Essay Competition ELSA Milano 2020/2021

The Damaging Potential of Biased Automated Weapon Systems: A Call for the Regulation of the Development Phase in Military Uses of AI

List of Abbreviations

AGI	Artificial General Intelligence
AI	Artificial Intelligence
AR	Augmented Reality
CCW	Convention on Certain Conventional Weapons
DoD	US Department of Defence
ICRC	International Committee of the Red Cross
iPRAW	International Panel on the Regulation of Autonomous Weapons
LAWS	Lethal Autonomous Weapon Systems
Lt. Col.	Lieutenant Colonel
NGO	Non-Governmental Organisation
US	United States of America

Essay Competition ELSA Milano 2021/2021

The Damaging Potential of Biased Automated Weapon Systems: A Call for the Regulation of the Development Phase in Military Uses of AI

1 Introduction

It has been 66 years since the term 'artificial intelligence' (AI) was first coined by Professor John McCarthy[1], a Dartmouth researcher who, along with a group of fellow academics and scientists, proposed a summer research project on the subject[2], and to say we have come far ever since is an understatement. From the AI-powered algorithms that sort spam e-mails out of our inbox folders[3] to the sophisticated neural networks that can assess the metastasis risk of skin cancer as accurately as a dermatologist would[4], AI has become ubiquitous in our day to day lives.

Just like several other scientific breakthroughs that preceded it in history, AI is on the verge of changing the way armed conflict is conducted by automating certain military tasks, a development that concerns many in the international sphere.

When we think of automation in warfare, it may be easy for our mind to stray to extreme scenarios where rifle-bearing robots strike terror in the battlefield, but the reality is that armies are working on AI and machine learning solutions that could revolutionise how they operate in ways that are less sensational, yet worthy of great caution.

In a report that originated from the discussions held at a roundtable meeting with AI researchers on June 2018, the International Committee of the Red Cross (ICRC) defined autonomous weapon system as "any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e., search for or detect, identify, track, select) and attack (i.e., use force against, neutralize, damage or destroy) targets without human intervention."[5]. The

[1] Stanford University, Professor John McCarthy: Father of AI, http://jmc.stanford.edu/general/index.html, (accessed on 27.02.2021).
[2] J. McCarthy et al., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, in: AI Magazine 27 (4) (2006), https://doi.org/10.1609/aimag.v27i4.1904, p. 12.
[3] Information Age, Can Artificial Intelligence Spot Spam Quicker Than Humans?, https://www.information-age.com/artificial-intelligence-spam-machine-learning-123481368/, (accessed on 27/02/2021).
[4] ScienceDaily, Algorithm that performs as accurately as dermatologists, https://www.sciencedaily.com/releases/2021/02/210212111902.htm, (accessed on 27.02.2021).
[5] International Committee of the Red Cross, Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control, https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control, (accessed on 24.02.2021), p. 5.

lack of human intervention is determined by the fact that, once activated, autonomous weapon systems use their sensors, software, and connected weaponry to identify and attack targets autonomously[6]: therefore, the human operator does not have awareness of the target attacked, nor the timing and location of the attack, as opposed to what happens with remotely controlled weapon systems, where the target, location and time of attack are chosen by a human user[7]. The lack of human involvement in these decisions is a main point of concern for the ICRC[8], and it informs its human-centred agenda regarding the use and development of AI in warfare[9].

It is very relevant to this essay's discussion that remotely controlled weapons, as the ICRC's report points out, could easily "become tomorrow's autonomous weapons with just a software upgrade"[10]: subsequently, "(…) the question is not so much whether we will see more weaponised robots, but whether and by what means they will remain under human control."[11].

2 Opposing Views on the Regulation of Autonomous Weapon Systems, from the Requirement of Human Control and Towards a Ban on LAWS

The idea of human control is based around the notion that AI-powered weapon functions must serve as a tool that augments human decision-making rather than replace it[12], a requirement that is also crucial to ensure that combatants comply with international and humanitarian law and maintain legal responsibility for the decisions they enact[13].

In the light of this same human-centric spirit the European Parliament, in a resolution adopted on January 2021, has called for the adoption of a common European position on lethal autonomous weapon systems (LAWS) aimed at preventing altogether the "development, production and the use of LAWS capable of attack without meaningful human control, as well as the initiation of effective negotiations for their prohibition (…)"[14], suggesting even a ban on "so-called 'killer robots'"[15].

---

[6] *Ibid.*
[7] *Ibid.*
[8] International Committee of the Red Cross, Artificial Intelligence and Machine Learning in Armed Conflict: a Human-Centered Approach, https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach, (accessed on 24.02.2021), p. 5.
[9] Id., p. 7.
[10] International Committee of the Red Cross, supra note 5, p. 6.
[11] *Ibid.*
[12] *Ibid.*
[13] *Ibid.*
[14] European Parliament, European Parliament Resolution on Artificial intelligence: Questions of Interpretation and Application of International Law, 20 January 2021, P9_TA(2021)0009.
[15] *Ibid.*

The plea to adopt a pre-emptive ban on LAWS was first advanced in 2012 by Campaign to Stop Killer Robots, a coalition of non-governmental organisations (NGOs) and international actors[16] that advocates for the adoption of an international treaty that would not only enshrine the requirement of meaningful human control in weapons with automated functions in international law, but effectively prohibit the use and development of fully automated weapons[17].

According to a 2020 report by Human Rights Watch, a majority of the 97 countries that have publicly expressed their stance on fully autonomous weapons agree, on varying levels, that their use must be subjected to human control and judgement[18], with a considerable number of countries and several state groups even calling for a ban of fully autonomous weapons[19]. Furthermore, in a recent meeting of the Convention on Certain Conventional Weapons (CCW) on lethal autonomous weapons, a call for the adoption of a "legally-binding instrument to prohibit and restrict such weapon systems"[20] was endorsed by more than 65 High Contracting Parties of the Convention[21].

However, the efforts of these countries are bound to be halted by the adversarial position that has been constantly adopted over time from the likes of Russia and the United States[22], who deem moves to create a treaty banning LAWS "premature"[23].

The United Kingdom's position on the matter, who has also consistently been that of opposing the adoption of a ban on LAWS[24], offers food for thought: while there is no intention, on part of the United Kingdom, to develop fully autonomous weapons, the opinion on weaponry with automated functions is that "international humanitarian law and the existing regulatory

[16] Campaign to Stop Killer Robots, About, https://www.stopkillerrobots.org/about/, (accessed on 04.04.2021).
[17] Campaign to Stop Killer Robots, Learn: the Solution, https://www.stopkillerrobots.org/learn/#problem, (accessed on 04.03.2021).
[18] Human Rights Watch, Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control, https://www.hrw.org/news/2020/08/10/killer-robots-growing-support-ban, (accessed on 04.03.2021), p. 3.
[19] Id., p. 4.
[20] Campaign to Stop Killer Robots, Diplomatic Talks Re-Convene, https://www.stopkillerrobots.org/2020/09/diplomatic2020/, (accessed on 04.03.2021).
[21] *Ibid*.
[22] Human Rights Watch, supra note 14, p. 5.
[23] *Ibid*. This position was reportedly reiterated by the United States at the September 2020 meetings, Campaign to Stop Killer Robots, supra note 20.
[24] United Nations, United Kingdom's Opening Remarks at the First Session of the GGE on LAWS (21 to 25 September 2020), https://documents.unoda.org/wp-content/uploads/2020/09/LAWS-Sept-GGE-2020-UK-Opening-Statement.pdf, (accessed on 13.03.2021).

framework for the development, procurement and use of weapons systems remain more than sufficient to regulate new capabilities"[25].

The reasons that inform the United Kingdom's LAWS policy were eloquently delineated by Lt. Col. John Stroud-Turp of the Ministry of Defence during a 2016 ICRC Expert Meeting on the subject[26]. When called upon to discuss the possible consequences of an expansion of autonomous functions in weapon systems, Stroud-Turp argued that is very unlikely that a machine will ever be capable of emulating the multitude of intellectual aptitudes inherent to the human mind that are fundamental to making key military decisions[27], mainly the ability to combine analytical thinking with intuition[28].

While analytical thinking consists in the use of external information to identify the ideal solution to a problem from an array of alternatives, an intuitive approach to decision making is based on a more subjective perception of reality, informed by prior knowledge, experience, character and judgement[29]. Given that the level of understanding and autonomous learning required to make sensible decisions in a battlefield is, Stroud-Turp argues, "beyond the learning predicted for machines"[30], a wider, pre-emptive ban on LAWS would only negatively affect the research currently conducted on target precision tools and other non-lethal AI tools that could contribute to the reduction of collateral damage caused by existing weapons, as well as improve their efficiency[31]. Simply put, a pre-emptive ban would not make sense if weapons with autonomous functions that emulate perfectly human intelligence do not (and may never) exist.

AI scientists define the artificially intelligent systems that would be sophisticated enough to be indistinguishable from the human mind as general artificial intelligence (AGI), currently a mere theoretical concept[32]. Opinions on when and if these systems could be developed diverge[33]: when writer Martin Ford asked a group of AI experts to give their best estimate on when, in

---

[25] *Ibid.*
[26] Lt. Col. J. Stroud-Turp, Lethal Autonomous Weapon Systems (LAWS): Speaker's Summary, in: Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons, Expert Meeting, Versoix, Switzerland, 15-16 March 2016, https://www.icrc.org/en/publication/4283-autonomous-weapons-systems, (accessed on 04.03.2021), p. 57.
[27] Id., p. 58.
[28] Id., p. 57.
[29] *Ibid.*
[30] Id., p. 58.
[31] Id., p. 59.
[32] IBM Cloud Learn Hub, Strong AI, https://www.ibm.com/cloud/learn/strong-ai, (accessed on 08.03.2021).
[33] J. Vincent, This is When AI's Top Researchers Think Artificial Intelligence Will be Achieved, https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book, (accessed on 08.03.2021).

the future, there will be at least a 50% chance of AGI being built, the average response was year 2099, although Ray Kurzweil of Google believes AGI will be achievable as soon as 2029[34]. Opinions also diverged on what it will take for AGI to be developed, with some researchers claiming that we already have most of the technical tools needed[35] and others arguing that we are "still missing a great number of the fundamental breakthroughs needed to reach this goal."[36]

With regards to the potential risks posed by AGI, a majority of the scientists interviewed by Mr. Ford believe the problem to be "(…) extremely distant compared to problems like economic disruption and the use of advanced automation in war."[37] As professor Barbara Grosz of Harvard University holds, "The real point is we have any number of ethical issues right now, with the AI systems we have (…) I think it's unfortunate to distract attention from those because of scary futuristic scenarios."[38].

In the light of the scientific discourse around the practical attainability of AGI and Professor Grosz's warning, one may have reason enough to wonder whether the United Kingdom's LAWS policy is built around a premature assumption that draws the focus away from the potential danger that weapons with automated functions being developed today already pose.

2.1 The potential for regulation in Article 36-inspired weapon review processes

As mentioned previously, the United Kingdom deems existing international humanitarian law, combined with the exercise of human control over their use, to be sufficient to regulate autonomous weapon systems[39]: in particular, as Lt. Col. Stroud-Turp points out, a thorough Article 36 review of every newly developed weapon, as required by Protocol I Additional of the Geneva Conventions, represents a legal threshold robust enough to regulate the development of novel autonomous weapon systems[40].

Article 36 obliges High Contracting Parties to the Geneva Conventions to determine, at the time of study, development, acquisition, or adoption of new weapons, means or methods of

[34] *Ibid.*
[35] *Ibid.*
[36] *Ibid.*
[37] *Ibid.*
[38] *Ibid.*
[39] United Nations, supra note 24.
[40] Lt. Col. J. Stroud-Turp, supra note 26, p. 59.

warfare, whether their deployment in conflict would, in some or all circumstances, be prohibited under Protocol I Additional or any other rule of international law[41].

In and of itself, Protocol I Additional does not provide High Contracting States with a detailed explanation as to how Article 36 weapon reviews must be conducted, therefore it is the single State's duty to identify and adopt the legal, administrative, and regulatory instruments needed to conduct the review as soon as novel weapons are being studied, developed or their inclusion in the national arsenal considered[42].

Only 25 States are reportedly known to conduct proper Article 36 reviews[43], and Stroud-Turp believes that if that number were to increase following an international move towards the practice "(…) the fear, suspicion and misunderstanding surrounding the development of future weapon systems could be partly allayed. Article 36 reviews have been capable of dealing with advances in technology for close to 40 years; there is no reason to doubt their suitability for dealing with greater advances in autonomy."[44].

Despite them not being a party to Protocol I Additional [45], the United States have also shown their support for the adoption of a weapons review process[46]: in his closing statement at the 2015 CCW Meeting of Experts on Lethal Autonomous Weapons Systems, the Head of the US Delegation Michael W. Meier declared that "The United States would like to see the Meeting of High Contracting Parties agree to begin work on an interim outcome document that sets forth what is entailed by a comprehensive weapons review process, including the policy, technical, legal and operational requirements that would apply if a state were developing LAWS."[47], specifying that the document would not constitute an incentive to the adoption of

---

[41] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 UNTS 3.

[42] International Committee of the Red Cross, A Guide to the Legal Review of New Weapons, Means and Methods of Warfare Measures to Implement Article 36 of Additional Protocol I of 1977, https://www.icrc.org/en/publication/0902-guide-legal-review-new-weapons-means-and-methods-warfare-measures-implement-article, (accessed on 07.03.2021), p. 20.

[43] Lt. Col. J. Stroud-Turp, supra note 26, p. 59.

[44] *Ibid.*

[45] M.W. Meier, Autonomous Legal Reasoning?: Legal and Ethical Issues in the Technologies of Conflict: Lethal Autonomous Weapons Systems (LAWS): Conducting a Comprehensive Weapons Review, 30 Temp. Int'l & Comp. L.J. 119, Spring 2016, p. 5.
Meier reminds us that "For those States that are not a party to Additional Protocol I, such as the United States, Boothby notes that there is 'an applied obligation to conduct such review as noted by the practice of certain states prior to the adoption of Additional Protocol I."

[46] M.W. Meier, CCW LAWS Meeting: U.S. Closing Statement and the Way Ahead, https://geneva.usmission.gov/2015/05/08/ccw-laws-meeting-u-s-closing-statement-and-the-way-ahead/, (accessed on 06.03.2021).

[47] *Ibid.*

LAWS, but rather a safeguard against potentially dangerous, unregulated developments in the field of LAWS[48].

In a paper he authored shortly thereafter, Meier holds that, with the international debate on LAWS being far from a conclusion, focusing the CCW's discussions on the adoption of a document that outlines a model weapon review procedure would "help identify any specific issues related to evaluating LAWS"[49] as well as prove useful for conducting reviews on all other weapons systems[50]. He then sets out to review the questions that should be posed to conduct a legal review of autonomous weapons, as identified by the US Department of Defence's Law of War Manual[51] and Bill Boothby's 'Weapons and the Law of Armed Conflict'[52]. The questions identified read as follows:

(1) Is there a specific rule, whether in the form of a treaty obligation or customary international law, that prohibits or restricts the use of the weapon?[53]
(2) Is the weapon of a nature to cause superfluous injury or unnecessary suffering in its normal or intended circumstances of use?[54]
(3) Is the weapon inherently indiscriminate?[55]
(4) Is the weapon intended to, or may it expectedly, cause widespread, long term and severe damage to the natural environment?[56]
(5) Are there any likely future developments in the law of armed conflict that may be expected to affect the weapon subject to review?[57]

For the sake of brevity, this paper will not examine in detail each of the five questions identified by Attorney-Adviser Meier. However, attention must be paid to question three, since it emerges that the international community has raised questions as to whether LAWS can be used in compliance with the principle of distinction between civilians and combatants[58].

The principle of distinction is enshrined in Article 48 of Protocol I Additional to the Geneva Conventions, that reads: "In order to ensure respect for and protection of the civilian population

[48] *Ibid.*
[49] M.W. Meier, supra note 45, p. 5.
[50] *Ibid.*
[51] Id., p. 7.
[52] Id., p. 8.
[53] *Ibid.*
[54] Id., p. 9.
[55] Id., p. 10.
[56] Id., p. 11.
[57] Id., p. 12.
[58] Id., p. 10.

and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[59].

To better understand the underlying issues that are culprit of making the use of AI in autonomous weapon systems incompatible with the principle of distinction, a brief explanation as to how these technologies work is essential.

2.2 AI and machine learning: a brief technical introduction

The artificially intelligent systems that are being developed today, including those who constitute the autonomous functions of automated weapons as defined by the ICRC[60] are defined as narrow AI. Machine learning solutions designer and developer DeepAI defines artificial intelligence as the "application of rapid data processing, machine learning, predictive analysis and automation to simulate intelligent behaviour and problem-solving capabilities with machines and software"[61].

Despite the name it bears, there is nothing simple or limited about narrow AI. Artificial intelligence has reached a significant level of sophistication, mainly thanks to the advances made in the field of machine learning, a branch of artificial intelligence that empowers the artificial brain to develop autonomously as opposed to simply responding to an input the way it was programmed to[62]. This is achieved by using large amounts of data to teach the algorithm how to identify different patterns and characteristics, so that it can then do the same when presented with a new set of data[63]. The goal is to use the machine's ability to analyse large amounts of data fast so as to make informed predictions and suggest effective solutions to a particular problem[64].

There are three types of machine learning: supervised, semi-supervised and unsupervised[65].

---

[59] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 UNTS 3.
[60] International Committee of the Red Cross, supra note 5, p. 2.
[61] DeepAI, Artificial Intelligence, https://deepai.org/machine-learning-glossary-and-terms/artificial-intelligence, (accessed on 08.03.2021).
[62] IBM Cloud Learn Hub, Machine Learning, https://www.ibm.com/cloud/learn/machine-learning#toc-what-is-ma-qhM6PX35, (accessed on 08.03.2021).
[63] *Ibid.*
[64] *Ibid.*
[65] *Ibid.*

While in supervised learning the algorithm is trained with data that has already been labelled by a human[66], unsupervised models autonomously work through larger amounts of unlabelled data to find new patterns and similarities between different data points, all without the need for human intervention[67]. Their ability to spot similarities and patterns in raw data makes unsupervised learning models ideal for the purpose of grouping data together (a task known as 'clustering') and identifying relationships between different variables[68]. Lastly, semi-supervised learning models, as the word may suggest, use datasets that are only partially labelled: they take advantage of the labelled portion of data to guide the algorithm's classification activity on unlabelled data[69].

The latest frontier in machine learning is deep learning: systems that are powered by deep learning are distinguished from 'non-deep' machine learning systems because of their ability to learn autonomously through large, unlabelled datasets by making use of neural networks[70] (more precisely, deep neural networks – that is, neural networks with more than three layers[71]).

To try and put it as simply as possible, artificially intelligent systems that are powered by neural networks work through the data via an extremely intricated process that emulates the human brain[72]. The multiple layers that constitute a deep neural network can be visible (this is the case for input and output layers) or hidden[73]: the visible input and output layers are respectively where the data enters the neural network and where the final outcome of the data elaboration process is determined[74]. The calculations needed to obtain a final output occur between the various hidden layers that make up the neural network, wherein the data is processed by progressively more complex algorithms in a procedure known as forward propagation[75]. Deep learning neural networks even learn from their 'mistakes' via a process known as backpropagation, in which erroneous outcomes are assigned weight and biases, then sent back

---

[66] IBM Cloud Learn Hub, What is Supervised Learning?, https://www.ibm.com/cloud/learn/supervised-learning#toc-how-superv-A-QjXQz-, (accessed on 08.03.2021).
[67] IBM Cloud Learn Hub, Unsupervised Learning, https://www.ibm.com/cloud/learn/unsupervised-learning#toc-what-is-un-MP1gM75c, (accessed on 08.03.2021).
[68] *Ibid.*
[69] IBM Cloud Learn Hub, supra note 62.
[70] IBM Cloud Learn Hub, AI vs. Machine Learning vs. Neural Networks: What's the Difference?, https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks, (accessed on 09.03.2021).
[71] *Ibid.*
[72] *Ibid.*
[73] IBM Cloud Learn Hub, Deep Learning, https://www.ibm.com/cloud/learn/deep-learning, (accessed on 09.03.2021).
[74] *Ibid.*
[75] *Ibid.*

to previous layers[76]: weights and biases play a crucial role in a neural network's learning process, because they represent the fundamental parameters that dictate how input data moves through the neural network's layers[77]. While weights dictate how much influence the input data will have on the final outcome, biases (that is, the bias neurons of a neural network, not to be confused with the generic meaning of the word bias[78]) enable the model to be flexible and adapt itself, if need be, to "fit the data"[79]. When this intricate web of back-and-forth interactions plays out in the context of a very large amount of data, the learning capacity of machine learning becomes exponentially more sophisticated and accurate[80]. Among the multiple practical applications of deep neural networks there is computer vision, a field of AI that enables computers to extract information from still images and videos[81].

3 The role of human operators in tainting training datasets: the potential for disaster in future war zones

As mentioned previously, the principle of distinction introduces an obligation for the parties to the conflict to only engage in offensive actions against military objectives and combatants, and clearly distinguish them from the civilian population and their objects[82].

Keeping in mind the tidbits of 'technical' knowledge we have acquired so far, how would technology potentially enable lethal autonomous weapons to make that distinction in a battlefield?

In the hypothetical case of a weapon system analogous to a missile drone that autonomously identifies targets and initiates attacks against them, said autonomous weapon would most likely utilise computer vision to single out targets in the battlefield. Being powered by machine learning, computer vision uses large amounts of data to learn how to discern different elements in an image and, ultimately, recognise objects or persons[83] (in the case of our missile drone, military objects and combatants). Its machine learning component allows it to use a large set of data to learn how to tell different images apart, while a convolutional neural network helps

---

[76] *Ibid.*
[77] AI Wiki, Weights and Biases, https://docs.paperspace.com/machine-learning/wiki/weights-and-biases, (accessed on 09.03.2021).
[78] Ł. Gebel, Why We Need Bias in Neural Networks, https://towardsdatascience.com/why-we-need-bias-in-neural-networks-db8f7e07cb98, (accessed on 09.03.2021).
[79] *Ibid.*
[80] IBM Cloud Learn Hub, supra note 73.
[81] IBM, Computer Vision, https://www.ibm.com/topics/computer-vision, (accessed on 09.03.2021).
[82] Protocol Additional to the Geneva Conventions, supra note 59.
[83] IBM, supra note 81.

it recognise images by breaking them down in smaller, labelled portions, based on which the convoluted neural network will make predictions about what the image consists of, reiterating the process until the predictions are accurate[84].

Machine learning models, including those who empower computer vision tools, rely heavily on data to learn and improve their performance over time, therefore training datasets need to be prepared properly to ensure the machine learning model gives out accurate responses[85].

However, bad training datasets create problems that go well beyond technical inaccuracy: the human beings tasked with collecting and labelling training data can taint the process with their own biases and preconceptions, resulting in models that effectively perpetrate racism and prejudice, Deborah Raji notes in an article for the MIT Technology Review[86]. As machine learning researcher Christabelle Pabalan points out regarding biased data, "the real hazard in machine learning has less to do with robotic conscious entities and more to do with another type of conscious entity - human beings."[87].

The errors caused by biased datasets in commercial uses of computer vision today give us a glimpse into the catastrophic consequences those may potentially have if deployed in a battlefield free of human supervision. In an experiment conducted by AlgorithmWatch it was discovered that, when presented with a picture of a black man holding a thermometer, Google's computer vision service Vision Cloud labelled it with the tags 'hand', 'gun' and 'firearm'[88]. However, when that same image was submitted with an overlay that turned the skin of the man white, the tags it was labelled with were 'hand' and 'monocular'[89]. The erroneous outcome in question, AlgorithmWatch argues, can most likely be drawn up to the fact that individuals with darker complexion were more prominently featured in images from the training dataset that depict violence, an occurrence that the machine learning model may have interpreted as a correlation between dark skin tones and violent behaviour[90]. It goes without saying that such

---

[84] *Ibid.*
[85] IBM Cloud Learning Hub, supra note 62.
[86] D. Raji, How Our Data Encodes Systematic Racism, https://www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion/, (accessed on 11.03.2021).
[87] C. Pabalan, Our Machine Learning Algorithms are Magnifying Bias and Perpetuating Social Disparities, https://towardsdatascience.com/our-machine-learning-algorithms-are-magnifying-bias-and-perpetuating-social-disparities-6beb6a03c939, (accessed on 06.03.2021).
[88] N. Kayser-Bril, Google Apologises After Its Vision AI Produced Racist Results, https://algorithmwatch.org/en/google-vision-racism/, (accessed on 06.03.2021).
[89] *Ibid.*
[90] *Ibid.*

errors could potentially translate into serious violations of basic principles of international humanitarian law, such as the principle of non-discrimination[91] and distinction[92].

Even though computer vision is far from being precise enough to be used in a battlefield, the problem of racism and bias in AI must be faced right away. With the use of image recognition technologies set to become more and more prominent in aiding human operators to analyse hours of drone footage faster[93], paying close attention to who labels that data and how will be of utmost significance.

The data gathered by military forces today could become the training data of the autonomous weapon systems of tomorrow. Now more than ever in the history of warfare, attention must be paid to the workers in military innovation, from the data scientists and machine learning engineers to those who are in charge of collecting and labelling data.

While talks on the regulation of future autonomous weapon systems continue, the international community has a duty to nip the problem at the bud by ensuring that all countries investing in automation for their military services and their tech workers operate with a clear understanding of the consequences their work may have in a distant – although we don't know exactly how – future.

4 AI uses in the military today: a concerning glimpse into the future

The world is already looking at AI, machine learning and neural networks with growing interest. Although the United States[94], United Kingdom[95] and the European Union's[96] member states do not currently have LAWS in their arsenals, nor are they intently working on their development, it is no secret that most major military powers are actively investing in new technologies set to introduce automation in the daily lives of their military forces.

---

[91] J. Pejic, Non-Discrimination and Armed Conflict, in: International Review of the Red Cross 83 (841) (2001), p. 184.
[92] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), supra note 59.
[93] K. Atherton, Targeting the Future of the DoD's Controversial Project Maven Initiative, https://www.c4isrnet.com/it-networks/2018/07/27/targeting-the-future-of-the-dods-controversial-project-maven-initiative/, (accessed on 09.03.2021).
[94] Congressional Research Service, Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, https://crsreports.congress.gov/product/pdf/IF/IF11150, (accessed on 07.03.2021), p. 1.
[95] Lt. Col. J. Stroud-Turp, supra note 26, p. 59.
[96] European Parliament, European Parliament Resolution on Autonomous Weapon Systems, 12 September 2018, P8_TA(2018)0341.

In November 2020, the United Kingdom's Prime Minister announced a £16.5 billion investment in the modernisation of the Armed Forces that will contribute, among others, to the development of "autonomous vehicles, swarm drones, and cutting-edge battlefield awareness systems […] for military use."[97]

Meanwhile, the US Department of Defence (DoD) published its artificial intelligence strategy on February 2019[98], wherein it expressed its intention to "[…] harness the potential of AI to transform all functions of the Department positively, thereby supporting and protecting U.S. servicemembers, safeguarding U.S. citizens, defending allies and partners, and improving the affordability, effectiveness, and speed of our operations."[99]. The strategy advocates, in particular, for the use of AI to "enhance military decision-making and operations across key mission areas"[100], for example by "[…] improving situational awareness and decision-making, increasing the safety of operating equipment, implementing predictive maintenance and supply, and streamlining business processes."[101].

Work to implement said strategy is already underway: take, for example, the $479 million contract that Microsoft has signed with the US Army to supply servicemen with 100,000 HoloLens augmented reality (AR) headsets, to be used in training and combat to "increase lethality by enhancing the ability to detect, decide and engage before the enemy."[102]. Several Microsoft employees expressed strong disapproval for the project in an open letter, demanding that their CEO drop the contract immediately and the company refrain from developing weapon technologies in the future[103].

A passage from that open letter strikes a sour note: as it is noted, the existence of an internal review process that ensures ethical standards in AI proved insufficient to prevent the work of hundreds of engineers being used for unethical purposes they did not agree upon[104]. As the

---

[97] Ministry of Defence, Defence Secures Largest Investment Since the Cold War, https://www.gov.uk/government/news/defence-secures-largest-investment-since-the-cold-war, (accessed on 07.03.2021).
[98] T. Moon-Cronk, US Dept. of Defense, DOD Unveils its Artificial Intelligence Strategy, https://www.defense.gov/Explore/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/, (accessed on 03.03.2021).
[99] United States Department of Defense, Summary of the 2018 Department of Defense Artificial Intelligence Strategy, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF, (accessed on 03.03.2021), p. 4.
[100] Id., p. 7.
[101] *Ibid.*
[102] C. Lecher, Microsoft Workers' Letter Demands Company Drop Army HoloLens Contract, https://www.theverge.com/2019/2/22/18236116/microsoft-hololens-army-contract-workers-letter, (accessed on 10.03.2021).
[103] *Ibid.*
[104] *Ibid.*

letter states, the CEO's suggestion that "employees concerned about working on unethical projects 'would be allowed to move to other work within the company' ignores the problem that workers are not properly informed of the use of their work. There are many engineers who contributed to HoloLens before this contract even existed, believing it would be used to help architects and engineers build buildings and cars, to help teach people how to perform surgery or play the piano […]. These engineers have now lost their ability to make decisions about what they work on, instead finding themselves implicated as war profiteers."[105].

In 2018, Google workers have expressed concern over the company's involvement in Project Maven, a DoD programme committed on the development of a computer vision model that will assist military personnel in the analysis and extraction of data from drone footage[106]. In an open letter to CEO Sundar Pichai, employees called for the withdrawal from the project and the implementation of a policy stating that Google will never build warfare technology[107]: consequently, Google announced that it will not renew the contract[108] and published a set of principles for AI development, wherein it pledged to never design AI for weapon technologies[109], only to clarify, in the following passage, that it will not cease to work with the military in other areas, such as cybersecurity and training[110].

Behind the stage on which Google employees expressed their rightful concerns on the impact of their work on Project Maven, thousands of ghost workers remained unheard of. As it has been unveiled by The Intercept, in October 2017 Google delegated the labelling process of the raw images that would then be used to train the algorithm developed for Project Maven to Figure Eight, a company that employs thousands of tech gig workers to perform small tasks that contribute to the development of AI systems, such as data labelling[111]. An executive at Figure Eight confirmed that the workers did not know they were working for Google or the US

---

[105] *Ibid.*

[106] C. Pellerin, Project Maven to Deploy Computer Algorithms to War Zone by Year's End, https://www.defense.gov/Explore/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/, (accessed on 10.03.2021).

[107] D. Deahl, Google Employees Demand the Company Pull Out of Pentagon AI Project, https://www.theverge.com/2018/4/4/17199818/google-pentagon-project-maven-pull-out-letter-ceo-sundar-pichai, (accessed on 10.03.2021).

[108] N. Statt, Google Reportedly Leaving Project Maven Military AI Program After 2019, https://www.theverge.com/2018/6/1/17418406/google-maven-drone-imagery-ai-contract-expire, (accessed on 10.03.2021).

[109] S. Pichai, AI at Google: Our Principles, https://blog.google/technology/ai/ai-principles/, (accessed on 10/03/2021).

[110] *Ibid.*

[111] L. Fang, Google Hired Gig Economy Workers to Improve Artificial Intelligence in Controversial Drone-Targeting Project, https://theintercept.com/2019/02/04/google-ai-project-maven-figure-eight/, (accessed on 10.03.2021).

Army, nor did they know what their work would contribute to, as the case usually is for every task they execute in the platform[112]: as one of the workers shared with The Intercept, they "[…] are given a reason for why they are doing a task, like, 'Draw boxes around a certain product to help machines recognize it,' but they are not given the company that receives the data."[113].

The fact that millions of tech workers do not know who or what their work benefits is not only morally questionable[114], but it signifies that a part of the work needed to develop AI systems as significant as the preparation of training data goes almost completely unregulated.

5 The way forward: the case for better regulation in the development phase of autonomous weapons

Our first steps in warfare's AI revolution should be accompanied by the knowledge that, now more than ever, how weapons are developed and by who is of the utmost importance. Regulating that process meaningfully, ideally by extending the ICRC's call to ensure human control over the functioning of automated weapons[115] to their development phase, would be a good start.

This has already been suggested by the International Panel on the Regulation of Autonomous Weapons (iPRAW), which recommended the CCW to recognise the principle of human control over autonomous weapons both in design and use[116]. The iPRAW then goes on to suggest the possible regulatory outcomes for the CCW's discussion, including (but not limited to) pre-emptive bans on the development and use of LAWS[117] and soft law tools such as guidelines or best practices for the "appropriate development and use or implementation of the technology"[118].

However, these tools could prove useless (as Project Maven teaches us) if the NGOs, tech companies and States involved in the discussion do not address openly how today's 'non-lethal' applications of AI in the military could affect future developments.

---

[112] *Ibid.*
[113] *Ibid.*
[114] *Ibid.*
[115] International Committee of the Red Cross, supra note 5.
[116] iPRAW, Concluding Report: Recommendations to the GGE, https://www.ipraw.org/publications/recommendations/, (accessed on 12.03.2021), p. 14.
According to the iPRAW, human control by design would imply the human's operator knowledge of the environment in which the weapon operates and the system itself, as to the inform human intervention in the targeting process.
[117] Id., p. 19.
[118] Id., p. 21.

The data that is being collected and labelled today for Project Maven and the US Army's HoloLens AR headsets will not go down the drain. Given that a poorly executed data collection and labelling process can cause AI systems to become perpetrators of existing racism and prejudice[119], and with non-discrimination being a fundamental tenet of international humanitarian law[120], the potential of AI to amplify social injustice in war zones must be openly addressed by the CCW if the technology is to be used in military equipment or even, as the ICRC has suggested, to aid humanitarian action[121].

Regardless of whether LAWS will be banned or even developed in the first place, AI has already taken position in the military's strategy plans. Are we ready for what is to come?

[119] D. Raji, supra note 86.
[120] J. Pejic, supra note 91.
[121] International Committee of the Red Cross, supra note 8, p. 6.

Bibliography

AI Wiki, Weights and Biases, https://docs.paperspace.com/machine-learning/wiki/weights-and-biases.

Campaign to Stop Killer Robots, About, https://www.stopkillerrobots.org/about/.

Campaign to Stop Killer Robots, Diplomatic Talks Re-Convene, https://www.stopkillerrobots.org/2020/09/diplomatic2020/.

Campaign to Stop Killer Robots, Learn: the solution, https://www.stopkillerrobots.org/learn/#problem.

C. Pellerin, Project Maven to Deploy Computer Algorithms to War Zone by Year's End, https://www.defense.gov/Explore/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

C. Pabalan, Our Machine Learning Algorithms are Magnifying Bias and Perpetuating Social Disparities, https://towardsdatascience.com/our-machine-learning-algorithms-are-magnifying-bias-and-perpetuating-social-disparities-6beb6a03c939.

C. Lecher, Microsoft Workers' Letter Demands Company Drop Army HoloLens Contract, https://www.theverge.com/2019/2/22/18236116/microsoft-hololens-army-contract-workers-letter.

Congressional Research Service, Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems, https://crsreports.congress.gov/product/pdf/IF/IF11150.

D. Deahl, Google Employees Demand the Company Pull Out of Pentagon AI Project, https://www.theverge.com/2018/4/4/17199818/google-pentagon-project-maven-pull-out-letter-ceo-sundar-pichai.

D. Raji, How Our Data Encodes Systematic Racism, https://www.technologyreview.com/2020/12/10/1013617/racism-data-science-artificial-intelligence-ai-opinion/.

DeepAI, Artificial Intelligence, https://deepai.org/machine-learning-glossary-and-terms/artificial-intelligence.

European Parliament, European Parliament Resolution on Artificial Intelligence: Questions of Interpretation and Application of International Law, 20 January 2021, P9_TA(2021)0009.

European Parliament, European Parliament Resolution on Autonomous Weapon Systems, 12 September 2018, P8_TA(2018)0341.

Human Rights Watch, Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control, https://www.hrw.org/news/2020/08/10/killer-robots-growing-support-ban.

IBM Cloud Learn Hub, Machine Learning, https://www.ibm.com/cloud/learn/machine-learning#toc-what-is-ma-qhM6PX35.

IBM Cloud Learn Hub, Strong AI, https://www.ibm.com/cloud/learn/strong-ai.

IBM Cloud Learn Hub, What is Supervised Learning?, https://www.ibm.com/cloud/learn/supervised-learning#toc-how-superv-A-QjXQz-.

IBM Cloud Learn Hub, AI vs. Machine Learning vs. Neural Networks: What's the Difference?, https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks.

IBM Cloud Learn Hub, Deep Learning, https://www.ibm.com/cloud/learn/deep-learning.

IBM Cloud Learn Hub, Unsupervised Learning, https://www.ibm.com/cloud/learn/unsupervised-learning#toc-what-is-un-MP1gM75c.

IBM, Computer Vision, https://www.ibm.com/topics/computer-vision.

Information Age, Can Artificial Intelligence Spot Spam Quicker Than Humans?, https://www.information-age.com/artificial-intelligence-spam-machine-learning-123481368/.

International Committee of the Red Cross, A Guide to the Legal Review of New Weapons, Means and Methods of Warfare Measures to Implement Article 36 of Additional Protocol I of 1977, https://www.icrc.org/en/publication/0902-guide-legal-review-new-weapons-means-and-methods-warfare-measures-implement-article.

International Committee of the Red Cross, Artificial Intelligence and Machine Learning in Armed Conflict: a Human-Centered Approach, https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach.

International Committee of the Red Cross, Autonomy, Artificial Intelligence and Robotics: Technical Aspects of Human Control, https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control.

iPRAW, Concluding Report: Recommendations to the GGE, https://www.ipraw.org/publications/recommendations/.

J. Vincent, This is When AI's Top Researchers Think Artificial Intelligence Will be Achieved, https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book.

J. Pejic, Non-Discrimination and Armed Conflict, in: International Review of the Red Cross 83 (841) (2001).

J. McCarthy et al., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, in AI Magazine 27 (4) (2006), p. 12, https://doi.org/10.1609/aimag.v27i4.1904.

K. Atherton, Targeting the Future of the DoD's Controversial Project Maven Initiative, https://www.c4isrnet.com/it-networks/2018/07/27/targeting-the-future-of-the-dods-controversial-project-maven-initiative/.

L. Fang, Google Hired Gig Economy Workers to Improve Artificial Intelligence in Controversial Drone-Targeting Project, https://theintercept.com/2019/02/04/google-ai-project-maven-figure-eight/.

Lt. Col. J. Stroud-Turp, Lethal Autonomous Weapon Systems (LAWS): Speaker's Summary, in: Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons. Expert Meeting, Versoix, Switzerland, 15-16 March 2016, p. 57, https://www.icrc.org/en/publication/4283-autonomous-weapons-systems.

Ł. Gebel, Why We Need Bias in Neural Networks, https://towardsdatascience.com/why-we-need-bias-in-neural-networks-db8f7e07cb98.

M. W. Meier, Autonomous Legal Reasoning?: Legal and Ethical Issues in the Technologies of Conflict: Lethal Autonomous Weapons Systems (LAWS): Conducting a Comprehensive Weapons Review, 30 Temp. Int'l & Comp. L.J. 119, Spring 2016.

M. W. Meier, CCW LAWS Meeting: U.S. Closing Statement and the Way Ahead, https://geneva.usmission.gov/2015/05/08/ccw-laws-meeting-u-s-closing-statement-and-the-way-ahead/.

Ministry of Defence, Defence Secures Largest Investment Since the Cold War, https://www.gov.uk/government/news/defence-secures-largest-investment-since-the-cold-war.

N. Statt, Google Reportedly Leaving Project Maven Military AI Program After 2019, https://www.theverge.com/2018/6/1/17418406/google-maven-drone-imagery-ai-contract-expire.

N. Kayser-Bril, Google Apologises After Its Vision AI Produced Racist Results, https://algorithmwatch.org/en/google-vision-racism/.

Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977, 1125 UNTS 3.

ScienceDaily, Algorithm that Performs as Accurately as Dermatologists, https://www.sciencedaily.com/releases/2021/02/210212111902.htm.

Stanford University, Professor John McCarthy: Father of AI, http://jmc.stanford.edu/general/index.html.

S. Pichai, AI at Google: Our Principles, https://blog.google/technology/ai/ai-principles/.

T. Moon-Cronk, DOD Unveils its Artificial Intelligence Strategy, https://www.defense.gov/Explore/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/.

United Nations, United Kingdom's Opening Remarks at the First Session of the GGE on LAWS (21 to 25 September 2020), https://documents.unoda.org/wp-content/uploads/2020/09/LAWS-Sept-GGE-2020-UK-Opening-Statement.pdf.

US Dept. of Defense, Summary of the 2018 Department of Defense Artificial Intelligence Strategy, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF