



Volume 39 No 3, September 2024

DOI: <https://doi.org/10.20473/ydk.v39i3.55188>

Fakultas Hukum Universitas Airlangga, Jalan Dharmawangsa Dalam Selatan Surabaya, 60286 Indonesia, +6231-5023151/5023252

Fax +6231-5020454, E-mail: yuridika@fh.unair.ac.id

Yuridika (ISSN: 0215-840X | e-ISSN: 2528-3103)

by <http://e-journal.unair.ac.id/index.php/YDK/index> under a Creative Commons Attribution 4.0 International license.

FAKULTAS HUKUM UNIVERSITAS AIRLANGGA



Article history: Submitted 16 February 2024; Accepted 15 June 2024; Available Online 25 September 2024.

Lethal Autonomous Weapon Systems and Automation Bias

Gökhan Güneysu

gguneysu@anadolu.edu.tr
Anadolu University, Turkey

Abstract

Autonomy in weapon systems is already a genuine concern. States try to come up with their own definitions of these systems and pay utmost effort to impose their own understanding of these systems upon other states. For a fairly high number of states barring a total ban on such weapons would be the ideal solution; however, such states that are anxious about the increase in autonomy in war-making capabilities, adopts as a second-best scenario to contain risks created by the deployment of such systems. To this end, placing them under meaningful human control emerges as an important political and legal objective. The author believes that placing autonomous weapons under human supervision, despite its initial promise, will yield negative results. This is due to the fact that humans tend rather to be too willing to follow the solutions generated by autonomous systems. First observed in other industries of civilian nature like aviation or health, automation bias has the potential to negate most if not all of supervision measures expected to ensure proper implementation of international humanitarian law.

Keywords: Lethal Autonomous Weapon Systems; Killer Robots; International Humanitarian Law; Meaningful Human Control; Automation Bias.

Introduction

Algorithmic decision-making is becoming common in many societal domains, such as the military, criminal justice system, and law enforcement, and this raises serious ethical concerns.¹ This paper aims to shed some light onto one of these concerns, which is a very important human trait with grave possible consequences. This human tendency is the so-called automation bias. First detected in civilian sectors, this bias can briefly be defined as a human inclination to depend too heavily on and to believe in information from autonomous systems, even when contradicting

¹ Denise Garcia, 'Algorithms and Decision-Making in Military Artificial Intelligence' [2023] Global Society.[1].

or differing information from other sources is available or could easily be found with the right search.²

In the upcoming years, one feels justified to take for granted that systems with different levels of autonomy will serve various objectives during armed conflicts.³ Since the autonomous weapon systems seem to be the new weapons of choice in warfare,⁴ this bias will have warfare-related impacts as well. This impact first and foremost jeopardizes a proper implementation of IHL, which is undoubtedly essential for protecting civilians and non-military objects.

This paper is an endeavor to elaborate an innate human tendency and its impact on machine-human interaction. It will first set out to present the development of debate on autonomy in weaponry. Following that, a definition will be presented for the purposes of this paper. Human-machine interaction shall be then briefly investigated. Then, the author will elaborate on the concept of meaningful human concept. Finally, the author aims to offer a sounder opinion as to how a satisfying degree of control can, if ever, be formed.

It is the fundamental stance of this paper that overly optimistic assumptions as to the reliability of autonomous weapons and the feasibility of a meaningful human control over them may prove insubstantial in the face of many factors, including automation bias. This bias carries in itself the risk of turning human operators into complete automatons. The existence of a supervising human operator who is merely accepting the automated solutions without ever bothering to question them would actually come to mean that there is no real human control left over the LAWS.⁵ This insufficient level of oversight and supervision comes in essence to mean nothing but

² Antonio Coco, 'Exploring the Impact of Automation Bias and Complacency on Individual Criminal Responsibility for War Crimes' [2023] *Journal of International Criminal Justice* mqad034. [2].

³ Filippo Santoni de Sio and Jeroen van den Hoven, 'Meaningful Human Control over Autonomous Systems: A Philosophical Account' (2018) 5 *Frontiers in Robotics and AI* <<https://www.frontiersin.org/articles/10.3389/frobt.2018.00015>>.

⁴ Garcia (n 1).[2].

⁵ Heather Roff and Richard M Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (Article 36 2016) *Briefing Paper 1* <<https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>>.

a very shaky ground for the protection of humanitarian values as well as the opening of the gates of impunity for transgressions like war crimes. Being cognizant of this alarming fact, the author believes that one safe way to secure humanity against autonomous weapons may be to introduce a comprehensive restraint or a ban on the development and deployment of such systems and weapons.

Chronological Development of the LAWS Problematique

A specter has been haunting the diplomatic negotiations and academic discussions for a long time now and it is the tormenting problem of how humanity will cope with the so-called lethal autonomous weapon systems (LAWS). This has actually been a hot topic since the beginning of the 2010's.⁶ A significant number of individuals who are very knowledgeable in robotics or international law contested the very idea of developing such systems that can decide on their own to kill human beings, showing no remorse for or even no appreciation of the consequences of their lethal solutions.

Another group of pundits defended the development of such systems on the grounds that these systems lacked the deficiencies and weaknesses human beings are prone to. In this line of thinking, LAWS embody a *supermensch on diesel* which will keep its mandate as programmed and discharge its soldierly obligations without the meddling of any humane conditions or vulnerabilities like exhaustion, rage or revenge.⁷

There is now a large body of literature on the issue with a view to especially probing into their compatibility with international humanitarian law and, to a lesser extent, international human rights law. This paper does not aim to add anything to

⁶ Peter Asaro, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 *International Review of the Red Cross* 687, 688; Anzhelika Solovyeva and Nik Hynek, 'Going Beyond the "Killer Robots" Debate: Six Dilemmas Autonomous Weapon Systems Raise.' (2018) 12 *Central European Journal of International & Security Studies*. [167].

⁷ Mustafa Can Sati, 'The Attributability of Combatant Status to Military AI Technologies under International Humanitarian Law' [2023] *Global Society* 1, 1–2; Solovyeva and Hynek (n 6). [180].

the already existing body of works on these crucial points, rather it deals mainly with automation bias and its effects in relation with autonomous weapon systems. It will endeavor to elaborate on the question of whether meaningful human control is ever attainable. This matter of control over LAWS has assumed a central position not only in diplomatic discussions but also in academic works. For the sake of this paper, the author will use the LAWS definition by the US Department of Defense.⁸ According to this, LAWS are such weapons systems, which ‘...once activated, can select and engage targets without further intervention by a human operator’, which is a commonly adopted definition.⁹

First Debates about LAWS

In 2013 the topic of LAWS first made it to the top the international political agenda and has so far kept that hot agenda status.¹⁰ Since 2014 experts have been meeting to discuss the possible legal or political reactions on behalf of states and their ramifications at large.¹¹ First kickstarted as a non-governmental meeting and then transformed into a full-blown governmental experts’ get-together in 2017,¹² these talks have been a great learning experience for the young and the uninitiated in terms of how international law-making among states proceeds. Diplomats, concerned individuals, learned academics and international NGOs have offered additional insights during these talks. These contributions not only facilitated a larger dissemination of the ideas of those individuals but also played an influential role in the improvement of the discussions held in and outside the power circles, including, among others, academic outputs. There is now a vast academic literature

⁸ (Department of Defense 2012).[13].

⁹ Thomas Weigend, ‘Convicting Autonomous Weapons?: Criminal Responsibility of and for AWS under International Law’ [2023] *Journal of International Criminal Justice* mqad037.[2].

¹⁰ Merel Ekelhof, ‘Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation’ (2019) 10 *Global Policy*. [343].

¹¹ Kenneth Anderson, Daniel Reisner and Matthew Waxman, ‘Adapting the Law of Armed Conflict to Autonomous Weapon Systems’ (2014) 90 *International Law Studies*. [386].

¹² Shane R Reeves, Ronald TP Alcalá and Amy McCarthy, ‘Challenges in Regulating Lethal Autonomous Weapons under International Law Fighting in the Law’s Gaps’ (2021) 27 *Southwestern Journal of International Law*. [101].

on what these LAWS amount to and what kind of impact they might have on international relations in the future.

There have been a series of recurring themes in these discussions, including, but not limited to, meaningful human control, applicability of the rules of the law of the armed conflict (LOAC), machine-human interaction and lastly human dignity. In the discussions that ensued, a great number of states as well as other international actors and scholars have highlighted the importance of maintaining at all times a *meaningful* human control¹³ over the different stages of LAWS's functioning with a view to safeguarding the full implementation of humanitarian rules.

Human control has been an important catchword that alludes to the essential fact that international humanitarian law addresses real persons and tries to protect their interests from armed conflicts.¹⁴ Whether a fully autonomous weapon system can ever adhere to humanitarian standards as foreseen by LOAC has proven to be a thorny question, upon which countries apparently have hitherto failed to reach a consensus. One important fact stands out, and it is the fact that almost all states have referred to meaningful human control during the formal talks. This convergence in the national official positions as to human control may be a good indicator for the probable emergence of a new humanitarian law of customary nature. Yet, the process is far from being complete, nor can this probable emergence be taken for granted,

¹³ Edward Hunter Christie and others, 'Regulating Lethal Autonomous Weapon Systems: Exploring the Challenges of Explainability and Traceability' [2023] AI and Ethics <<https://doi.org/10.1007/s43681-023-00261-0>>; Armin Krishnan, 'Enforced Transparency: A Solution to Autonomous Weapons as Potentially Uncontrollable Weapons Similar to Bioweapons' in Jai Galliot, Duncan MacIntosh and Jens David Ohlin (eds), *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021).[220]. <<https://doi.org/10.1093/oso/9780197546048.003.0015>> accessed 18 July 2023; Sehoon Park, 'Analysis of the Positions Held by Countries on Legal Issues of Lethal Autonomous Weapons Systems and Proper Domestic Policy Direction of South Korea' (2020) 32 *The Korean journal of defense analysis*. [393-400]; Ilse Verdiesen, Filippo Santoni de Sio and Virginia Dignum, 'Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight' (2021) 31 *Minds and Machines*. [137 - 138].

¹⁴ Vincent Chetail, 'The Fundamental Principles of Humanitarian Law through the Case Law of the International Court of Justice' (2002) 21 *Refugee Survey Quarterly*. [211]. <<https://repository.graduateinstitute.ch/record/5034/files/Refugee%20Survey%20Quarterly-2002-Chetail-199-211.pdf>>; Jacques Meurant, 'Inter Arma Caritas: Evolution and Nature of International Humanitarian Law' (1987) 24 *Journal of Peace Research*. [237].

since there is as much agreement on this control requirement as disagreement as to its substantial content. The goal in reserving such a central position to MHC is to respond to this human-centric nature of IHL. Obviously such a control will be established in order to reduce the unpredictability of LAWS.¹⁵

Lethal Autonomous Weapon Systems: Defining the Future

Keeping an edge in military technology has been a priority for every military establishment. Political and military decision-makers have been attaching great significance to technological innovation in their war-making capabilities.¹⁶ In this vein, the United States Department of Defense has been spending huge sums of money on new military technologies with a view to safeguarding their dominant position.¹⁷ Haner and Garcia (2019, s. 331) remind that ‘global military spending on AWS and AI, narrowly defined, is projected to reach \$16 and \$18 billion respectively by 2025’. One of the most striking points of competition has been that of lethal autonomous weapons. A sizeable group of states, already with the strongest armed forces in the world, have been apparently very zealous in their quest for better capabilities in their autonomous weapons arsenals.¹⁸ The US, Russia, China, and the UK are reportedly among these states striving for a larger degree of autonomy in weapon systems. Par this course, Hoffman and others remind us that ‘in fact, very little in the broad Defense Department portfolio, is not aimed at generating more automation’.¹⁹

LAWS offer undeniable advantages for those parties that can afford to deploy them.²⁰ These advantages turn them into coveted commodities of huge interest for the willing and able states. These systems are able to scan and interpret their

¹⁵ Krishnan (n 13).[220].

¹⁶ Ekelhof (n 10).[343–348].

¹⁷ Michael T Klare, ‘The Challenges of Emerging Technologies’ [2018] Arms Control Today.

¹⁸ Krishnan (n 13).[225].

¹⁹ Robert R Hoffman and others, ‘Myths of Automation and Their Implications for Military Procurement’ (2018) 74 Bulletin of the Atomic Scientists.[255].

²⁰ Diego Badell and Lewin Schmitt, ‘Contested Views? Tracing European Positions on Lethal Autonomous Weapon Systems’ (2022) 31 European Security.[242].

environment faster than human operators can owing to the sensor technology.²¹ They can react to changes in this environment and this they do in high speed²² along with considerably improved precision in comparison to human soldiers. They reduce the financial and political costs of war-making, i.e. they may eventually open the gates of political as well as legal impunity, turning war into a more viable political option. All in all, it is, as far as this author is concerned, very understandable that some states unequivocally displayed an eagerness to further develop such systems and weapons. It must be the expectation of these states to turn this aggressive capability into a political position of strength and thus influence other states' behaviors more assertively. A group of concerned scientists and NGOs voiced their concerns about this trend and suggested a blanket ban on all autonomous weapons.²³ The 'stop the killer robots' campaign played an indubitably important role in raising awareness as to inherent dangers of such weapons.²⁴

Philip Alston was one of the first individuals to highlight these risks.²⁵ As he put it in 2010, 'the rapid growth of these technologies, especially those with lethal capacities and those with decreased levels of human control, raise serious concerns that have been almost entirely unexamined by human rights or humanitarian actors'.²⁶ It is remarkable that a plethora of roboticists, legal scholars, diplomats have been in rather unproductive talks for years now almost completely within framework of Alston's initial tabulation of problem areas. Alston demarcated with great success the problematic nature of decreasing human control over such weapons. What is also remarkable is the fact that he presented the notion of autonomy using a

²¹ Agnieszka Szpak, 'Legality of Use and Challenges of New Technologies in Warfare – the Use of Autonomous Weapons in Contemporary or Future Wars' (2020) 28 *European Review*. [118-119].

²² ICRC, 'ICRC Position on Autonomous Weapon Systems: ICRC Position and Background Paper' (ICRC 2021). [6].

²³ Elvira Rosert and Frank Sauer, 'Prohibiting Autonomous Weapons: Put Human Dignity First' (2019) 10 *Global Policy*. [370].

²⁴ Ondrej Rosendorf, Michal Smetana and Marek Vranka, 'Autonomous Weapons and Ethical Judgments: Experimental Evidence on Attitudes toward the Military Use of "Killer Robots".' (2022) 28 *Peace and Conflict: Journal of Peace Psychology*. [177-178].

²⁵ Verdiessen, Santoni de Sio and Dignum (n 13). [145].

²⁶ Philip Alston, 'Interim Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions' (United Nations Human Rights Council 2010) UN Doc A/65/32.

spectrum-like approach, which was to be adopted later by governmental experts and other representatives of a variety of NGOs.

Following in the footsteps of Alston, Christof Heyns proved to be very adamant about the far-reaching concerns that had to be raised, especially those in connection with the protection of life during armed conflicts as well as international peace, if autonomous weapon systems were to put to use in armed conflicts.²⁷ Heyns describes his *lethal autonomous robots* as ‘machines that are built upon the sense-think-act paradigm: they have sensors that give them a degree of situational awareness; processors or artificial intelligence that “decides” how to respond to a given stimulus; and effectors that carry out those decisions’.²⁸ Heyns raised the significant question of the extent to which LAWS ‘can be programmed to comply with the requirements of international humanitarian law and the standards protecting life under international human rights law’. In addition to them, accountability seems to emerge as a thorny issue for Heyns. Having taken into cognizance all the problematic areas and concerns, Heyns suggested all states establish national moratoria until an international policy on lethal autonomous robots could be devised.²⁹ Robert Sparrow is one of the thinkers that find the deployment of autonomous weapon systems deeply precarious. In his view, in order for one to be able to talk about the existence of a just war, someone must be available to be eventually held morally responsible for the deaths of real persons.³⁰ This requisite is even more acutely stressed in cases of civilian deaths. Sparrow adds ‘(t)he least we owe our enemies is allowing that their lives are of sufficient worth that someone should accept responsibility for their deaths’.³¹ Who will then be held morally or legally responsible, though, if autonomous systems are given free rides at combat roles replacing the human soldiers and enjoying the new role as the ultimate judge

²⁷ Christof Heyns, ‘Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions’ (United Nations Human Rights Council 2013) UN Doc A/HRC/23/47.

²⁸ *ibid.*[8].

²⁹ *ibid.*[22].

³⁰ Robert Sparrow, ‘Killer Robots’ (2007) 24 *Journal of Applied Philosophy*. [62].

³¹ *ibid.*[67].

who decides on which persons will live and which ones will die? In Sparrow's line of thinking, the responsibility gap thus created might be construed to come to mean the end of any ethical or humane debate (Misselhorn, 2019, s. 324). Until someone is properly designated or this designation is do-able, Sparrow labels all kinds of wars as unethical, if autonomous weapon systems are deployed during their course.³² In this respect, Rosert and Sauer³³ strongly emphasize that 'delegating the decision to kill to algorithms is inhumane and unacceptable under any circumstances'.

There are also optimists. Harald Schaub opines that 'humans are too slow, cognitively way insufficient and basically too unreliable to understand the changes in their environment and react accordingly to them, when compared with autonomous and automated systems'.³⁴

Ronald Arkin, too, is one of the most articulate supporters of the idea that, contrary to all the worrying reports and narratives, robots and autonomous weapons could actually play a crucial role in facilitating the proper implementation of humanitarian rules and principles.³⁵ Autonomous weapon systems would not display the weaknesses like anger, fatigue, rage or revenge human beings are constantly prone to.³⁶ LAWS do not have any emotions which humans do experience severe hardships in controlling.³⁷ This would, in turn, enable them to execute with no personal motives playing a role therein, flawlessly following the pre-determined algorithms that allowed or prohibited certain acts on the field.

In addition, it must be reminded that LAWS are such systems that learn from their experience and function more efficiently, with the eventual probability to go beyond their original programming.³⁸ The International Committee of the Red Cross

³² *ibid.*[74].

³³ (n 23).[372].

³⁴ Harald Schaub, 'Der Einsatz Autonomer Waffensysteme Aus Psychologischer Perspektive' (2020) 13 Zeitschrift für Außen- und Sicherheitspolitik.[337].

³⁵ Ronald C Arkin, *Governing Lethal Behavior in Autonomous Robots* (CRC Press 2009). [71].

³⁶ Ronald C Arkin, 'The Case for Ethical Autonomy in Unmanned Systems' (2010) 9 Journal of Military Ethics.[332 - 333].

³⁷ Erica Ma, 'Autonomous Weapons Systems Under International Law' (2020) 95 New York University Law Review.[1435 - 445].

³⁸ Krishnan (n 13).[219].

(ICRC), accordingly does not seem to enjoy a similar optimism and enthusiasm regarding LAWS like Arkin and others do. They warn the public that LAWS cause more risks than the advantages they may offer, due to the fact that humans are not in a position to specifically select or de-select targets when LAWS are taking part in hostilities.³⁹

The ICRC underlines among the defining features of LAWS their sensor-based capacity to select and engage targets after the initial activation.⁴⁰ From this flows a certain degree of unpredictability, which basically means that human operators and the designers of LAWS could not enjoy many reliable estimations as to where, when and whom these weapons could militarily engage.⁴¹ Yet, the underlying logic of legal rules of armed conflict seems to require a certain degree of knowledge and control so that real persons could be responsible for the breaches of law. There is an unbridgeable incongruity here between two opposing poles of complete unpredictability of LAWS and the innate legal need to hold someone eventually responsible. Maybe with a view to enabling law to function as it was originally geared to or with a view to assuaging all those concerns as to legal and ethical consequences of unfettered deployment of LAWS, many actors have thus far come up with ways and suggestions to keep these systems under constant human oversight. It was their way out in the face of pressing fears and anxiety. Meaningful human control has gained popularity within this context.

Even a short literature review enables one to notice that there are some differences as to what LAWS amount to and how humanity must go on with them, if ever. One approach embraced by some states is that these weapons are solely the weapons of the future. As far as this author is concerned, this Star-Wars approach is not a solid policy option. This approach instantly curtails any need to probe deeper into these systems, now that they allegedly do not

³⁹ ICRC (n 22).[6].

⁴⁰ *ibid.*[1].

⁴¹ *ibid.*[5]; Solovyeva and Hynek (n 6).[172].

exist. LAWS are already in use, though. This last claim about the existence of LAWS is in harmony with the ‘critical functions’ definition endorsed by ICRC. A full autonomy should not be seen as a pre-requisite for one system to be labelled as LAWS. There must be an essential analysis though, which must be run with a view to finding out whether the systems enjoy autonomy in two critical functions, targeting and engaging.⁴²

Another point of contestation would be whether these systems must be banned all together or not. In this equation there are a very high number of variables. But the fact that these weapon systems are a revolutionizing novelty in warfare calls for diligence on how we operate them. When operational issues are brought to the fore, meaningful human control assumes a central position, since it is almost ubiquitously seen as an indispensable measure to uphold and implement IHL.

Human Machine Interaction and Meaningful Human Control

Technological innovations inherently give rise to new challenges for society in general and the law, in particular.⁴³ Human-machine collaboration is a troublesome issue of an ever-increasing significance, in this digital age.⁴⁴ The subject of human-machine interaction has been a focal point in the diplomatic discussions held on LAWS.⁴⁵ It has been a matter of agreement for all those actors taking part in discussions about LAWS that the international humanitarian law, under all circumstances, will continue to find application for autonomous weapon systems. Another issue closely related to the application of humanitarian rules is the matter of the so-called meaningful human control (MHC) upon the significance

⁴² Rosendorf, Smetana and Vranka (n 24).[177].

⁴³ Susanne Beck, ‘Der Rechtliche Status Autonomer Maschinen’ (2017) 26 PJA/AJP.[183 - 184].

⁴⁴ Michał Boni, ‘The Ethical Dimension of Human–Artificial Intelligence Collaboration’ (2021) 20 European View.[182].

⁴⁵ Daniele Amoroso and Guglielmo Tamburrini, ‘The Ethical and Legal Case Against Autonomy in Weapons Systems’ (2018) 18.[5]. <<https://doi.org/10.1515/gj-2017-0012>> accessed 21 September 2023; Ángel Gómez de Ágreda, ‘Ethics of Autonomous Weapons Systems and Its Applicability to Any AI Systems’ (2020) 44 Artificial intelligence, economy and society 101953.[9].

of which there seems to exist a unanimity among the states.⁴⁶ The term, first coined by the NGO Article 36 in 2013,⁴⁷ was initially used as a warning call by Article 36 to remind states a positive international obligation to place individual attacks under meaningful human control.⁴⁸ Specifically alarmed by the UK Ministry of Defense's viewpoint that 'attacks without human assessment of the target, or a subsequent human authorization to attack, could still be legal', Article 36 contended that 'it is the moral agency that require of humans, coupled with the freedom to choose to follow the rules or not, that are the basis of the normative power of law'.⁴⁹ According to this opinion, human operators have to be the sole ruler as regarding the whole hostile engagement stage during armed confrontations.

States representatives have remarkably easily come to an understanding as to the value of an MHC over LAWS. Yet, there is a serious confusion regarding what it should amount to. The principle seems to be like a memorable motto that somehow sticks, but its practical impact is questionable at best. Moreover, this author has serious doubts on whether a meaningful human control is ever attainable. In order to delve deeper into the MHC-related debates, the author feels obliged to first scrutinize the loop-models that should facilitate a better understanding of the interaction between machines and humans. This interaction is exactly where the MHC should be ensured.

Three modalities have been constantly highlighted which also offered a very comprehensible framework to access the problems with technological and legal

⁴⁶ Rebecca Crootoof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 Temple International and Comparative Law Journal.[53].; Anja Dahlmann, Elisabeth Hoffberger-Pippan and Lydia Wachs, 'Autonome Waffensysteme und menschliche Kontrolle - Konsens über das Konzept, Unklarheit über die Operationalisierung'.[3].; Merel AC Ekelhof, 'Lifting the Fog of Targeting' (2018) 71 Naval War College Review 61.[24].

⁴⁷ Thompson Chengeta, 'Defining the Emerging Notion of "Meaningful Human Control" in Weapon Systems' (2017) 49 New York University Journal of International Law and Politics.[833-854].

⁴⁸ Tim McFarland, *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law* (Cambridge University Press 2020) <<https://www.cambridge.org/core/books/autonomous-weapon-systems-and-the-law-of-armed-conflict/09BFF6BB-5B88E34935678B5A0606A8A7>>.

⁴⁹ *ibid.*

aspects. The first one is the *human-in-the-loop* type of interaction, where the human operator continues to plan an integral role in the system.⁵⁰ She has the final say especially on critical military functions such as targeting and engagement. In this *modus operandi*, LAWS still discharge a high number of its functions autonomously. Yet, in the engagement phase, i.e. where whether the designated target is to be attacked or not is to be decided, the final say falls in the decision-making sphere of the human operator.

In a slightly altered interaction mode, *human-on-the-loop* interaction preserves a supervisory position for the human operator.⁵¹ The primary task originally reserved for the human operator is to monitor the activities of the autonomous weapon system.⁵² Here, the final decision about whether or not attack the potential targets normally belongs to the autonomous weapon system. However, this remains only to be the case until the human operator decides to override that autonomously-rendered decision. The difference between this and the *in-the-loop* modality is that humans are allocated solely an overriding authority in the latter one, whereas the former interaction mode dons the sole decision-making power to human operators.

In the *on-the-loop* mode, system is allowed to function unimpededly, as it is geared to within its predetermined algorithmic boundaries, unless the human operator decides to interfere. Override function, out of necessity, has to take place in a very short time span. As reported in a fratricide case, a patriot crew, that initially operated the air defense system autonomously was granted, as per its algorithm, a ten second interval to override the autonomously generated solution to attack an approaching vehicle.⁵³ The crew refrained from overriding the decision and this decision proved later to be fatal, since a friendly fighter jet was downed by the

⁵⁰ Leila Methnani and others, 'Let Me Take Over: Variable Autonomy for Meaningful Human Control' (2021) 4 *Frontiers in Artificial Intelligence* 2 <<https://www.frontiersin.org/articles/10.3389/frai.2021.737072>>.

⁵¹ *ibid.*

⁵² Caitlin Mitchell, 'When Laws Govern LAWS: A Review of the 2018 Discussions of the Group of Governmental Experts on the Implementation and Regulation of Lethal Autonomous Weapons Systems' (2020) 36 *Santa Clara High Technology Law Journal*. [412].

⁵³ Hoffman and others (n 19).

system.⁵⁴ As can be seen there is a relatively well-documented human tendency at work in similar scenarios where human beings only too happily accept the solutions offered by the computerized systems.

The third possible interaction type is the one which places the human totally *out of the loop*.⁵⁵ In this mode, the system goes on a totally autonomous mode and decides about its functions without any chance for outsiders to interfere or override. There is an easily observable school of thought that sees human control on LAWS as an indispensable requisite for the avoidance of legal transgressions. In-the-loop and on-the-loop modalities of human-machine interaction, correspondingly, are deemed to be appropriate choices to realize this humane goal.

In an ideal setting, a human operator with adequate training and sufficient time to successfully analyze their environment, could run or supervise a LAWS with flying colors, guaranteeing concomitantly due implementation of humanitarian rules.⁵⁶ This outcome though, is dependent upon the existence of a number of prerequisites like the training component, and a relatively relaxed theatre of war creating no time pressure for the human operator. In addition to these *positive* prerequisites, it must be added that all the components of LAWS should function *bug-free* and the essential ontological problem for the LAWS of brittleness must have somehow been defeated. Yet, in essence, the very assumption of such a care-free and efficient operator is the crux of this problem, since the armed conflict can always come up with that now-proverbial fog of war and will create other such instances, in which the need for a quick decision could be pressing. In these moments of desperation with stakes so high, it is extremely unrealistic to expect the human operator to keep a balanced, sane and reliable watch over the multitude of inflowing variables about the LAWS and the battlefield.⁵⁷ Now that the only decision aid she can benefit from is an autonomous system and that she has about a few seconds to render a critical

⁵⁴ Coco (n 2).[2].

⁵⁵ Rosendorf, Smetana and Vranka (n 24).[178].

⁵⁶ Crootof (n 46).[56].

⁵⁷ Nathan Gabriel Wood, 'Autonomous Weapon Systems and Responsibility Gaps: A Taxonomy' (2023) 25 Ethics and Information Technology.[16].

decision, the human operator *in* or *on* the loop would develop only too welcoming an attitude about the suggestions of the system concerned. This is the very instance automation bias kicks in. Automation bias has got to do with the tendency across many sectors of the society toward ever-more autonomous algorithmic decision-making, where the algorithms lead and people do nothing but just submit to the decisions produced by the algorithms.⁵⁸

Automation Bias

Delegating vital tasks to automated or autonomous systems is a process that had previously been observed in a number of civilian sectors.⁵⁹ It is only natural that the range of these delegated activities will expand in line with improvements in technology.⁶⁰ Time-tested technologies and systems will be used in other sectors as well, needless to say, with a view to managing the problems and processes more successfully and swiftly.

There is also another factor observable at the relevant bureaucratic circles including but not limited to military establishments, which exerts an undeniable influence on technology-related administrative decisions. This factor is the unrestrained enthusiasm about what the technology may potentially offer. Some dub this marvel-laden state of technological enchantment ‘technophilic hubris’.⁶¹

With the trend of delegation so strong and the optimism as to the promise of autonomy so overwhelming, it is a common assumption that human involvement will diminish in many activities where autonomous systems could be utilized.⁶² The so-called technophilic hubris causes the anthropomorphic conceptualization of autonomous systems and technologies and contributes vastly to the accumulation of great expectations that autonomous systems will one day surpass human

⁵⁸ Garcia (n 1).[1].

⁵⁹ David D Woods, ‘The Risks of Autonomy: Doyle’s Catch’ (2016) 10 Journal of Cognitive Engineering and Decision Making.[131].

⁶⁰ *ibid.*

⁶¹ Hoffman and others (n 19).[255].

⁶² Woods (n 59).[131].

capabilities.⁶³ As long as those great expectation are efficaciously met, there is no problem to talk about.⁶⁴ However, things may get very complicated extremely quickly, if the over-confidence in the capabilities of those systems proves to be unwarranted, which would be an especially distressing scenario due to devastating consequence in a hostility setting.

Literature Review on the Definitions of Automation Bias

Two major traits of delegation and unbridled optimism depend upon a seemingly unshakeable confidence about the future effects of technological revolutions.⁶⁵ This confidence, however, could be the ground zero for an extremely hazardous human behavior, namely automation bias, which is nothing but a ‘heuristic re-placement for vigilant information seeking and processing’.⁶⁶ This bias carries an inherent risk to turn human operators into automatic operators.⁶⁷

Strauß defines automation bias as ‘the tendency to uncritically accept the computer-generated outcome’.⁶⁸ Since the 1980’s actually, there have been continuous warnings as to the development of an out-of-the-loop kind of the loss of human familiarity within autonomous operations.⁶⁹ Vagle defines automation bias shortly as ‘over-reliance on and over-confidence in the results of’ automated information systems.⁷⁰ Vagle⁷¹ unmistakably highlights the fact that automation bias

⁶³ Hoffman and others (n 19).[255]; James Johnson, ‘Finding AI Faces in the Moon and Armies in the Clouds: Anthropomorphising Artificial Intelligence in Military Human-Machine Interactions’ *Global Society* (2023) [2023] *Global Society*.

⁶⁴ Coco (n 2).[4].

⁶⁵ Woods (n 59).

⁶⁶ Kathleen L Mosier and others, ‘Aircrews and Automation Bias: The Advantages of Teamwork?’ (2001) 11 *The International Journal of Aviation Psychology*. [1].

⁶⁷ Afonso Seixas-Nunes, *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective* (Cambridge University Press 2022).[55] <<https://www.cambridge.org/core/books/legality-and-accountability-of-autonomous-weapon-systems/FE880FD3F459B29A495D79D0C8347D79>>.

⁶⁸ Stefan Strauß, ‘Deep Automation Bias: How to Tackle a Wicked Problem of AI?’ (2021) 5 *Big Data and Cognitive Computing*. [7].

⁶⁹ Hoffman and others (n 19).[256].

⁷⁰ Jeffrey L Vagle, ‘Tightening the OODA Loop: Police Militarization, Race, and Algorithmic Surveillance’ (2016) 22 *Michigan Journal of Race and Law*. [101-128].

⁷¹ *ibid.* [128].

has created negative effects in many sectors, including, healthcare, transportation, power distribution, and defense.

One of the leading researchers to point at the undesirable outcomes of technology is Bainbridge, who calls these side-effects *the ironies of automation*. She warns that automation imposes new tasks on human operators and paves the way for new types of human error, thus making the whole process even more burdensome than it originally was.⁷²

Kirlik, in his 1993 paper, warns about how automation creates new task demands.⁷³ This stacking up of new tasks and challenges may later cause the human operator not to use the automation in question at all, which was initially put in place to facilitate the off-loading of the operators' tasks.⁷⁴ Kirlik opines that the operator *in situ* should always call the time and place to enter into autonomous phase. Dependent upon the levels and use of autonomy then, he foresees a controlling human operator. In his approach, the operator dynamically decides when to use autonomy as a task off-loading mechanism. It is then the personally-developed and adapted strategy of the human operator which would have a conclusive impact on the whole process.⁷⁵ If the operators are talented and sufficiently trained, it may be a realistic expectation that such personal strategies be in place. However, this too contributes immensely to increases in time spent and resources allocated for the training and recruitment of the highly intelligent and/or qualified personnel.

Parasuraman and Manzey⁷⁶ highlight the misleading characteristics of automated decision aids. Cues automatically generated by such systems are just too salient and flashy to allow the human operator to thoroughly examine all the relevant data accessible at that time. Secondly, the human operators expect these

⁷² Bainbridge, 'Ironies of Automation' (1983) 19 *Automatica* 775, 777; Hoffman and others (n 19).[255].

⁷³ 'Modeling Strategic Behavior in Human-Automation Interaction: Why an "Aid" Can (and Should) Go Unused' (1993) 35 *Human Factors*. [221-222].

⁷⁴ *ibid.*[222].

⁷⁵ Kirlik (n 73).

⁷⁶ 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52 *Human Factors*. [381-391].

aids to be sources of unmistakable advice, for they ascribe them an illusional power and anthromorphic authority.⁷⁷

As regards automation bias, Mary L. Cummings⁷⁸ offers a very clear-cut taxonomy of different types of automation bias. According to her, automation bias is simply there when human personnel disregards or quits searching for any information which may essentially be contradictory to the solution offered by the autonomous system.⁷⁹ She differentiates between three different types of automation bias, which are: automation bias in computer-assisted route planning, automation bias in critical event diagnosis and action, and finally automation bias in time-sensitive resource allocation.⁸⁰

Automation Bias in the Military Sphere

This author believes in the utility of Cummings' taxation of automation bias cases in analyzing the emergence of this bias in military setting during the deployment autonomous weapon systems in hostilities.

Automation bias in time-sensitive resource allocation is easily applicable for LAWS. In a given combat situation, the human operator can be in a situation where she will be called upon to render a decision with conceivable lethal ramifications. This only too naturally creates a paramount load of stress and anxiety on the shoulders of the personnel concerned, who is supposed to filter a considerably high number of variables from the field and come up with an adequate solution in a very short time span, trying at the same time to avoid miscalculations that may bring about death and destruction. This third bias, i.e., that one in the time-sensitive allocation of resources especially deals with such time-pressure situations as one would be subject to during the course of armed confrontations.⁸¹ Any operator that

⁷⁷ *ibid.*[391].

⁷⁸ 'Automation Bias in Intelligent Time Critical Decision Support Systems', *AIAA 1st Intelligent Systems Technical Conference* (American Institute of Aeronautics and Astronautics 2004) <<https://doi.org/10.2514/6.2004-6313>> accessed 22 September 2023.

⁷⁹ *ibid.*[2].

⁸⁰ *ibid.*[3–4].

⁸¹ *ibid.*[4].

faces the distant possibility of going through such baptism by fire would cherish the idea of making good use of autonomous decision aids. In LAWS' case, this would generally be either in-the-loop or on-the-loop modality for the time being. In this respect, Asaro⁸² opines that 'it will be difficult or impossible to design systems capable of dealing with the fog and friction of war', at least on its own a long time.

As Amoroso and Tamburrini⁸³ remind us though, 'the human tendency to over-trust machine decision-making—is demonstrably exacerbated when the human role consists solely of the ability to override decisions that have already been autonomously made by machines'. In the light of this, it is obvious that on-the-loop modality will cause even greater risk than in-the-loop functioning mode. Yet, the difference might be strikingly smaller than initially considered, solely due to pressure-laden and risk-rich environment of a military confrontation. With the cyber warfare emerging as a new method of warfare and quasi-ubiquitous deployment of new and deadlier technologies of warfare, the reaction time to the new types of attack is getting considerably shorter, likely to get out of control in the wake of mankind's natural capabilities and limitations. For this author, the risk caused by the in-the-loop interaction is arguably comparable to, if not the same with that of on-the-loop type of interaction.

In addition, the second bias type as to critical event diagnosis and action is also very relevant to military deployments of LAWS, since it is the kind of automation that helps human operators by delivering recommendations for actions. In cases where automated systems have failed to notice and/or to tip off on critical events or have proposed unfitting solutions as to the emerging situations, it has more often than not been the case that humans have opted to stick with these faulty suggestions.⁸⁴ This inclination has the potential to ultimately lead all that importance attached to the notion of meaningful human control into downright triviality. As a

⁸² (n 6).[692].

⁸³ Daniele Amoroso and Guglielmo Tamburrini, 'Toward a Normative Model of Meaningful Human Control over Weapons Systems' (2021) 35 *Ethics & International Affairs*. [245-260].

⁸⁴ Cummings (n 78).[4].

matter of fact, whether any kind of control on autonomous weapons by humans is a very legitimate question.

Linda Skitka has been another pioneering figure in the study of automation bias. After reminding her readers about the evident push, among others, in aviation industry, Skitka warns that ‘the presence of automated decision aids was probably associated with two kinds of errors: commission errors and omission errors’.⁸⁵ In omission errors, human operator fails to react to the changes in the environment due to belated or sloppy detection or complete lack thereof by the autonomous systems.⁸⁶ In commission errors, systems chime in and feed the human operator with false information or guidance.⁸⁷ These errors are, in addition to other factors, also attributable to three distinct features of humans in their interactions with automated machines.

The first underlying reason is the fact that a majority of people would be only too willing to follow the automatically generated solution sets, so that they will remain in their own cognitive comfort zone. Skitka⁸⁸ dubs them ‘the cognitive misers’. The second largely observable characteristics of humans is the so-called ‘slacking-off effect’. Working in groups, individuals usually refrain from allocating their complete energy and concentration to the task at hand, due the expectation that working in numbers would make a smaller amount of effort necessary for the fulfilling of the imminent task. In a similar vein, humans tend to perceive computers and other autonomous systems as team members, which cause them to go lightly about their individual areas of responsibility, which in turn causes errors.

Lastly, people suffer from their own propensity to perceive autonomous systems as ‘decision-making authorities’.⁸⁹ Reminiscent, among others, of those individuals taking part at the Stanford experiment who did not shy away from torturing their fellow citizens, people will evidently follow the decision aid provided

⁸⁵ Linda J Skitka, Kathleen L Mosier and Mark Burdick, ‘Does Automation Bias Decision-Making?’ (1999) 51 *International Journal of Human-Computer Studies*. [991 - 993].

⁸⁶ Vagle (n 70). [128].

⁸⁷ *ibid.* [129].

⁸⁸ (n 85). [992].

⁸⁹ *ibid.* [992].

by computer, which they would regard as smarter and *authoritative*.⁹⁰ It has been a common response to claim that more training of military personnel could be the key in battling automation bias and its consequences. It is now established that training and instructions do not have a big impact upon this kind of flawed decision-making process.⁹¹ Worryingly enough, functioning in teams does not automatically come to mean that one will be well shielded against the ill-effects of automation bias.

These last typologies mentioned employ a type of personnel with the best intentions, who is trying their best to fulfil their obligations within the legal and administrative framework. Yet they fail to discharge these obligations properly because of an inherent human tendency. This may be seen as an example of recklessness since there is no actual *mens rea* available in such cases to commit a crime. However, there may be such instances where human personnel may just opt to close their eyes to false suggestions the LAWS are generating.⁹² In a given specific case, this would not amount to a bias, for the personnel does not really act or refrain from acting due to an overconfidence in the capabilities of the system concerned. Here, they simply chose to ignore any autonomously generated suggestions so that a war crime could for example be committed. If, however, this personal choice to ignore these failures turns into an institutionalized ignorance at least for a considerable period of time, the newer batches of personnel will be kept in the dark as to these shortcomings and will in all probability be indoctrinated into believing that that faulty course of action (or rather omission) on behalf of the human operators must be the endorsed course of action in that specific unit. This now rather institutionalized willingness to take advantage of the impunity created by the autonomous qualities of these systems may work as an in-built factor to produce and reproduce a careless attitude vis-à-vis LAWS.

Another risk that may emerge due to criminal omissions of some human personnel is that this flawed and unlawful *modus operandi* may be taken in by the

⁹⁰ *ibid.*[993]; Vagle (n 70).[129].

⁹¹ Parasuraman and Manzey (n 76).[381].

⁹² Coco (n 2).[4].

system via *machine learning*. Having been allowed to engage civilians and non-military objects for a long time, the system may be taught to take such targets as perfect military targets to engage and decimate. This in turn will pave the way for another mistaken behavior of other personnel that will only too gladly follow the solutions generated by the autonomous systems, since the inner-functioning of the whole system is just too opaque for ordinary personnel to follow and understand. Though not directly an example of bias, this scenario is a very palpable menace of originally criminal nature. This criminal mode of functioning will be augmented by automation bias. In addition, especially in such systems where artificial neural networks are allowed to roam freely to facilitate a machine learning, there is a very probable risk that the system itself may find it necessary to change its own original programming to get better results in its functioning.⁹³ Such a development practically means that any contemplated prohibition inserted in absolute terms into the algorithm of the system could be jettisoned anytime during the deployment of the system. Any trust that was created because of the success of the system will get someone to continue heeding the solutions generated by the systems, yet the system and its operational framework will have been altered without any permission and/or approval by the competent authorities.

Regardless of the specific type of error and of its underlying reason or motives, automation bias is here on board with us, just as the LAWS already are. It is just natural that the former will follow the latter. Unfortunately, there has already been instances of automation bias with fatal consequences.⁹⁴ As noted above, there were two distinct cases of fratricide brought about by the autonomous air defense systems deployed during hostilities in Iraqi War. These systems on two different occasions took incoming fighter jets for missiles fired by the enemy and engaged the aerial vehicles causing loss of life. As this case vividly shows, there is potentially a huge insecurity caused by LAWS and a human supervisor might be just a convincing accessory to assuage the related worries but end up insufficient to tackle real risks on the field.

⁹³ Krishnan (n 13).[220].

⁹⁴ Parasuraman and Manzey (n 76).[395].

Conclusion

In light of the foregoing, it is safe to assume that LAWS are going to be among the most coveted means of warfare for the foreseeable future. Big powers with industrial capabilities and militaries seem just too willing to develop these systems. In their quest for game-changing LAWS, these mighty states wish to be intruded by no or little restraints, whereas other states against the LAWS are few in numbers and negligible in power and international influence. In light of these facts, the pro-ban campaign has a long road to walk, before they could ever claim victory in their pursuit to champion humanitarian values over the obvious military advantages offered by the deployment and development of LAWS.

Meaningful human control has emerged as a pivotal concept in diplomatic discussions. There is limitless trust invested in this concept, which however needs further concretization. Lack of common understanding is an important deficiency in terms of law-making and proper implementation. Many international actors depend on the further acknowledgement and adoption in legal documents of this control so that humanitarian values can be promoted and safeguarded. Nevertheless, automation bias has all the potential to off-set any imaginable precautions likely to be provided by a human supervisor and their presumed control. All the researchers and states' representatives who try to delineate between different modalities of human-machine interaction and ultimately favoring the human-in-the-loop modality so that a certain degree of human control may be kept at all times over the LAWS, could have been gravely misled, should they ignore automation bias and its effects. The issue is a familiar problem for a relatively long time especially in medical and aviation industries. This bias to comply with the solutions generated by these systems have caused accidents in the past in mentioned sectors, where problems particularly occurred after the systems failed to generate accurate solutions.

What the literature has so far highlighted in this respect should create serious doubts as to the feasibility of insisting as a *panacea* on a continuous human control upon LAWS. In the wake of all the ethical opposition to LAWS coupled with the scientific findings about automation bias, one must feel obliged to reiterate that

meaningful human control is not something easily attainable and any future legal construction must avoid investing heavily in it.

It is justifiable and understandable to try to see in human control an important pillar or shield, for the upholding of humanitarian rules. Yet, this control might be potentially doomed to failure in many instances, if not all, of the deployment of LAWS. One must rather put the blame on our tendency to overzealously heed to the suggestions and/or solutions offered by autonomous systems, which are undeniably faster and more intelligent in computation and are perceived to be simply flawless in execution. The risk created by autonomous systems seems to be more substantial and perilous after the introduction of LAWS into the conflict theatres. The author will claim that the only way to prevent this inherent risk might be a ban on the development and deployment of LAWS altogether. However, this is not a realistic expectation, since many states are keen on developing these systems.

This paper aims to highlight how misplaced our trust in MHC could be. Until we may safely be talking about machines having cognitive capabilities comparable to that of humans and thus about the possibility of full implementation of humanitarian rules, we should refrain from blindly overvaluing any human control as a reliable benefactor of humanitarian concerns. It will in all probability fail due to our innate tendency to over-trust automated processes. As for now, this author feels obliged to agree with George Jain⁹⁵ about that ‘it remains unclear whether and how humans can exercise meaningful control over technologies that are intended to exceed human capabilities in terms of speed or cognitive capacity’. This is a vital question we must all ponder over in the future.

Acknowledgments

The improvement and the revision of a previous draft of this paper was facilitated by a scholarship received within the framework of TUBITAK 2219 Programme, Batch 2023/.

⁹⁵ Abhimanyu George Jain ‘Autonomous Weapon Systems, Errors and Breaches of International Humanitarian Law’ (2023) 21 (3) *Journal of International Criminal Justice*.

Bibliography

- Alston P, 'Interim Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions' (United Nations Human Rights Council 2010) UN Doc A/65/32.
- Amoroso D and Tamburrini G, 'The Ethical and Legal Case Against Autonomy in Weapons Systems' (2018) 18 <<https://doi.org/10.1515/gj-2017-0012>> accessed 21 September 2023.
- , 'Toward a Normative Model of Meaningful Human Control over Weapons Systems' (2021) 35 *Ethics & International Affairs*.
- Anderson K, Reisner D and Waxman M, 'Adapting the Law of Armed Conflict to Autonomous Weapon Systems' (2014) 90 *International Law Studies*.
- Arkin RC, *Governing Lethal Behavior in Autonomous Robots* (CRC Press 2009)
- , 'The Case for Ethical Autonomy in Unmanned Systems' (2010) 9 *Journal of Military Ethics*.
- Asaro P, 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making' (2012) 94 *International Review of the Red Cross*.
- AU - Misselhorn C, 'Autonome Waffensysteme/Kriegsroboter', *Handbuch Maschinenethik* (Springer VS, Springer Fachmedien Wiesbaden 2019) <https://www.researchgate.net/publication/336597731_Autonome_WaffensystemeKriegsroboter>.
- Badell D and Schmitt L, 'Contested Views? Tracing European Positions on Lethal Autonomous Weapon Systems' (2022) 31 *European Security*.
- Bainbridge, 'Ironies of Automation' (1983) 19 *Automatica*.
- Beck S, 'Der Rechtliche Status Autonomer Maschinen' (2017) 26 *PJA/AJP*.
- Boni M, 'The Ethical Dimension of Human–Artificial Intelligence Collaboration' (2021) 20 *European View*.
- Chengeta T, 'Defining the Emerging Notion of “Meaningful Human Control” in Weapon Systems' (2017) 49 *New York University Journal of International Law and Politics*.
- Chetail V, 'The Fundamental Principles of Humanitarian Law through the Case Law of the International Court of Justice' (2002) 21 *Refugee Survey Quarterly*

<<https://repository.graduateinstitute.ch/record/5034/files/Refugee%20Survey%20Quarterly-2002-Chetail-199-211.pdf>>.

Christie EH and others, 'Regulating Lethal Autonomous Weapon Systems: Exploring the Challenges of Explainability and Traceability' [2023] *AI and Ethics* <<https://doi.org/10.1007/s43681-023-00261-0>>.

Coco A, 'Exploring the Impact of Automation Bias and Complacency on Individual Criminal Responsibility for War Crimes' [2023] *Journal of International Criminal Justice* mjad.

Crootof R, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 *Temple International and Comparative Law Journal*.

Cummings M, 'Automation Bias in Intelligent Time Critical Decision Support Systems', *AIAA 1st Intelligent Systems Technical Conference* (American Institute of Aeronautics and Astronautics 2004) <<https://doi.org/10.2514/6.2004-6313>> accessed 22 September 2023.

Dahlmann A, Hoffberger-Pippan E and Wachs L, 'Autonome Waffensysteme und menschliche Kontrolle - Konsens über das Konzept, Unklarheit über die Operationalisierung'.

'Department of Defense Directive 3000.09' (Department of Defense 2012).

Ekelhof M, 'Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation' (2019) 10 *Global Policy*.

Ekelhof MAC, 'Lifting the Fog of Targeting' (2018) 71 *Naval War College Review*.

Garcia D, 'Algorithms and Decision-Making in Military Artificial Intelligence' [2023] *Global Society*.

George Jain A, 'Autonomous Weapon Systems, Errors and Breaches of International Humanitarian Law' [2023] *Journal of International Criminal Justice* mjad.

Gómez de Ágreda Á, 'Ethics of Autonomous Weapons Systems and Its Applicability to Any AI Systems' (2020) 44 *Artificial intelligence, economy and society*.

Haner J and Garcia D, 'The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development' (2019) 10 *Global Policy*.

Heyns C, 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions' (United Nations Human Rights Council 2013) UN Doc A/HRC/23/47.

Hoffman RR and others, 'Myths of Automation and Their Implications for Military Procurement' (2018) 74 *Bulletin of the Atomic Scientists*.

ICRC, 'ICRC Position on Autonomous Weapon Systems: ICRC Position and Background Paper' (ICRC 2021).

Johnson J, 'Finding AI Faces in the Moon and Armies in the Clouds: Anthropomorphising Artificial Intelligence in Military Human-Machine Interactions' *Global Society* (2023) [2023] *Global Society*.

Kirlik A, 'Modeling Strategic Behavior in Human-Automation Interaction: Why an "Aid" Can (and Should) Go Unused' (1993) 35 *Human Factors*.

Klare MT, 'The Challenges of Emerging Technologies' [2018] *Arms Control Today*.

Krishnan A, 'Enforced Transparency: A Solution to Autonomous Weapons as Potentially Uncontrollable Weapons Similar to Bioweapons' in Jai Galliot, Duncan MacIntosh and Jens David Ohlin (eds), *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare* (Oxford University Press 2021) <<https://doi.org/10.1093/oso/9780197546048.003.0015>> accessed 18 July 2023.

Ma E, 'Autonomous Weapons Systems Under International Law' (2020) 95 *New York University Law Review*.

McFarland T, *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law* (Cambridge University Press 2020) <<https://www.cambridge.org/core/books/autonomous-weapon-systems-and-the-law-of-armed-conflict/09BFF6BB5B88E34935678B5A0606A8A7>>.

Methnani L and others, 'Let Me Take Over: Variable Autonomy for Meaningful Human Control' (2021) 4 *Frontiers in Artificial Intelligence* <<https://www.frontiersin.org/articles/10.3389/frai.2021.737072>>.

Meurant J, 'Inter Arma Caritas: Evolution and Nature of International Humanitarian Law' (1987) 24 *Journal of Peace Research*.

Mitchell C, 'When Laws Govern LAWS: A Review of the 2018 Discussions of the Group of Governmental Experts on the Implementation and Regulation of Lethal Autonomous Weapons Systems' (2020) 36 *Santa Clara High Technology Law Journal*.

Mosier KL and others, 'Aircrews and Automation Bias: The Advantages of Teamwork?' (2001) 11 *The International Journal of Aviation Psychology*.

Parasuraman R and Manzey DH, 'Complacency and Bias in Human Use of Automation: An Attentional Integration' (2010) 52 *Human Factors*.

Reeves SR, Alcalá RTP and McCarthy A, 'Challenges in Regulating Lethal Autonomous Weapons under International Law Fighting in the Law's Gaps' (2021) 27 *Southwestern Journal of International Law*.

Roff H and Moyes RM, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons' (Article 36 2016) Briefing Paper <<https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf>>.

Rosendorf O, Smetana M and Vranka M, 'Autonomous Weapons and Ethical Judgments: Experimental Evidence on Attitudes toward the Military Use of "Killer Robots".' (2022) 28 *Peace and Conflict: Journal of Peace Psychology*.

Rosert E and Sauer F, 'Prohibiting Autonomous Weapons: Put Human Dignity First' (2019) 10 *Global Policy*.

Santoni de Sio F and van den Hoven J, 'Meaningful Human Control over Autonomous Systems: A Philosophical Account' (2018) 5 *Frontiers in Robotics and AI* <<https://www.frontiersin.org/articles/10.3389/frobt.2018.00015>>.

Sati MC, 'The Attributability of Combatant Status to Military AI Technologies under International Humanitarian Law' [2023] *Global Society*.

Schaub H, 'Der Einsatz Autonomer Waffensysteme Aus Psychologischer Perspektive' (2020) Volume 13 *Zeitschrift für Außen- und Sicherheitspolitik*.

Sehoon Park, 'Analysis of the Positions Held by Countries on Legal Issues of Lethal Autonomous Weapons Systems and Proper Domestic Policy Direction of South Korea' (2020) 32 *The Korean journal of defense analysis*.

Seixas-Nunes A, *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective* (Cambridge University Press 2022) <<https://www.cambridge.org/core/books/legality-and-accountability-of-autonomous-weapon-systems/FE880FD3F459B29A495D79D0C8347D79>>.

Skitka LJ, Mosier KL and Burdick M, 'Does Automation Bias Decision-Making?' (1999) 51 *International Journal of Human-Computer Studies*.

Solovyeva A and Hynek N, 'Going Beyond the "Killer Robots" Debate: Six Dilemmas Autonomous Weapon Systems Raise.' (2018) 12 *Central European Journal of International & Security Studies*.

Sparrow R, 'Killer Robots' (2007) 24 *Journal of Applied Philosophy*.

- Strauß S, 'Deep Automation Bias: How to Tackle a Wicked Problem of AI?' (2021) 5 *Big Data and Cognitive Computing*.
- Szpak A, 'Legality of Use and Challenges of New Technologies in Warfare – the Use of Autonomous Weapons in Contemporary or Future Wars' (2020) 28 *European Review*.
- Vagle JL, 'Tightening the OODA Loop: Police Militarization, Race, and Algorithmic Surveillance' (2016) 22 *Michigan Journal of Race and Law*.
- Verdiesen I, Santoni de Sio F and Dignum V, 'Accountability and Control Over Autonomous Weapon Systems: A Framework for Comprehensive Human Oversight' (2021) 31 *Minds and Machines*.
- Weigend T, 'Convicting Autonomous Weapons?: Criminal Responsibility of and for AWS under International Law' [2023] *Journal of International Criminal Justice* mqad.
- Wood NG, 'Autonomous Weapon Systems and Responsibility Gaps: A Taxonomy' (2023) 25 *Ethics and Information Technology*.
- Woods DD, 'The Risks of Autonomy: Doyle's Catch' (2016) 10 *Journal of Cognitive Engineering and Decision Making*.

--This page is intentionally left blank--