Introduction

Increasing progress in the field of Artificial Intelligence (AI) and greater understanding of its many potential applications in warfare have, for some time now, given rise to much heated international debate. While drone technology has already demonstrated that modern weapons of war can be 'uninhabited' (no human operator is to be found inside the drone) and remotely controlled, current research on AI has begun investigating the feasibility of producing *autonomous weapon systems*.[1] Seducing some and scaring others, weapon systems capable of *identifying, selecting and engaging military targets without human intervention* could in the not too distant future be deployed on the battlefield.

The debate became demanding, as well as confusing, when NGOs, states, scholars and roboticists began raising questions about whether the behaviour of future AWS would be sufficiently predictable to be safe, or indeed lawful. AWS tend to evoke in the public consciousness such robots as *The Terminator*, no longer existing merely as science fiction horrors, but in reality; and with this comes the terrifying corollary that humankind could become enslaved by its own inventions.[2] It has even been suggested that if such machines were to replace human soldiers, able to gather their own data, make their own deductions and take lethal targeting decisions without human intervention, it would constitute a 'third revolution in military affairs'.[3]

[1] Alex Leveringhaus, *Ethics and Autonomous Weapons* (Palgrave Pivot 2016) 47–49; PAX, 'Slippery Slope. The Arms Industry and Increasingly Autonomous Weapons' (PAX International 2019) ch 2 <https://paxforpeace.nl/what-we-do/publications/slippery-slope>

[2] Neil M Richards and William D Smart, 'How Should the Law Think About Robots?' [2013] SSRN 1, 1–2; PW Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century* (Reprint edn, Penguin 2011); Mark A Lemley and Bryan Casey, 'You Might Be a Robot' [2019] *Cornell Law Review* 1, 7; Yonah Jeremy Bob, 'Scientists Warn AI Control of Nukes Could Lead to "Terminator-Style" War' *The Jerusalem Post* (25 December 2019) <www.jpost.com/International/Nuke-scientists-warn-AI-control-could-lead-to-Terminator-style-nuke-war-612123>.

[3] Future of Life, 'Open Letter on Autonomous Weapons' (28 July 2015) <https://futureoflife.org/open-letter-autonomous-weapons/>; Future of Life, 'An Open Letter to the United Nations

Notwithstanding the gravity of the questions prompted by the emergence of AWS, the topic has been surrounded, and to a certain extent adversely affected, by the plethora of terms coined to describe the systems, and the snowstorm of policy documents all desperately trying to come to terms with the implications of introducing lethal weapons that are, to all intents and purposes, capable of making their own decisions and acting of their own volition. Will it still be realistic to categorise them as weapons, or might they rather, as some experts have suggested, become a new and independent quasi-agent on the battlefield? Some authors are of the opinion that autonomous systems already exist; in their view, AWS are just more sophisticated weapons than any we have seen hitherto. Others sternly defend the opposing argument that, because of their capacity for machine learning and independent decision-making, AWS must inherently be unpredictable – and therefore unlawful. These advisers urge all the states concerned to call for a total, pre-emptive ban on the study, development and acquisition of such systems.[4]

The aim of this thesis is to explore and understand the challenges to the rule of International Humanitarian Law (IHL) that would surely follow the introduction of AWS into the battlespace. These 'autonomous' weapons, as a subject of study and discourse, can be approached from a number of perspectives and divergent fields of knowledge; this aspect of the matter has certainly proved to be one of the most fascinating and complex issues with which this research has had to grapple.

There is no doubt that AWS would deliver enormous advantages in the theatre of war. First and foremost, they would guarantee the physical and psychological distancing of soldiers from lethal risk on the battlefield. Second, the increasing sophistication of modern communication technology means that computers can access and process high-quality data in quantity, and at a speed that would simply be impossible for any human operator. Thus, AWS promise to be safer, faster, and more efficient than human personnel.[5] The

Convention on Certain Conventional Weapons' (21 August 2017) <https://futureoflife.org/autonomous-weapons-open-letter-2017/>.

4 Metdi Haji-Janev and Kiril Hristovski, 'Beyond the Fog of War: Autonomous Weapons Systems in the Context of International Law of Armed Conflicts' (2017) 57 *Jurimetrics* 325, 326, 330–332; Kenneth Anderson and Mathew C Waxman, 'Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation Under International Law' in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation, and Technology* (Oxford University Press 2017) 1098–1100 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978359>; Simon Chesterman, 'Artificial Intelligence and the Problem of Autonomy' (2020) 1 *Notre Dame Journal on Emerging Technologies* 211, 229.

5 The Russian Federation, '2019 Group of Governmental Experts: Potential Opportunities and Limitations of Military Uses of Lethal Autonomous Weapons Systems' (2019) CCW/GGE.1/

problems, for those who question the legitimacy of AWS, all relate to the potential consequences of employing weapon systems that are no longer supervised by human operators.[6]

The entire debate was initiated by the release of two very different documents: one published in the United States, *USA Department of Defence Directive 3000.09*, in 2012, and one in the United Kingdom, the *Joint Doctrine Publication 0–30.2 Unmanned Aircraft Systems*, in 2017.[7] While the US directive defines AWS as 'weapon systems that operate without human intervention', the UK's definition is 'weapon systems that operate without human control'. Moreover, in the 2011 directive, the *UK Approach to Unmanned Aircraft Vehicles*, the UK's policy was to regard AWS as 'self-aware' systems.[8] Thus, even a factor as fundamental as the role of human operators at the human–machine interface has been, from the beginning, subject to misinterpretation and disagreement, rather than providing common grounds for a consensual and productive debate.

In 2013, an international agreement was reached among all states privy to the Convention on Certain Conventional Weapons (CCW), which would come to be the most appropriate international forum in which states could investigate, analyse and debate the many questions and interests surrounding AWS.[9] In the same year, Special Rapporteur Christof Heyns called all states 'to declare and implement national moratoria' until an international agreement on the future of AWS was reached.[10] In 2016, at the CCW meeting it was

https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2019/gge/Documents/GGE.2-WP1.pdf.

6 Yonah Jeremy Bob, 'The Future of AI in Warfare and Counterterrorism' *The Jerusalem Post* (23 January 2020) <www.jpost.com/Jpost-Tech/The-future-of-AI-in-warfare-and-counterterrorism-615112>.

7 Anderson and Waxman, 'Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation Under International Law' (n 4) 1097–1098; US Department of Defense, 'DoD 3000.09. Autonomy in Weapon Systems' (21 November 2012) <www.hsdl.org/?abstract&did=726163>; UK Ministry of Defence, 'Joint Doctrine Publication 0–30.2 Unmanned Aircraft Systems' (2017) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf>.

8 UK Ministry of Defence, 'The UK Approach to Unmanned Aircraft Vehicles, JDN 2/11' (2011) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/644084/20110505-JDN_2-11_UAS_archived-U.pdf>.

9 Vincent Boulanin and Maaike Verbruggen, 'Mapping the Development of Autonomy in Weapon Systems' (Sipri – Stockholm International Peace Research Institute 2017) 1 <www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf>.

10 Christof Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns' (Human Rights Council 2013) A/HRC/23/47 <www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf>.

decided that a Group of Governmental Experts (GGE) should be appointed to look at all aspects of the debate, identify problematic issues and point out which questions demanded legal responses.[11] At the time of writing, the GGE has met five times: first in August 2017, and more recently in August 2019 and September 2020.[12] To date, however, little consensus has been achieved.[13] Until today, consensus was achieved in what concerns the CCW as 'the most appropriate forum to discuss emerging technologies in the area of Lethal Autonomous Weapons Systems',[14] and that 'International Humanitarian Law (IHL) fully applies to existing and emerging weapons systems and that states remain responsible and accountable for their development, deployment and use in situations of armed conflict'.[15]

Considering the proliferation of reviews, opinions and debates that have now been aired between engineering and robotics experts, NGOs, scholars and State representatives, there is still a lack of common agreement on the precise definition, the true nature and the legal status of an AWS. Most worrying of all is the question that has now been raised as to whether the deployment of AWS might, in legal terms, open the doors to a disastrous 'responsibility gap'.[16] Although these machines are expected

[11] 'Fifth Review Conference of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (CCW/Conf.V/10)' (2016) <www.unog.ch/80256EDD006B8954/(httpAssets)/B80134C5E97FB90AC125814F0047CCB1/$file/FinalDocument_FifthCCWRevCon.pdf>.

[12] Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, 'Report of the 2019 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (2019) CCW/GGE.1/2019/3, para 2 <https://undocs.org/en/CCW/GGE.1/2019/3>. The GGE Meetings in 2020, due to the COVID-19 pandemic, have only made available recording of the ten State members' meetings. '2020 Convention on Certain Conventional Weapons – Group of Governmental Experts on Lethal Autonomous Weapons Systems Recordings' (2020) <https://meetings.unoda.org/section/ccw-gge-2020_reports_10635_press-releases_10640/>.

[13] 'UN Talks on Killer Robots End in a Stalemate' (*NEWEUROPE*, 25 November 2019) <www.neweurope.eu/article/un-talks-on-killer-robots-end-in-a-stalemate/>.

[14] Israel, 'Group of Experts Meeting on Lethal Autonomous Weapons Systems Convention on Certain Conventional Weapons (CCW) 2018' (United Nations 2018).

[15] '2019 EU Statement. 5(a) An Exploration of the Potential Challenges Posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems to International Humanitarian Law' (2019) <www.unog.ch/80256EDD006B8954/(httpAssets)/EA84B3C2340F877DC12583CB003727F3/$file/ALIGNED+-+LAWS+GGE+EU+statement+IHL.pdf>; 'Final Report. National Security Commission on Artificial Intelligence' (2021) 91 <www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>.

[16] Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 *Ethics and Information Technology* 175; Marc Champagne and Ryan Tonkens, 'Bridging the Responsibility Gap in Automated Warfare' (2015) 28 *Philosophy & Technology* 125; Cortney Weinbaum, 'The Ethics of Artificial Intelligence in Intelligence

to be capable of gathering their own data, evaluating the data and selecting targets – sometimes involving lethal force – based on their own deductions, without human oversight, no one has yet found a satisfactory way of apportioning blame should IHL violations result. These are the concerns that have determined the structure of this thesis, which now, modestly, seeks to provide some clarity, articulate the questions that need to be discussed. In light of the aforementioned discussion, Chapter 1 deals with the scientific and technological elements that facilitate the design and construction of AWS: the author believes, as a necessary precursor to any discussion of the weapons themselves, that a rigorous understanding of AWS is required. Concepts such as 'dissociation of risk' and 'dissociation of communication' become vitally important to explain modern weapon systems, as they develop from those with inbuilt deterministic algorithms to a new model of machine 'deep learning'. According to this shift, algorithms are designed and programmed to 'learn' and to 'adapt' to rapidly changing circumstances, always in accordance with a pre-established goal.[17]

The concept of AWS and its specific machine–human interface is crucial to determine the status of such systems. Some scholars have argued that AWS blur the distinction between 'combatants' and 'weapons', hypothesising the advent of 'new agents' on the battlefield.[18] Such a possibility would have an enormous impact on IHL, but the understanding in this thesis is that an AWS will operate according to an algorithm designed and programmed by human operators *for the mission*. In a nutshell, AWS should be deployed to accomplish a preset human goal.

In order to understand how an *algorithm for the mission* operates, it is important to consider the steps involved in carrying out a military mission: the collection, evaluation and selection of data; the organising and planning of the mission; and, finally, the attack. Thus, it is argued that, amidst the complexity of a military operation, the *algorithm for the mission* can usefully be viewed as a tripartite structure: the *algorithm for situation assessment*, the

Agencies' (*National Interest*, 18 July 2016) <https://nationalinterest.org/blog/the-buzz/the-ethics-artificial-intelligence-intelligence-agencies-17021>.

[17] '2019 EU Statement. 5(a) An Exploration of the Potential Challenges Posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems to International Humanitarian Law' (n 15).

[18] Hin-Yan Liu, 'Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016) 327; Heyns, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns' (n 10) para 38; Lemley and Casey (n 2) 7.

selection algorithm and *the algorithm for situation management*, each algorithm being 'responsible' for a stage of the targeting process.

Once the way in which an AWS operates has been explored, the author will be in a better position to analyse the current state of the debate; this will be done in Chapter 2. The first step is to look at the different definitions of AWS that have been put forward by some of the states involved, for example Ireland, the Netherlands, the Russian Federation, the United Kingdom and the United States, and gain an understanding of how widely their approaches towards AI and AWS diverge. The second step is to examine all the arguments (legal, ethical and political) promulgated by those who advocate a total ban and/or try to prevent states from investing in further research and development and, above all, from arming themselves with AWS.

Chapter 3 will address the way in which confusion about the concept of 'autonomy' has been at the heart of the arguments for and against AWS as expressed by scholars and NGOs – and illustrated by the two policy documents mentioned earlier. But the main question that has to be addressed is 'what does "autonomy" mean when applied in the context of weapons technology?' Chapter 3 will confirm the definition of what will be called 'functional autonomy'.

An agreed interpretation of 'autonomy' will provide a platform from which to address questions relating to the status of AWS, and avoid the introduction of emotive terms such as 'quasi-agents', 'independent actors' and 'new agents on the battlefield', albeit favoured by some authors. Once the meaning of 'autonomy' has been determined, it will be proposed that AWS constitute a new type of weapon system, designed, programmed and developed to perform autonomously at all stages of a selection-making and targeting process. This will be examined in the context of what is known as the 'OOSA Loop', an adaptation of the 'OODA Loop', the well-known cycle observe–orient–decide–act, developed by military strategist and US Air Force Colonel John Boyd.

Chapter 4 deals with the long- and well-established legal principles of IHL regarding the use of weapons and the laws of targeting. Since AWS are expected to carry out their operations without human intervention, two sets of these rules are of particular relevance. First, there are those concerning the freedom of states to legitimately develop new weapons technology (covered by Articles 35 and 36 of the API). Second, there are the precautionary measures (enshrined in Article 57 API) that must in law be respected by those who 'plan and decide upon an attack', namely the Principle of Distinction and the Principle of Proportionality. Our approach will be to examine the complexities and requirements of those rules – complexities that are, in themselves,

frequently the object of contentious debate. It will indeed be a considerable challenge for their champions to provide convincing proof that AWS can comply with these rigorous and far-reaching legal requirements.

Finally, Chapter 5 will look in detail at the problems of accountability and the implications of the 'responsibility gap' alluded to earlier. In order to approach the problem, it is necessary to consider the multifarious factors that may contribute to a violation of IHL. In this connection, it will be important to clearly distinguish between situations resulting from 'system malfunction', from 'accidents' and from 'errors'. After the distinctions between these causal factors are made, it will become clear that human operators should be held responsible only for 'accidents' because these, and only these, can be traced back to human fault.

It will be posited that 'malfunctions' and 'errors' do not result from human fault, but rather from unexpected causes, and, therefore, the customary and conventional forms of individual accountability are not applicable. Malfunctions will not constitute a new problem: just like any other weapon or weapon system, AWS will be, on occasion, susceptible to technical and hardware failure, and this is an issue that has already been addressed by the academy.[19] The true novelty, and the source of all the ensuing legal complexities, is the potential for the system to make its *own errors* – situations, that is, in which an AWS, through misinterpretation of incoming data or any other cause, subverts the pre-given mission and initiates events that cause violations of IHL. In this chain of causation there is no link whatsoever with human behaviour, but with the system's 'reasoning' process. As Thompson Chengeta argues, 'the possibility of AWS acting in an unpredictable manner, they may represent an unresolved challenge as far as the establishment of the accused person's *mens rea* is concerned'.[20]

In light of the aforementioned discussion, looking at the complexity of AWS and at the situations in which those systems might be involved, it should be a legal requirement that designers, programmers, technicians and commanders take on a higher level of responsibility. If there is human fault at the level of designing, programming or maintenance that results in an IHL violation (*accident*), human operatives will have to be prepared to be held accountable. This obligation will necessitate an examination of different levels of *mens rea*

[19] Ian Henderson, Patrick Keane and Josh Liddy, 'Remote and Autonomous Warfare Systems: Precautions in Attack and Individual Accountability' in Jens David Ohlin (ed), *Research Handbook on Remote Warfare* (Edward Elgar Press 2016).

[20] Thompson Chengeta, 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law' (2016) 45 *Denver Journal of International Law and Policy* 1, 3.

to establish an appropriate link of causation between unlawful outcomes and AWS' human operators.

In cases of 'error', it will be impossible to identify any human operative to be held accountable. However, there is no reason to posit a 'responsibility gap', because in these cases State responsibility would be invoked. There are ample grounds, already entrenched in Conventional and Customary International Law, to sustain the argument that the AWS' deploying State bears direct responsibility for violations of IHL caused by system failures ('errors').

In order for the distinctions just discussed to be clear, to guarantee the transparency of the operation and to determine the appropriate form of accountability, it will be necessary for the factors according to which an AWS has selected its course of action to be clearly identifiable at the end of the operation. As will be explained, '*factual algorithms*', that is, algorithms that provide clear information about which factors the system has weighed in the balance in making its decisions, will need to be designed and installed at the time of programming the system. Only in this way will a court of law be able to ascertain with certainty *who* or *what* has caused a violation: *how* to achieve this will be the next challenge.