Weaponized Artificial Intelligence (AI) and the Laws of Armed Conflict (LOAC)– the RAiLE© Project

Morgan M. Broman[(1)], Pamela Finckenberg-Broman[(2)]

[(1)]Queensland University of Technology, Brisbane, Australia

[(2)]Charles Darwin University, Darwin, Australia

ABSTRACT

This article critically examines the area of and contributes to the knowledge around laws and policy for the emerging technology - military application of autonomous weapon systems (AWS). Some argue that any attempt to outright ban AWS is pointless, as they are considered to be in their early concept stage, and the shape or look these may take in the future are currently unknown.(Noone and Noone, 2015) The debate on AWS can be divided into three broad approaches within the literature; '*total ban*', '*wait and see*' and the '*pre-emptive*' approach. Relevant literature for the subject matter is used strive to answer the question; How do we achieve AWS/AI programming which adheres to the LOAC's intentions of the '*core principles of distinction, proportionality, humanity and military necessity"*?

Keywords: Autonomous Weapon Systems, AWS, Laws of Armed Conflict, LOAC,

International Humanitarian Law, IHL, Software, Programming

INTRODUCTION

Our human history is full of examples of new means of warfare, the introduction of chariots, cavalry, gunpowder, mines and nuclear missiles to mention a few, and according to some authors this development is implicitly assumed within the LOAC.(Reeves and Johnson, 2014) Alan L. Schuller stresses the importance of evolving the discussion around potential ramifications of increasingly autonomous AWS' from theory to practice.(Schuller, 2017) In line with Schuster this article argues that it is of vital importance today to both discuss and act upon the emergence of application of lethal force by AWS'. Designers and users need clarification of the AWS' relationship to the Laws of Armed Conflict (LOAC) (Levin Institute-The State University of New York, 2016), International Humanitarian Law (IHL)(International Committee of the Red Cross (ICRC), 2014) and the unstructured Rules of Engagement (RoE). Further, this article argues that provided with the right design parameters it is highly plausible that AWS' ultimately will prove to be at least as compliant and at times superior in many circumstances in the application of *distinction, proportionality, humanity and military necessity* in comparison to their human counterparts. Thus, the planned outcome of research based on these arguments is; to provide a platform for further development of a methodological framework for current and future innovations implementing international law into the design, application and deployment phases of new technological products, in particular for AWS'. The intent at this stage is to improve our understanding and management of new AWS technologies that can be of benefit to humanity.

From a civilian aspect the initial legal issue has been a focus on liability in relation to accidents involving autonomous entities.(Marchant and Lindor, 2012, Douma and Palodichuk, 2012) But there is another side to the ongoing development of autonomous entities, the military utilization of both civilian and military autonomous entities in armed conflicts. This article's focus on the intersection between increased deployment of semi-& fully autonomous military technology(Reeves and Johnson, 2014) and international regulations of weapon systems with related social and legal issues that leave many questions unanswered with the respect to the development and deployment of AWS.(Noone and Noone, 2015) This causes its own legal issues already in the first stage, the definition of AWS, which is inconsistent. (Conn, 2016, Horowitz and Horowitz, 2015, Etzioni and Etzioni, 2017)

The Human Rights Watch (HRW) utilize a popular 3-step listing – '1. Human- in-the-loop or semi-autonomous systems, 2. Human-on-the loop or human-supervised autonomous systems, 3. Human- out- of-the-loop or fully autonomous weapon systems.'(Human Rights Watch (HRW), 2012) But how ever we define these entities, the question remains; can we and if so how, achieve legislation to enable AWS/AI programming which adheres to the LOAC's intentions of the '*core principles of*

distinction, proportionality, humanity and military necessity'? In a search for tools enabling AWS/AI programming to include the abovementioned essential principles a natural starting point is to explore the debate on the subject matter in relevant literature.

THE DEBATE WITHIN THE AI, REGULATION AND ETHICS LITERATURE

Introduction. With the natural starting point for this article being the debate within the literature on AI, regulations and ethics, we choose to start with an arguably well-established authority on the subject matter.

Beginning in 1986 Professor Ronald Arkin published his research(Arkin, 1986) in the area of integrated system for visual data interpretation and robot navigation. In 2009 Arkin co-wrote an article with Assistant Professor Alan R. Wagner titled '*Robot Deception: Recognizing when a Robot Should Deceive*'(Wagner and Arkin, 2009). The research presented robot control software capable of deception, making the robot able to discern if and when to deceive. This research raised ethical questions around deception which the authors did acknowledge(Kite-Powell, 2012, Wagner and Arkin, 2010), and it triggered an extended discussion on the capability to insert ethics into autonomous systems. That same year Arkin published a book, strongly relating to the subject of this article, titled '*Governing Lethal Behaviour in Autonomous Robots*'(Arkin, 2009). Referring to James Canton at the Institute for Global Futures, Arkin states that autonomy for '*armed robots*' will happen, leading to the machine, hunting, identifying, authenticating a target and possibly neutralize or kill it without any human in the decision loop.(Arkin, 2009)

For the purpose of this article, it is important to note that Arkin on the subject of AWS programmable behavior always refers to documents such as LOAC and IH. This indicates that we can infer from his writings that he does not think that the decision on what or which rules for its behavior are to be programmed into the AWS lies either with the military nor with the system designers, but within international law.

Ban entirely approach. The concerns within the international community around the advent of AWS' is understandable,(2018) with the direction of research, trust and accountability being issues surrounding them.(Human Rights Watch (HRW), 2015) The 'ban' advocates, such as the Human Rights Watch (HRW) (Human Rights Watch (HRW), 2012, Human Rights Watch (HRW), 2014, Human Rights Watch (HRW), 2015), see the outright ban on AWS' as the simplest and most efficient way to eliminate issues in the areas of AWS development regarding research and development, technology and law, and policy implications.(Noone and Noone, 2015) The HRW does empathically drive the message of '*killer robots*'(Human Rights Watch (HRW), 2012) being a serious concern as they, in their view, are unable to meet either legal or essential non-legal standards for protection of civilians in times

of war.(Human Rights Watch (HRW), 2012) The argument is that the system(s) lack desirable human qualities and the ability to relate to humans as well as the capacity to apply human judgement(Human Rights Watch (HRW), 2012), traits they consider necessary for an AWS to comply with the law.(Noone and Noone, 2015) They also argue that even if the AWS could comply with IHL, moral and ethical issues must be considered. Key among these if we should allow a life and death decision to a machine with little or no human control.(International Committee of the Red Cross (ICRC), 2014) This argument carries particular weight when considering checks on causing collateral damage in a military setting.

However, they still struggle with the issue of accurate definitions. The United Nations Office for Disarmament Affairs (UNODA) in October 2017 held a session titled '*Pathways to Banning Fully Autonomous Weapons*'(Linden, 2018). During this session it was noted that the lack of clear definitions in this area makes discussions on banning of *fully* Autonomous Weapon Systems complicated. In his closing statement Mr. Camilo Serna, from Seguridad Humana en Latinoamérica y el Caribe, also commented on the lack of *'...proper definition with legal clarity make it difficult to fix this problem*.' (Linden, 2018) A US Department of Defence report from 2007 does not even contain the word 'ethics' or 'morals'. It does however bring up *'...safety concerns, including legal issues, associated with the rapid development and use of a diverse family of unmanned systems...*'.(US Department of Defense, 2007) While the HRW support the stand-point of ban on all '*fully autonomous weapons*'.(Human Rights Watch (HRW), 2012) This currently looks like an untenable position as our history shows that technological development will continue to happen, driven by market and political forces demanding technological benefits provided by inventions.(Noone and Noone, 2015)

Wait and see approach. The argument for 'wait and see' is that a lack of deployed AWS or similar systems is making it premature to conclude anything about the legality of their existence and/or if they should be banned as a matter of policy.(Schmitt, 2013) Thus, there is a lack of an advanced enough understanding to make any conclusions as to their cost v. benefit in legal, moral, and operational terms highly doubtful.(Schmitt, 2013) However, statistics from the U.S.A on numbers of deployed unmanned entities in conflicts contradicts this. In 2009, Singer an American political scientist and international relations scholar, wrote about the growing use of unmanned systems in the U.S. armed forces.(Singer, 2009) He showed that in the beginning of the 2003 invasion of Iraq, U.S. forces had only a handful of airborne drones, with the ground forces having none in a tactical sense. By 2008 U.S. unmanned airborne systems numbered 5,331, while the ground forces number of armed ground robots had reached over 12,000.(Singer, 2009) This use of unmanned vehicles is a growing trend in today's asymmetrical combat situations.

Waiting may not be a viable option today as we already see a significant impact on our daily lives from automated/autonomous and semi-automated/autonomous systems.(Noone and Noone, 2015) For instance, financial institution (transactions), corporations (Business Intelligence), utilities-water, electricity etc. (grid

management). There is already a de facto delegation of responsibility for decisions, with legal implications, to supportive computer systems leading to system support becoming system decision-making by default(Noone and Noone, 2015). Failure of these systems may at times have lethal effects for humans, e.g., overloading electricity grids and prompting blackouts during heatwaves(Singer, 2009) or aircraft landing systems.(Anderson and Waxman, 2013)

Pre-emptive legislative approach. There is limited precedence for weapon bans prior to their development. Historically most international regulation on weapon systems has come into place after their actual deployment in the field, e.g., poisonous gas(Geneva Conference for the Supervision of the International Traffic in Arms, 1925), cluster mines(Convention on Cluster Munitions (CCM), 2008) and blinding lasers(Review Conference CCW, 1995). For AWS' the work has begun with efforts to establish definitions and categorisation to be utilized in later legislation.(Schmitt, 2013) The International Committee of the Red Cross (ICRC) define an AWS as a weapon with an autonomous '*critical function*' that can 'independently select and attack targets', making the autonomous capability of the AWS targeting system in acquiring, tracking, selecting and attacking targets the key(Noone and Noone, 2015), indirectly relating to the ICRC's paradigm for military activities, the '*conduct of hostilities*'(International Committee of the Red Cross (ICRC), 2013). This option can be seen as too little, too late, as it was shown here previously, modern warfare already deploys substantial amounts of semi-autonomous robotic weapons in different roles. Some authors contend that if this development is left unchallenged the likelihood of the trend of development and use of sophisticated military hardware will only increase.(Tonkens, 2012) However, as mentioned in this paper already, history shows that this is nothing new and nothing indicates any substantial change in the near future.

AI PROGRAMMING AND LOAC'S CORE PRINCIPLES

So, how can an AI adhere to the LOAC's intentions of the 'core principles of distinction, proportionality, humanity and military necessity'? AWS' run by AI's naturally have the same inherit weaknesses of other computer technology, hardware and/or software malfunction, potential hostile hacking by outside forces if connected to a network. One serious Achilles' heel of the AWS in an armed conflict setting is information/intelligence that the system has available for tasking and deployment.(Noone and Noone, 2015) Thus, an important component of the system design, to build a reliable and LOAC compliant AWS, is to enable it to control who provides the necessary data, how is it done and how to interpret the data provided. There is a substantial methodological gap between '*data to be interpreted*' and the '*interpretation of data*'. While '*data to be interpreted'*, looks at available data, performs an interpretation and aligns a relevant action (objective) with this

information, the '*interpretation of data*', begins with a purpose (objective) and analyses the data, adjusting the interpretation to suit a purpose. This decision parameter will control how the AWS autonomously will interpret its mission in relation to pre-programmed values, such as distinction, proportionality, humanity and military necessity.

Ryan Calo, while not specifically mentioning LOAC, bring up the issue with determining which objectives and values should be applied or '*imported into the context of machines*'.(Calo, 2017) He provides the important note that certain decisions with moral and ethical values involved, such as taking a human off life support very possibly cannot be dealt with by an objectively well-designed machine.(Calo, 2017) His discussion around the use of force focuses on policymakers need to provide a '*framework for responsibility around Al and force that is fair and satisfactory to all stakeholders*.'(Calo, 2017) Including some criticism around the ethical code of conduct as developed by industry.(Calo, 2017)

Alan L. Schuller comes close in writing about the complexity surrounding the coding of AWS.(Schuller, 2017) Based on the OODA (Observe, Orient, Decide, Act) Loop(Marra and McNeil, 2013) used by the military as a model for evaluating the human decision-making process, he brings up the issue with the AI's learning ability increasing the lack of predictability of the AWS's behaviour.(Schuller, 2017) He establishes 5 principles to avoid any 'unlawful' autonomy(Schuller, 2017) – 1) Decision to kill not functionally delegated to a computer, 2) AWS may be lawfully controlled through programming alone, 3) IHL does not require human interaction with an AWS prior to lethal kinetic action, 4) Reasonable predictability is required only with respect to IHL compliance, dependant of specific fragments of the OODA loop granted to the AWS, 5) Limitations imposed on an A WS may compensate for performance shortfalls.(Schuller, 2017) These underlaying principles around AWS programming may provide guidance for high-level system design. Schuller also states that AI's learning capability using factored and structured representations of its surroundings is crucial to the AWS for actions such as navigation, object recognition, and fire-control which depends on what the AWS's purpose and objectives are.(Schuller, 2017)

CONCLUSIONS

This article argues that it is highly plausible that AWS', provided with the right design parameters, ultimately will prove to be at least as compliant and at times superior in many circumstances in the application of the LOAC core principles of *distinction, proportionality, humanity and military necessity* evaluation in their application of LOAC than their human counterparts. For this work the concept of '*artificial agents*' for the AWS' can be a good tool, as defined by Luciano Floridi and J. W. Sanders, utilizing three important key features of the agent - interactivity, autonomy and adaptability.(Floridi and Sanders, 2004) Indeed, the interaction of humans of

automated devices has become so common that today there exist case law *'…that anticipates the legal principles that may come to govern displacement of human activity by intelligent artifacts.'*(Wein, 1992) From this case law we learn that robots and AI are considered '*mindless*' and hence to have no will of their own.[1]

Bonnie Docherty wrote an article in The Guardian in 2018, titled '*We're running out of time to stop killer robot weapons'*(Docherty, 2018), in which she states that; '*Legally, the so-called "killer robots" would lack human judgment, meaning that it would be very challenging to ensure that their decisions complied with international humanitarian and human rights law. For example, a robot could not be preprogrammed to assess the proportionality of using force in every situation, and it would find it difficult to judge accurately whether civilian harm outweighed military advantage in each particular instance.*'(Docherty, 2018) The problem with this statement is that it applies in equal measure to the *human lethal autonomous weapon system*. The individual human soldier in the field does not always know or understand, due to lack of information, either the '*why*' or the consequences of a certain action taken or decision made by superior officers in regards to hers/his own activity in an armed conflict. In that regard the human acts as an agent for the superior officer, just like could be the case with an AWS in the same situation.

Docherty further emphasise the importance of responsibility; *'…who would be responsible for attacks that violate these laws if a human did not make the decision to fire on a specific target? In fact, it would be legally difficult and potentially unfair to hold anyone responsible for unforeseeable harm to civilians*.'(Docherty, 2018) Fortunately, there is an existing international law, the Convention of 14 March 1978 on the Law Applicable to Agency(Hague Conference on Private International Law, 1978), which (even though it is private international law) could be adapted to apply for electronic agents. As well as in order to establish the responsibility for the AWS' action we could look at the military chain-of-command concept as a parallel to the human agent situation. A selection of relevant key arguments impacting this article, excerpted from those presented above, are; a) proposals and/or demands for legislation on *or* a ban of AWS' is too premature and too speculative at this stage; b) the ability/possibility of utilizing LOAC to control AWS development and future

[1]See analogously German Federal Supreme Court (Bundesgerichtshof judgement of 16 October 2012 – X ZR 37/12, 2012) paras 130, 133, 145, 147, 154 the court denied compensation for an airline ticket booked online for an "unknown" person. Accordingly, there was no valid contract due to lack of declaration of intent. An "unknown" person was not specific enough to identify as a part for the contract and even though the online system had registered the ticket for "unknown", this ticket had not been accepted and concluded by the airline. (United States of America v. Athlone Industries, Inc., 1984) The mindlessness argument is also behind the decision of (Software Solutions Partners Ltd, R (on the application of) v HM Customs & Excise [2007] EWHC 971, 2007) at paragraph 67 from the UK on restrictions of whom can act as an agent.

operations should not be underestimated; c) the development of AWS' is inevitable, in line with other military technological development, so passing on the opportunity to harness their capacity to act within the core principles of LOAC; *distinction, proportionality, humanity and military necessity*, would be irresponsible, from any perspective.

Some authors, while stating our current lack of knowledge around AWS', still argue that the research and development behind them should not be banned, as they see a potential humanitarian risk involved in a prohibition and the possibility that AWS technology possibly could become ethically preferable to alternatives.(Reeves and Johnson, 2014, Anderson et al., 2014) It may be argued that the key to develop the ongoing discussion on the potential for an international legal and development framework for AWS' should focus on the activities leading up to the final decision to '*pull-the-trigger*'. To shift the focus on the issue of the AWS' deployment in an armed conflict, approaching the responsibility issue from the perspective of the "*decision-to-deploy*" whether human or autonomous.

REFERENCES

2018. Emerging Security Issues-The Weaponization of Increasingly Autonomous Technologies (Phase III). *www.unidir.org*.

ANDERSON, K., REISNER, D. & WAXMAN, M. 2014. Adapting the Law of Armed Conflict to Autonomous Weapon Systems. *International Law Studies U.S. Naval War College,* 90**,** 386-411.

ANDERSON, K. & WAXMAN, M. 2013. Law and Ethics for Autonomous Weapon Systems-Why a Ban Won't Work and How the Laws of War Can. *Task Force on National Security and Law***,** 1-32.

ARKIN, R. 2009. *Governing Lethal Behavior in Autonomous Robots,* New York, USA, Chapman and Hall/CRC.

ARKIN, R. C. 1986. Path Planning For A Vision-Based Autonomous Robot. Amherst, MA, USA: University of Massachusetts Amherst, MA, USA ©1986.

CALO, R. 2017. Artificial Intelligence Policy: A Primer and Roadmap. *U.C. Davis Law Review,* 51**,** 399-436.

CONN, A. 2016. The Problem of Defining Autonomous Weapons. *Future of Life Institute.*

CONVENTION ON CLUSTER MUNITIONS (CCM) 2008. Convention on Cluster Munitions. *http://www.clusterconvention.org*.

DOCHERTY, B. 2018. We're running out of time to stop killer robot weapons. *In:* VINER, K. (ed.) *the Guardian.* Guardian Media Group.

DOUMA, F. & PALODICHUK, S. A. 2012. Criminal Liability Issues Created by Autonomous Vehicles. *Santa Clara Law Review,* 52**,** 1157-1169.

ETZIONI, A. & ETZIONI, O. 2017. Pros and Cons of Autonomous Weapons Systems. *Military Review***,** 72-81.

FLORIDI, L. & SANDERS, J. W. 2004. On the Morality of Artificial Agents. *Minds and Machines*, 349-379.

GENEVA CONFERENCE FOR THE SUPERVISION OF THE INTERNATIONAL TRAFFIC IN ARMS 1925. Protocole concernant la prohibition d'emploi à la guerre de gaz asphyxiants, toxiques ou similaires et de moyens bactériologiques, fait à Genève le 17 juin 1925. Web.archive.org: United Nations Office for Disarmament Affairs (UNODA).

HAGUE CONFERENCE ON PRIVATE INTERNATIONAL LAW. 1978. HCCH | #27 Convention on the Law Applicable to Agency. *Hcch.net* [Online]. Available: https://www.hcch.net/en/instruments/conventions/full-text/?cid=89 [Accessed 03 14].

HOROWITZ, M. C. & HOROWITZ, M. C. 2015. An Introduction to AUTONOMY in WEAPON SYSTEMS. United States of America: Center for a New American Security (CNAS).

HUMAN RIGHTS WATCH (HRW) 2012. Losing Humanity-The Case against Killer Robots. *In:* DOCHERTY, B. (ed.). Human Rights Watch (HRW).

HUMAN RIGHTS WATCH (HRW) 2014. Shaking the Foundations-The Human Rights Implications of Killer Robots. United States of America: Human Rights Watch.

HUMAN RIGHTS WATCH (HRW) 2015. Mind the Gap-The Lack of Accountability for Killer Robots. Human Rights Watch.

INTERNATIONAL COMMITTEE OF THE RED CROSS (ICRC) 2013. The use of force in armed conflicts-Interplay between the conduct of hostilities and law enforcement paradigms. International Committee of the Red Cross (ICRC),.

INTERNATIONAL COMMITTEE OF THE RED CROSS (ICRC) 2014. Report - Autonomous weapon systems technical, military, legal and humanitarian aspects. Geneva: International Committee of the Red Cross (ICRC).

KITE-POWELL, J. 2012. Teaching Robots To Deceive. *https://www.forbes.com*.

LEVIN INSTITUTE-THE STATE UNIVERSITY OF NEW YORK. 2016. *Law of Armed Conflict* [Online]. Globalization101-Law of Armed Conflict: Globalization101.org. Available: https://www.globalization101.org/law-of-armed-conflict/ [Accessed May 15 2018].

LINDEN, G. 2018. Pathways to Banning Fully Autonomous Weapons – UNODA. *Un.org*.

MARCHANT, G. E. & LINDOR, R. A. 2012. The Coming Collision Between Autonomous Vehicles and the Liability System. *Santa Clara Law Review*, 52, 1321-1340.

MARRA, W. C. & MCNEIL, S. K. 2013. Understanding "The Loop": Regulating the Next Generation of War Machines. *Harvard Journal of Law and Public Policy*, 1139-1145.

NOONE, G. P. & NOONE, D. C. 2015. The Debate over Autonomous Weapons Systems. *Case Western Reserve Journal of International Law*, 47, 25-36.

REEVES, S. R. & JOHNSON, W. J. 2014. Autonomous weapons: are you sure these are killer robots? Can we talk about it? *Army Lawyer*, 1, 25-31.

REVIEW CONFERENCE CCW 1995. CCW-Protocol (IV) on Blinding Laser Weapons. Geneva: International Committee of the Red Cross (ICRC),.

SCHMITT, M. N. 2013. Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics. *Harvard National Security Journal Features***,** 1-37.

SCHULLER, A. L. 2017. At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law. *Harvard National Security Journal,* 8**,** 379-425.

SINGER, P. W. 2009. Wired for War? Robots and Military Doctrine. *Joint Force Quarterly***,** 104-110.

TONKENS, R. 2012. The Case Against Robotic Warfare: A Response to Arkin. *Journal of Military Ethics,* 11**,** 149-168.

US DEPARTMENT OF DEFENSE 2007. Unmanned Systems Roadmap 2007-2032. US Department of Defense.

WAGNER, A. R. & ARKIN, R. C. Robot Deception: Recognizing when a Robot Should Deceive. 2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA),, 2009 Daejeon, South Korea. IEEE, 46-54.

WAGNER, A. R. & ARKIN, R. C. 2010. Acting Deceptively: Providing Robots with the Capacity for Deception. *International Journal of Social Robotics*.

WEIN, L. E. 1992. The Reponsibility of Intelligent Artifacts: Toward an Automation Jurisprudence. *Harvard Journal of Law & Technology,* 6**,** 103-154.