarland & McCormack, *Mind the Gap*, *supra* note 276, at 195. He considers that using a weapon when it is impossible to take sufficient precautions is an illegal act in itself, and if they are taken, there is control bridging any accountability gap; Sassóli, *Autonomous Weapons and IHL, supra* note 52, at 324-5: "I do not think that the possession of autonomous decision-making capacity breaks the causal chain allowing attribution and responsibility, because I assume that it is always humans who define how this autonomy will function."; *cf.,* McDougall, *AWS and Accountability, supra* note 156, at 16-7, she criticizes McFarland views and notes that despite his conclusion, he identified scenarios in which it would not be possible to establish individual criminal responsibility were weapons with very high levels of autonomy to be deployed.

[357] Charles J Dunlap Jr essentially asserts that AWS could only lawfully be deployed in scenarios that would allow for accountability. He says the "belief that there can be no accountability because, in their view, autonomous weap-

It is the author's view that the first of these assumptions lacks either information or foresight on the exponential growth of technological developments[358] and that the second relates more to the legality of the weapon *per se* rather than on the accountability for the consequences of their use. Needless to say, the legality of these weapons is a critical question as well, however, its analysis involves different factors and is thus addressed as the subject matter of the IHL portion of this study.

In any case, after considering all of the above, it is the author's view that it is seemingly possible to establish individual criminal accountability in situations where the human is "in or on the loop". This expression refers to the situation in which the human is still in control of both the operation and the technology.[359]

However, the accountability gap occurs when the human is "out of the loop". This scenario entails the deployment of a lethal autonomous system that can either operate in a structured environment (such as target identification on the basis of pre-programmed criteria) or in an open and unstructured environment (equipped with a degree of learning capacity).[360] In light of this possibility, it is clear that international criminal law must be revised, preferably *ex*

ons can act 'unforeseeably' is obviously wrong because deploying a weapon that is expected to launch attacks 'unforeseeably' is itself a punishable breach of the responsibilities of commanders, operators, and the nations they represent." Charles J Dunlap Jr, *Accountability and Autonomous Weapons: Much Ado about Nothing*, 30(1) Temple International and Comparative Law Journal at 63, 71 (2016).

358 As Jonathan Tapson has written: "Until we see an AI do the utterly unexpected, we don't even realise that we had a limited view of the possibilities. AIs move effortlessly beyond the limits of human imagination... How do you prevent an AI from using such methods when you don't actually know what its methods are?"; *see* Jonathan Tapson, *Google's Go Victory Shows AI Thinking Can Be Unpredictable, and That's a Concern,* The Conversation (18 March 2016) available at: *<https://theconversation.com/googles-go-victory-shows-ai-thinking-can-be-unpredictable- and-thats-a-concern-56209>* accessed 11 March 2021.

359 Schmitt & Thurnher, *Out of the Loop*, *supra* note 83, at 276-7.

360 ICRC, *Ethics and AWS, supra* note 122, at 9.

ante, in the normative domain in order to have preventative effects, otherwise, it will come in the form of an *ex post facto* interpretation in a jurisdictional context.

The author's personal view is that the first step in closing the accountability gap is to hold accountable the leaders who make irresponsible decisions on the development and deployment of AWS. This would be the most effective solution as it would *ab initio* foreclose the existence of the gap.

In fact, holding leaders accountable is a well-entrenched international commitment, as can be observed by the inclusion of Article 27 of the Rome Statute in 1998 relating to the irrelevance of official capacity of a Head of State or Government, a member of a Government or parliament, an elected representative or any kind of government official, which shall in no case exempt such persons from criminal responsibility under this Statute.

The same undertaking was reiterated in the year 2000 by the UN Security Council in Resolution 1329 which emphasized the prosecution of leadership figures for war crimes in the context of the *ad hoc* tribunals.[361] In the author's view, leaders only have two choices, to deploy AWS only when meaningful human control can be fully assured or to not authorize their use at all. Any other outcome should certainly attract criminal responsibility for them.

Therefore, I encourage every reader to demand that our leaders make the right decision by forestalling authorizations on their unlawful or premature use or deployment, thus preventing the crimes from being committed in the first place.

Nevertheless, it is also important to remember that individual criminal responsibility arises on various levels. Ascribing criminal responsibility to political leadership and other

[361] United Nations Security Council Resolution 1329, U.N. Doc. S/RES/1329 (30 November 2000). Taking note of the position expressed by the International Tribunals that civilian, military and paramilitary leaders should be tried before them in preference to minor actors.

high-ranking figures does not preclude the responsibility of the individual or individuals involved in the development and/or final deployment of the weapon.[362] Even though international courts and tribunals concentrate on the "big fish", as a matter of policy, "small fish" still need prosecution albeit in national courts.[363]

As a final reflection, the author shares the five rules that have been developed by practical ethicists and social theorists who insist on the principle that humans cannot be excused from moral responsibility for the design, development or deployment of computing artefacts.[364]

The rules provide as follows:

Rule 1: The people who design, develop or deploy a computing artefact are morally responsible for that artefact, and for the foreseeable effects of that artefact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artefact as part of a sociotechnical system.

Rule 2: The shared responsibility of computing artefacts is not a zero-sum game. The responsibility of an individual is not reduced simply because more people become involved in designing, developing, deploying or using the artefact. Instead, a person's responsibility

362 *See* ICTY, *Delalić's Judgment*, *supra* note 244, at 1280. In the ICC context, while complementarity bars the Court to prosecute *all* perpetrators, in domestic jurisdictions the principle of *aut dedere aut judicare* is still applicable for lower-level-perpetrators.

363 Jones & Powles, *International Criminal Practice, supra* note 199, at 412-14.

364 *See* Grodzinsky et al., *Moral Responsibility for Computing Artifacts*, *supra* note 238. The rules seem to follow a suggested notion of strict liability where responsibility is fully acknowledged before an autonomous weapon system is deployed; *see also* Ronald Arkin, *The Robot Didn't Do It,* Position Paper for a Workshop on Anticipatory Ethics, Responsibility and Artificial Agents, 1 (2013), available at <*http://www.cc.gatech.edu/ai/robot-lab/online-publications/positionpaperv3.pdf.*> accessed 10 March 2021.

includes being answerable for the behaviors of the artefact and for the artefact's effects after deployment, to the degree to which these effects are reasonably foreseeable by that person.

Rule 3: People who knowingly use a particular computing artefact are morally responsible for that use.

Rule 4: People who knowingly design, develop, deploy or use a computing artefact can do so responsibly only when they make a reasonable effort to take into account the sociotechnical systems in which the artefact is embedded.

Rule 5: People who design, develop, deploy, promote or evaluate a computing artefact should not explicitly or implicitly deceive users about the artefact or its foreseeable effects, or about the sociotechnical systems in which the artefact is embedded.

Even though these are not legally binding rules, the author finds them relevant as they include all the relevant elements to serve as guidelines in the construction of a framework that can be.