State responsibility in relation to military applications of artificial intelligence

Boutin, B.

DOI
10.1017/S0922156522000607
Publication date
2023
Document Version
Final published version
Published in
Leiden Journal of International Law
License
CC BY

Leiden Journal of International Law (2023), **36**, 133–150
doi:10.1017/S0922156522000607

ORIGINAL ARTICLE

INTERNATIONAL LAW AND PRACTICE

State responsibility in relation to military applications of artificial intelligence

Bérénice Boutin*

Asser Institute, R.J. Schimmelpenninclaan 20-22, 2517 JN, The Hague, The Netherlands
Email: b.boutin@asser.nl

Abstract

This article explores the conditions and modalities under which a state can incur responsibility in relation to violations of international law involving military applications of artificial intelligence (AI) technologies. While the question of how to attribute and allocate responsibility for wrongful conduct is one of the central contemporary challenges of AI, the perspective of state responsibility under international law remains relatively underexplored. Moreover, most scholarly and policy debates have focused on questions raised by autonomous weapons systems (AWS), without paying significant attention to issues raised by other potential applications of AI in the military domain. This article provides a comprehensive analysis of state responsibility in relation to military AI. It discusses state responsibility for the wrongful use of AI-enabled military technologies and the question of attribution of conduct, as well as state responsibility prior to deployment, for failure to ensure compliance of AI systems with international law at the stages of development or acquisition. Further, it analyses derived state responsibility, which may arise in relation to the conduct of other states or private actors.

Keywords: artificial intelligence; attribution of conduct; autonomous weapons systems; duty to ensure respect; state responsibility

1. Introduction

The question of how to attribute and allocate responsibility in case of wrongful conduct is one of the central contemporary challenges of AI. Due to their inherent characteristics, notably in terms of autonomy and unpredictability, advanced AI systems typically raise difficult issues of accountability, which have long been discussed in the fields of law, ethics of technology, and computer science.[1] Responsibility in relation to AI remains a prevalent issue in scholarly and policy debates, as advancements in AI research and the increasingly widespread use of AI technologies have highlighted the broad societal and policy implications of AI.[2]

*Research for this article was conducted as part of the project 'Conceptual and Policy Implications of Increasingly Autonomous Military Technologies for State Responsibility under International Law', which received funding from the Gerda Henkel Foundation (2018–2019), and the project 'Designing International Law and Ethics into Military Artificial Intelligence' (DILEMA), which received funding from NWO (2020–2024).

[1]Early works on the topic of AI responsibility include: B. Friedman and P. H. Kahn, 'Human Agency and Responsible Computing: Implications for Computer System Design', (1992) 17 *Journal of Systems and Software* 7; A. Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata', (2004) 6 *Ethics and Information Technology* 175.

[2]See, for instance, the work of the EU High-Level Expert Group on Artificial Intelligence, which places accountability for AI systems as a key requirement for socially beneficial AI: *Ethics Guidelines for Trustworthy Artificial Intelligence* (2019),

In the military context, discussions surrounding AI technologies have primarily focused on autonomous weapons systems (AWS), colloquially referred to as 'killer robots'. International legal scholarship on the topic has questioned whether AWS could be able to be used in compliance with international humanitarian law (IHL), in particular with the principles of proportionality and distinction, and whether the deployment of AWS could lead to a responsibility gap at the individual level.[3] By contrast, the question of state responsibility under international law[4] in relation to AWS and other AI-based technologies has remained relatively underexplored. Several authors have touched upon aspects of state responsibility and indicated that it is a useful mechanism to ensure compliance with international law and accountability in relation to AI,[5] but did not extensively analyse how the law of state responsibility applies to violations of international law caused with AI-enabled military technologies.

Although relatively underexplored in scholarship, state responsibility has an important role to play as part of an overall framework of accountability in relation to military AI. Indeed, upholding the responsibility of collective actors such as states acknowledges the structural forces that drive the development and use of AI.[6] State responsibility is a particularly important perspective in relation to military AI, because states constitute the primary structures within which such AI systems are developed, regulated, and deployed. While some authors have argued that state responsibility is not a satisfactory way to address accountability in relation to AWS,[7] or conversely that state responsibility is 'preferable' to individual liability,[8] both individual and state responsibility operate concurrently and complementarily.[9] Moreover, state responsibility

at 19–20; and the extensive study drafted by Professor Karen Yeung for the Council of Europe, *Responsibility and AI: A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework* (2019), Council of Europe Study DGI(2019)05.

[3]See, e.g., M. N. Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics', (2013) 4 *Harvard National Security Journal Features* 1; M. Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified', (2014) 90 *International Law Studies* 308; J. van den Boogaard, 'Proportionality and Autonomous Weapons Systems', (2015) 6 *Journal of International Humanitarian Legal Studies* 247; C. J. Dunlap, 'Accountability and Autonomous Weapons: Much Ado About Nothing?', (2016) 30 *Temple International and Comparative Law Journal* 63; E. Winter, 'Autonomous Weapons in Humanitarian Law: Understanding the Technology, Its Compliance with the Principle of Proportionality and the Role of Utilitarianism', (2018) 6 *Groningen Journal of International Law* 183.

[4]As authoritatively codified by the International Law Commission (ILC): ILC Draft Articles on the Responsibility of States for Internationally Wrongful Acts, 2001 YILC, Vol. II (Part Two), at 26–30 ('ILC Articles' or 'ARSIWA'); ILC Draft Articles on the Responsibility of States for Internationally Wrongful Acts with Commentaries, 2001 YILC, Vol. II (Part Two), at 31–143 ('ARSIWA commentaries').

[5]R. Crootof, 'War Torts: Accountability for Autonomous Weapons', (2016) 164 *University of Pennsylvania Law Review* 1347, at 1389–93; K. Anderson and M. C. Waxman, 'Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation under International Law', in R. Brownsword, E. Scotford and K. Yeung (eds.), *The Oxford Handbook of Law, Regulation, and Technology* (2017), 1097, at 1110; D. A. Lewis, G. Blum and N. K. Modirzadeh, 'War-Algorithm Accountability', (2016) *Research Briefing, Harvard Law School Program on International Law and Armed Conflict*, at 83; NATO JAPCC, *Future Unmanned System Technologies: Legal and Ethical Implications of Increasing Automation* (2016), at 30; J. G. Castel and M. E. Castel, 'The Road to Artificial Super-Intelligence Has International Law a Role to Play', (2016) 14 *Canadian Journal of Law and Technology* 1, at 9; Human Rights Watch and Harvard Law School's International Human Rights Clinic, *Mind the Gap: The Lack of Accountability for Killer Robots* (2015), at 13; R. Geiss, 'Autonomous Weapons Systems: Risk Management and State Responsibility', (2016) *Short Paper for the Third CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva*, at 3.

[6]H. Hazenberg, B. Boutin and J. van den Hoven, 'Artificial Intelligence and International Law: Exploring Issues of Responsibility and Regulation', (2018) *Position Paper, TU Delft Pilot Program on Responsible Innovation for the Sustainable Development Goals*, at 7.

[7]D. Amoroso, '*Jus In Bello* and *Jus Ad Bellum* Arguments against Autonomy in Weapons Systems: A Re-Appraisal', (2017) *QIL: Questions of International Law* 5, at 21–2.

[8]D. N. Hammond, 'Autonomous Weapons and the Problem of State Accountability', (2015) 15 *Chicago Journal of International Law* 652, at 656–7.

[9]A. Nollkaemper, 'Concurrence between Individual Responsibility and State Responsibility in International Law', (2003) 52 *International and Comparative Law Quarterly* 615.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

comports some specific features that are particularly relevant to address accountability in relation to AI. One is that state responsibility entails not only an obligation to provide reparation, but also an obligation to cease the wrongful conduct and to offer appropriate assurances and guarantees of non-repetition.[10] Another is that state responsibility can arise prior to the deployment of military AI, as the framework also applies at the stages of development or procurement of military technologies.[11] State responsibility is thus not only relevant for *ex post* liability for wrongful conduct on the battlefield, but more generally for the international governance of AI throughout its lifecycle.

The present article sets to comprehensively address issues of state responsibility that arise as states increasingly integrate AI technologies in their military apparatus. The overarching research question is to determine in which circumstances a state can incur responsibility in relation to violations of international law involving military AI. Although the article discusses the topic in relation to military AI, many of the arguments advanced are relevant more broadly to the use and regulation of AI by states beyond the military context. Before delving into the analysis of state responsibility, Section 2 provides some necessary background on AI technologies and their military applications. Sections 3 to 5 explore three dimensions of state responsibility in relation to military AI. Section 3 addresses attribution of conduct involving AI, questioning whether and on which basis wrongful conduct occurring when AI-enabled systems are deployed on the battlefield can be attributed to the state for the purpose of international responsibility. Analysing the function and conceptual basis of attribution of conduct against the background of scholarship on the interaction of human agents and AI technologies, it proposes a framework for attribution of conduct involving AI. Section 4 analyses responsibility prior to deployment, at the stage of development or procurement of AI. It argues in favour of a compliance-by-design approach, under which states incur responsibility if they fail to ensure that AI systems are developed in compliance with their international obligations and designed in a way that embed these obligations. Section 5 completes the analysis by discussing grounds of derived state responsibility, which can potentially arise in relation to the conduct of other states or private actors which develop or deploy military AI. Section 6 offers concluding remarks and perspectives on how to operationalize the legal framework of state responsibility in relation to military AI.

2. Artificial intelligence in the military context, beyond autonomous weapons systems

AI can be defined as computer systems able to perform tasks that traditionally only humans could perform, such as rational reasoning, problem-solving and decision-making. It is based on algorithms, which are sets of mathematical instructions aimed at performing a specific task.[12] AI technologies are being used in areas ranging from video games, finance, and online commercial targeting, to healthcare, public welfare policy, border control, and criminal justice. One of the most controversial applications of AI is in the military sphere. Fuelled by popular imaginary and fears of machines overtaking the battlefield, a broad public debate has been taking place on AWS,[13] commonly defined as weapons systems 'that, once activated, can select and engage

[10]See ARSIWA, *supra* note 4, at Arts. 30 and 31. The legal consequences of state responsibility arise irrespective of a claim invoking responsibility, see ARSIWA commentaries, *supra* note 4, at 91 (Commentary to Art. 31 ARSIWA, para. 4).

[11]See Section 4, *infra*.

[12]S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (2010), at 1–16; A. Rubel, C. Castro and A. Pham, *Algorithms and Autonomy* (2021), at 8.

[13]See, in particular, the multilateral debates taking place in the context of the Governmental Experts (GGE) on lethal autonomous weapons systems (LAWS) of the United Nations Convention on Certain Conventional Weapons (CCW), and the work of the 'Campaign to Stop Killer Robots', a coalition of non-governmental organizations actively lobbying for a ban on AWS.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

targets without further intervention by a human operator'.[14] The key characteristic of AWS, which is at the core of critical debates on autonomy in weapons and human control over AWS, is the inclusion of AI technology into weapons systems.[15] Indeed, AWS in their common understanding are weapons systems which embed algorithms specifically aimed at identifying military targets. However, current and potential applications of AI in the military domain are not limited to the relatively narrow category of weapons systems. As will be further elaborated on below, AI technologies are and could be integrated throughout the realm of military activities and in support of varied types of military functions. The term 'military AI' as used in this contribution includes AI-enabled weapon systems as well as various other potential uses of AI in the military.

AI technologies can be broadly classified under two methodological categories. Traditionally, AI has been developed with rule-based algorithms, where each instruction is explicitly programmed in lines of code, under the format 'if, then'. For instance, to develop an algorithm that can recognize images of apples, the programmer would encode each and every detailed characteristic and feature that identifies an apple (e.g., its shape, colour, how it differentiates from a pear or a peach, etc.). This technique has the advantage of being fully explainable and predictable, but it is limited in what can be achieved as it is impossible to define and code every possibility.[16] More recently, data-driven machine learning techniques have been developed to increase and broaden the capabilities of AI systems. Instead of explicit instructions, machine learning algorithms are developed by providing a computer with large sets of data (in our example, images of apples and other fruits) and letting the system identify potential correlations through a process known as training. Applying statistical analysis to the data provided, the algorithm identifies and generalizes common features and recurring patterns, and thereby can 'learn' by itself how to recognize apples.[17] Machine learning techniques have enabled major breakthroughs in recent years, with AI systems able to perform a growing range of tasks including object or facial recognition, language processing, strategic reasoning, and predictive analysis.

However, AI systems based on machine learning come with a number of novel and specific characteristics that raise critical ethical, legal, and policy challenges. First, machine learning models produce results that are usually not explainable. This is known as the 'black box' dilemma, whereby AI systems produce results that cannot be explained or justified by either developers or users.[18] Second, AI systems are, by purpose, operating with a degree of autonomy that can escape direct human control. Advanced self-learning algorithms further complicate the picture as they can exhibit unpredictable or unexpected behaviour.[19] Third, AI operates at a speed and scale that goes beyond human cognitive possibilities. It can process incommensurably vast amount of data in a split-second, so that, even if an AI system is formally under human control or supervision, its results are not intelligible to human operators. This characteristic of AI systems is known to lead to the issue of automation bias: when a human operator is vested with the formal role of approving or not recommendations made by a decision-support algorithm, they do not have the genuine capacity to evaluate the suggested outcome and therefore tend to simply follow the system

[14]US Department of Defense, 'Autonomy in Weapons Systems', (2012) *Directive 3000.09*, at 13. Similar terminology is used by the ICRC and Human Rights Watch and has become widely accepted, see M. Ekelhof, 'Complications of a Common Language: Why it is so Hard to Talk about Autonomous Weapons', (2017) 22 *Journal of Conflict and Security Law* 311, at 322.

[15]D. A. Lewis, G. Blum and N. K. Modirzadeh, 'War-Algorithm Accountability', (2016) *Harvard Law School Program on International Law and Armed Conflict Research Briefing*, at 10.

[16]Russell and Norvig, *supra* note 12, at 22–4.

[17]C. Shah, *A Hands-On Introduction to Data Science* (2020), at 210–14.; Rubel, Castro and Pham, *supra* note 12, at 10.

[18]D. Castelvecchi, 'Can We Open the Black Box of AI?', (2016) 538 *Nature News* 20; J. Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms', (2016) 3 *Big Data & Society* 1; A. Holland Michel, 'The Black Box, Unlocked: Predictability and Understandability in Military AI', (2020) *UNIDIR Report*.

[19]R. V. Yampolskiy, 'Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent', (2020) 7(1) *Journal of Artificial Intelligence and Consciousness* 109.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

recommendations.[20] Fourth, technology remains the produce of human choices,[21] and decisions made at the stage of AI design and development have significant implications.[22] Systems' requirements define the parameters within which an AI system operates and how it interacts with human operators. For instance, AI systems used for reasoning and decision-making are developed and trained against goals and values that are specified by programmers.[23] It is also from choices made during the design stage (e.g., which datasets are considered representative) that issues of discriminatory bias can arise.[24] Fifth, it is important to understand that AI technology is not intelligent in the way that humans are. Algorithms do not understand the results they produce, in the sense that they cannot understand what an apple is, and cannot interpret data and results.[25] The only thing AI can 'see' are pixels, and what it can perceive as an important characteristic might be insignificant or unreliable for a human observer.[26] This is also why image recognition algorithms are subject to spoofing through adversarial attacks that slightly modify an image to lead the algorithm to make mistakes.[27]

Applied to the military, the current advancements in AI research can be used for navigation, for surveillance and recognition, in automated defence systems, to analyse satellite or drone imagery, to perform terrain, object, or facial recognition during operations, to estimate the potential damage of an attack, to enable robotic swarming, to detect and predict adversary conduct, anomalous activities, potential risks, and possible counterstrategies, and to provide decision support at the operational but also strategic level. The potential applications of AI in the military go far beyond purely operational matters involving targeting and offensive lethal operations, and AI technologies could be deployed throughout military planning and operations.[28]

[20]M. L. Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems', (2004) *American Institute of Aeronautics and Astronautics, Intelligent Systems Technical Conference*, available at doi.org/10.2514/6.2004-6313; P. Scharre, 'Autonomous Weapons and Operational Risk', (2016) *CNAS Paper*, at 31; D. Amoroso and G. Tamburrini, 'What Makes Human Control over Weapons "Meaningful"?', (2019) *ICRAC Report*, at 9; Morgan et al., 'Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World', (2020) *RAND Corporation*, at 36.

[21]D. G. Johnson, 'Computer Systems: Moral Entities but not Moral Agents', (2006) 8(4) *Ethics and Information Technology* 195, at 197. See also J. Bryson, 'Patiency Is Not a Virtue: The Design of Intelligent Systems and Systems of Ethics', (2018) 20 *Ethics and Information Technology* 15, at 21.

[22]J. van den Hoven, 'Value Sensitive Design and Responsible Innovation', in R. Owen, J. Bessant and M. Heintz (eds.), *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (2013), 75.

[23]N.-M. Aliman and L. Kester, 'Requisite Variety in Ethical Utility Functions for AI Value Alignment', (2019) 2419 *CEUR Workshop Proceedings*.

[24]The issue of racial and gender bias in both data sets and programming choices has been widely reported. See, notably, J. Angwin et al., 'Machine Bias', *ProPublica*, 23 May 2016, available at www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; M. Garcia, 'Racist in the Machine: The Disturbing Implications of Algorithmic Bias', (2017) 33(4) *World Policy Journal* 111; N. Turner Lee, P. Resnick and G. Barton, 'Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms', *Brookings Report*, 22 May 2019, available at www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/; S. Wachter, B. Mittelstadt and C. Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI', *SSRN*, 27 March 2020, available at www.ssrn.com/abstract=3547922.

[25]By contrast, in an example provided in a conversation with Giovanni Sileno (University of Amsterdam), a toddler is able to recognize a giraffe in a zoo after seeing just a few simplified children's illustrations of the animal.

[26]For example, in an experiment, an algorithm trained to recognize and differentiate between wolves and huskies identified the presence of snow in the background as a defining feature (M. Tulio Ribeiro, S. Singh and C. Guestrin, '"Why Should I Trust You?": Explaining the Predictions of Any Classifier', 2016, available at www.arxiv.org/abs/1602.04938).

[27]OpenAI, 'Attacking Machine Learning with Adversarial Examples', 24 February 2017, available at www.openai.com/blog/adversarial-example-research/.

[28]NATO JAPCC, *Future Unmanned System Technologies: Legal and Ethical Implications of Increasing Automation* (2016), at 4–5; UNIDIR, *The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence* (2018), at 8; Morgan et al., *supra* note 20, at 17–20; Persi Paoli et al., 'Modernizing Arms Control: Exploring Responses to the Use of AI in Military Decision-Making', (2020) UNIDIR, at 5–6; K. McKendrick, 'The Application of Artificial Intelligence in Operations Planning', (2017) *NATO Science and Technology Organization, STO-MP-SAS-OCS-ORA-2017*; W. A. Branch, 'Artificial Intelligence and Operational-Level Planning: An Emergent Convergence', (2018) *US Army, School of Advanced Military Studies*; French Ministry of Defence, 'Artificial Intelligence in Support of Defence', (2019) *Report of the AI Task*

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

Recent military practice offers a number of examples of the integration of AI in diverse and complex ways. Project Maven is an initiative of the US Department of Defense, that seeks to use AI to alleviate the cognitive burden of mundane repetitive tasks such as video analysis. It uses machine learning to process drone footage, produce actionable data, and enhance military decision-making.[29] AI can also be used to automate or delegate certain tasks that would be difficult or dangerous for human soldiers. For instance, the Maritime Mine Countermeasures (MMCM) programme seeks to develop an AI-enabled unmanned underwater anti-mine warfare system that can detect, identify, and neutralize naval mines.[30] Another example is the ViDAR Maritime system, which is used to perform wide area optical search to detect and classify objects at sea.[31] Other military applications of AI often seek to enable more rapid (and allegedly more accurate) decision-making in a time-critical environment. For instance, Project Convergence is a multi-platform project aimed at integrating AI in battlefield management systems. It consists in drones feeding real-time reconnaissance data to network algorithms which combine this data with other information to update digital maps, identify and prioritize potential targets, match threats to the best weapon, and send targeting information to fire control systems. The system thereby dramatically increases the speed at which ground operators can act on the basis of a wide range of data.[32] While such decision-support systems provide recommendations that are formally subject to approval by a human operator, the characteristics of AI described above indicate that human oversight over complex AI systems may be superficial in practice.[33] Other potential applications of AI include predictive algorithms in support of decision-making regarding detention[34] or combat medical triage,[35] facial recognition algorithms to help identifying war casualties,[36] and the use of AI to generate deepfakes as part of information warfare.[37]

Force, at 15; Ministry of Defence of the Netherlands, *Defence Vision 2035: Fighting for a safer future* (2020), Annex 1, at III; National Security Commission on Artificial Intelligence (NSCAI), *Final Report* (2021), available at www.nscai.gov/2021-final-report, at 79; French Ministry of Defence, 'Artificial Intelligence in Support of Defence', (2019) *Report of the AI Task Force*, at 14–19.

[29]US Deputy Secretary of Defense, 'Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven),' *Memorandum, 26 April 2017*, available at www.govexec.com/media/gbc/docs/pdfs_edit/establishment_of_the_awcft_project_maven.pdf.

[30]Thales Group, 'The Maritime Mine Countermeasures Programme: The French and British Navies Blaze the Trail Towards a Global First with Their Revolutionary Autonomous System', 13 September 2019, available at www.thalesgroup.com/en/worldwide-defence/naval-forces/magazine/maritime-mine-countermeasures-programme-french-and-british; X. Vavasseur, 'French Navy's SLAMF Unmanned Mine Warfare System to be Qualified in December', *Naval News*, 10 August 2019, available at www.navalnews.com/naval-news/2019/08/french-navys-slamf-unmanned-mine-warfare-system-to-be-qualified-in-december/.

[31]'Sentient Vision Systems selected for US FCT contract', *Australian Defence Magazine*, 16 March 2022, available at www.australiandefence.com.au/defence/sea/sentient-vision-systems-selected-for-us-fct-contract.

[32]J. Lacdan, 'Project Convergence Aims to Accelerate Change in Modernization Efforts', *Army News Service*, 11 September 2020, available at www.army.mil/article/238960/project_convergence_aims_to_accelerate_change_in_modernization_efforts; S. J. Freedberg Jr, 'Target Gone in 20 Seconds: Army Sensor-Shooter Test', *Breaking Defense*, 10 September 2020, available at www.breakingdefense.com/2020/09/target-gone-in-20-seconds-army-sensor-shooter-test/; S. J. Freedberg Jr, 'Kill Chain In The Sky With Data: Army's Project Convergence', *Breaking Defense*, 14 September 2020, available at www.breakingdefense.com/2020/09/kill-chain-in-the-sky-with-data-armys-project-convergence/.

[33]See Section 3.2, *infra*.

[34]D. A. Lewis, 'AI and Machine Learning Symposium: Why Detention, Humanitarian Services, Maritime Systems, and Legal Advice Merit Greater Attention', *Opinio Juris*, 28 April 2020, available at www.opiniojuris.org/2020/04/28/ai-and-machine-learning-symposium-ai-in-armed-conflict-why-detention-humanitarian-services-maritime-systems-and-legal-advice-merit-greater-attention.

[35]'Military Researchers to Apply Artificial Intelligence (AI) and Machine Learning to Combat Medical Triage', Militaryaerospace, 18 March 2022, available at www.militaryaerospace.com/computers/article/14248148/artificial-intelligence-ai-machine-learning-combat-medical-triage.

[36]P. Dave and J. Dastin, 'Ukraine has started using Clearview AI's facial recognition during war', *Reuters*, 14 March 2022, available at www.reuters.com/technology/exclusive-ukraine-has-started-using-clearview-ais-facial-recognition-during-war-2022-03-13/.

[37]H. Nasu, 'Deepfake Technology in the Age of Information Warfare', *Lieber Institute*, 1 March 2022, available at www.lieber.westpoint.edu/deepfake-technology-age-information-warfare/.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

Military applications of AI are broad and diverse, whereby AI is integrated in a highly distributed and subtle manner that does not always fit the narrative of AWS. Rather than a single AI system attached to a specific platform for a specific task, AI is set to become integrated in a variety of military tasks and at multiple stages of the decision-making chain. As a result, it becomes difficult to draw a clear line between 'lethal' and 'non-lethal' AI applications. For instance, data generated through an AI system initially used for mere surveillance might later become part of a targeting process.

In view of the unique characteristics of AI technologies, and the diverse and increasing use of these technologies in the military context, the following sections discuss the implications of AI in terms of state responsibility under international law.

3. State responsibility in the deployment of artificial intelligence systems: Attribution of wrongful conduct

When the deployment of AI systems on the battlefield results in violations of international obligations,[38] the responsibility of the state can be engaged if it is demonstrated that the wrongful conduct in question is attributable to the state.[39] The notion of attribution of conduct is a cornerstone of the law of state responsibility. Essentially, attribution of conduct consists in attaching to the state the actions or omissions of individuals or entities acting on its behalf.[40] Indeed, it is an 'elementary fact that the State cannot act of itself',[41] and 'what constitutes an "act of the State" for the purposes of State responsibility'[42] necessarily results from the actions or omissions of human beings acting on behalf of the state. In order to develop a framework for attribution of conduct involving AI systems (Section 3.3), the following paragraphs embark in an analysis of the human dimension of attribution of conduct (Section 3.1), read against the background of scholarship on the interaction of human agents and AI technologies (Section 3.2).

3.1 The human dimension of attribution of conduct in the ARSIWA

The law of state responsibility unequivocally hinges upon actions or omissions by human beings. Attribution is based on the conduct of 'human beings',[43] 'agents and representatives',[44] 'persons',[45] 'individuals'.[46] While the ARSIWA also occasionally refer to the conduct of a 'group of persons',[47] 'corporations or collectivities',[48] or 'entities',[49] these concern collective entities that are constituted by human beings. The law of state responsibility does not envisage conduct other than the conduct of human beings, acting individually or collectively.[50] Under the ILC framework, the existence of human conduct is therefore a precondition for state responsibility. A harmful outcome that does not originates in an action or omission by one or more human being(s) cannot engage responsibility.

[38]This section focuses on secondary rules of responsibility and does not address in detail the primary norms that can be potentially violated when using military AI technologies. Relevant norms are primarily found in the law of armed conflict and complementarily or subsidiarily in other regimes such as international human rights law.

[39]See ARSIWA, *supra* note 4, Art. 2.

[40]See ARSIWA commentaries, *supra* note 4, at 36 (Commentary to Art. 2, para. 12).

[41]See ibid., at 35 (Commentary to Art. 2, para. 5).

[42]See ibid., at 35 (Commentary to Art. 2 para. 5).

[43]See ibid., at 38 (Commentary to Ch. II of Part I, para. 2) and 35 (Commentary to Art. 2, para. 5).

[44]See ibid., at 35 (Commentary to Art. 2, para. 5).

[45]Ibid.

[46]See Ibid., at 40 (Commentary to Art. 4, para. 1).

[47]See ARSIWA, *supra* note 4, at Arts. 8 and 9.

[48]See ARSIWA commentaries, *supra* note 4, at 38 (Commentary to Ch. II of Part I, para. 2).

[49]See ARSIWA, *supra* note 4, at Art. 5.

[50]See ARSIWA commentaries, *supra* note 4, at 40 (Commentary to Art. 4, para. 1) and 49 (Commentary to Art. 8, para. 9).

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

Pushing this line of reasoning further, it can be argued that attribution relies on a causal link between actions or omissions by a human being and the occurrence of a breach of international law. Responsibility can be engaged if human conduct (that is attributable to the state) caused or contributed to cause a breach. If the operator of an AI system has no control over the outcome, if the machine operates to a large extent autonomously so that actions or omissions of human operators are not causally linked to the breach, it can be argued that there is no human conduct on the part of the operator to form the basis of attribution. It is therefore important to assess whether violations of international law caused with the use of AI systems can be traced back to human actions and omissions, and in turn to the state.

Some authors consider that the question of attribution in the context of AI is straightforward,[51] since, under the law of state responsibility, any and all conduct of state organs (e.g., armed forces) performed in their official capacity is attributable to the state,[52] whether or not AI is used. However, the critical aspect when it comes to AI and state responsibility is precisely whether there exists a 'human conduct' in the first place. Even though, unlike individual criminal responsibility for war crimes and other violations of IHL,[53] state responsibility is objective in nature and does not require demonstrating subjective intent or fault on the part of the state organ,[54] it relies nonetheless on a human element, which might lead to specific challenges in relation to AI technologies.

Since the operation of attribution of conduct relies on human conduct in order to attach violations of international law to a state, the question is whether and how the characteristics of AI systems, including autonomy, speed, unpredictability, and opacity, could affect or complicate the operation of attribution of conduct. In other words, what counts as 'human conduct' for the purpose of attribution when AI is involved?

3.2 What counts as 'human conduct' in socio-technical entanglements

The question of what constitutes (human) 'conduct' is not directly addressed in the law of state responsibility.[55] The identification of human conduct in relation to a breach is usually self-evident from the facts. However, the use of complex, decentralized, and increasingly autonomous technologies such as AI reconfigures the question, and requires us to analyse what qualifies as 'human conduct' in the context of state responsibility.

Decades of studies in philosophy and ethics of technology have demonstrated that AI technologies can affect human autonomy and reduce human control over outcomes.[56] In the debates on AWS, this concern is reflected in the idea that AWS should remain under 'meaningful human control'.[57] In particular, AI systems that integrate machine learning algorithms – which generate

[51]Crootof, *supra* note 5, at 1391; R. Geiß, 'State Control Over the Use of Autonomous Weapon Systems: Risk Management and State Responsibility', in Bartels et al. (eds.), *Military Operations and the Notion of Control Under International Law* (2021), 439, at 448; Human Rights Watch and Harvard Law School's International Human Rights Clinic, *Mind the Gap: The Lack of Accountability for Killer Robots* (2015), at 13.

[52]See ARSIWA commentaries, *supra* note 4, at 40 (Commentary to Art. 4, para. 5); J. Crawford, *State Responsibility: The General Part* (2013), at 117.

[53]M. Bo, 'Autonomous Weapons and the Responsibility Gap in light of the *Mens Rea* of the War Crime of Attacking Civilians in the ICC Statute', (2021) 19(2) *Journal of International Criminal Justice* 275.

[54]See ARSIWA commentaries, *supra* note 4, at 34 (Commentary to Art. 2, para. 3); Crawford, *supra* note 52, at 60–1. Certain primary norms specify a requirement of intent of the organ or agent through which the State is acting (e.g., Genocide Convention, Art. II).

[55]Apart from being defined as 'an act or omission or a series of acts or omissions' (see ARSIWA commentaries, *supra* note 4, at 38 (Commentary to Ch. II of Part I, para. 1)).

[56]Friedman and Kahn, *supra* note 1, at 7; Matthias, *supra* note 1, at 175; Johnson, *supra* note 21, at 195.

[57]H. M. Roff and R. Moyes, 'Meaningful Human Control, Artificial Intelligence and Autonomous Weapons', (2016) *Briefing paper for delegates at the Convention on Certain Conventional Weapons (CCW) Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS)*; R. Crootof, 'A Meaningful Floor for "Meaningful Human Control"', (2016) 30 *Temple International and Comparative Law Journa*l 53; UNIDIR, 'The Weaponization of Increasingly Autonomous

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

predictions or recommendations on the basis of inference and patterns in data, at a scale and speed that the human brain cannot comprehend[58] – pose significant challenges. Direct operators can have limited control over the technology they use, while developers or decision-makers can be seen as having a too far-removed connection with possible harm occurring during deployment.[59] Besides, it is increasingly recognized that humans and technology interact as part of complex socio-technical entanglements, where the distinction between human conduct and machine conduct is not clear cut.[60] Specifically, the autonomous capabilities of AI systems can affect human behaviours and reshape human agency in relation to technological objects.[61] As a result, there is a strong argument to be made that certain conduct resulting from the use of AI systems do not qualify as action or omission of a human operator. For instance, when an AI system operates semi-autonomously under the formal supervisory control of a human operator, without the operator having any meaningful capacity to influence the outcome, it could be argued that there is no human conduct on the part of the operator. The more machines operate irrespective of the actions or omissions of direct human operators, the more it is difficult to convincingly argue that the conduct of human operators provides ground for attribution.

While the conduct of direct human operators can become less relevant in the context of AI, wrongful conduct involving AI can still and always be traced back to human choices and human actions or omissions.[62] In particular, human conduct in the form of decision-making at the stages of technology development, procurement, or strategic and operational planning can be particularly relevant for attribution of conduct involving AI.[63] For instance, the decision to adopt or deploy a given AI system without verifying its reliability can directly contribute to the occurrence of IHL violations on the battlefield.

3.3 A framework for attribution of conduct involving artificial intelligence

Having demonstrated that attribution of conduct hinges upon the existence of human actions or omissions which result in a violation of international law, and that the way human beings interact with AI technologies can reconfigure what counts as 'human conduct', the following section discusses whether and on which basis wrongful conduct involving AI can be attributed to a state. In essence, it is argued that attribution can be grounded in human conduct which contributed to cause a violation of international law, and that, increasingly, the actions or omissions of commanders, deciders, and developers, have more influence on the occurrence of violations than

Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward', (2014) *UNIDIR Resources No 2*.

[58]See Section 2, *supra*.

[59]P. M. Asaro, 'The Liability Problem for Autonomous Artificial Agents', (2016) *Association for the Advancement of Artificial Intelligence* 190; A. Matthias, *supra* note 1, at 177; H. Y. Liu, 'Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems', in Bhuta et al. (eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (2016), 325, at 326; Commissie Actualisering Autonome Wapensystemen (CAAW), 'Autonome Wapensystemen; Het Belang van Reguleren en Investeren', *AIV-advies 119, CAVV-advies 38*, 3 December 2021, at 36.

[60]L. Suchman and J. Weber, 'Human–Machine Autonomies', in Bhuta et al., ibid., 75 at 98–102; P. P. Verbeek, 'Toward a Theory of Technological Mediation: A Program for Postphenomenological Research', in J. K. B. O. Friis and R. P. Crease (eds.), *Technoscience and Postphenomenology* (2015), 189; L. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions* (2006).

[61]E. Schwarz, 'Autonomous Weapons Systems: Artificial Intelligence, and the Problem of Meaningful Human Control', (2021) V *Philosophical Journal of Conflict and Violence* 53, 55–7; M. Arvidsson, 'Targeting, Gender, and International Posthumanitarian Law and Practice: Framing The Question of the Human in International Humanitarian Law', (2018) 44(1) *Australian Feminist Law Journal* 9, at 23–4.

[62]Bryson, *supra* note 21, at 21; D. J. Gunkel, 'Mind the Gap: Responsible Robotics and the Problem of Responsibility', (2020) 22 *Ethics and Information Technology* 307, at 309.

[63]Geiß, *supra* note 51, at 448.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

the conduct of the operator of an AI system. Four main scenarios can be identified to draw the contours of a framework for attribution of conduct involving AI.

First, some AI systems operate under the direct and genuine control of an operator at the tactical level. For instance, if an object recognition system is used as a vision aid in a context where the operator also has direct visual perception,[64] the actions and omissions of the operator directly contribute to the outcome, and conduct in violation of international law can be attributed to the state on behalf of which the operator acts.[65]

Second, certain AI systems can operate fully autonomously once they are activated. For example, air defence systems operating in automatic mode can, within limited parameters, identify and automatically fire at incoming threats.[66] In this scenario, human conduct that most clearly contributes to subsequent harm lies on the side of decision-makers who decided to deploy and activate the AI system.[67] Indeed, even if the system operator has the possibility to abort an attack, they have limited time to intervene and very limited situational awareness. As a result, the influence that human conduct in the form of override functions has over the outcome of becomes meaningless.[68] By contrast, the decision to deploy a system operating autonomously once activated, and the act of defining and circumscribing the parameters within which the system can launch attacks, can be causally linked to the outcome. Therefore, wrongful conduct occurring in the use of almost fully autonomous systems can be attributed to the state on the basis of the acts and omissions of decision-makers at the at the military or political levels.[69]

The third scenario involves AI systems that operate in a grey area, under some degree of human control and supervision. Typically, such a system is formally under the direct control of its operator, who retains the decision-making capacity to follow or reject AI-generated recommendations. For instance, AI systems that are used in support of target acquisition can gather and analyse data from various sensors and sources, and suggest potential targets, while the operator ultimately remains in charge of the decision to launch or not an attack.[70] In such systems, a certain level of discretion is vested in the operator, who constantly assesses AI recommendations against their own judgement and degree of situational awareness. However, in practice, there is a very fine line between algorithmic decision-support and algorithmic decision-making. As discussed above,[71] AI systems function at a speed and scale that makes it difficult, if not impossible, for human operators to genuinely assess whether a given targeting recommendations should be followed. As a result, control of the operator over outcomes in the use of semi-autonomous systems may become superficial, and the actions and omissions of the operator may not provide sufficient ground for attribution. In that case, a stronger argument is to rely on the conduct of state organs who decided to adopt and deploy the system. Indeed, decision-makers are in a position to assess and enquire into capabilities, limitations, and risks of a system, and to exercise informed judgement over whether, in which operational circumstances, and under which degree of human control, the system should be deployed. Depending on the extent to which the operator exercised control over the outcomes, wrongful conduct involving human-supervised semi-autonomous systems can arguably be attributed pursuant to Article 4 ARSIWA on the basis of either the decision of state organs within the

[64]K. Osborn, 'How the Army Intends to Fight Using Augmented Reality Goggles', *The National Interest*, 6 January 2021, available at www.nationalinterest.org/blog/reboot/how-army-intends-fight-using-augmented-reality-goggles-175745.

[65]Pursuant to Art. 4 ARSIWA, as members of a state's armed forces qualify as state organs.

[66]I. Bode and T. Watts, 'Meaning-less Human Control: Lessons from Air Defence Systems for Lethal Autonomous Weapons', (2021) *Report, Centre for War Studies (University of Southern Denmark) and Drone Wars UK*, at 27.

[67]See also Geiß, *supra* note 51, at 448.

[68]Bode and Watts, *supra* note 66, at 28.

[69]Pursuant to Art. 4 ARSIWA, as members of a state's armed forces or governmental apparatus qualify as state organs.

[70]V. Boulanin and M. Verbruggen, 'Mapping the Development of Autonomy in Weapon Systems', (2017) *SIPRI Report*, at 24–6.

[71]See Section 2, *supra*.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

military chain-of-command to make use of an AI system, or the actions and omissions of the systems operator.

Fourth, it cannot be excluded that future AI systems exhibiting higher degrees of autonomy could be developed, and that such systems be conceptualized as independent and endowed with a degree of autonomous agency. This article does not address the controversial and speculative question of whether such highly autonomous systems should be developed and used in the military context or beyond, nor whether it is ethically appropriate to conceptualize advanced AI as autonomous agents,[72] but seeks to prospectively explore the possible implications of highly autonomous AI systems for attribution of conduct. Should future advanced AI be developed and conceptualized as autonomous, independent, perhaps sentient entities, which could evolve and behave beyond human control, the link to any human conduct could be too vague and weak to ground attribution of conduct.[73] However, it could still be argued that conduct involving such AI could be attributed directly to the state, without the intermediation of human conduct. In this construction, advanced AI could be conceptualized as itself being an agent of the state, 'acting on the instructions of, or under the direction or control of, that State',[74] with wrongful conduct attributed pursuant to Article 8 ARSIWA. Article 7 ARSIWA, which provides that the conduct of an organ or agent can be attributed 'even if it exceeds its authority or contravenes instructions', could also be relevant if AI would be conceptualized as an autonomous agent. While relying on Article 8 ARSIWA allows to solve the problem of attribution for any conduct emanating from AI systems considered to be acting on behalf of the state, it is problematic as it implies that AI systems can and should be conceptualized as independent agents endowed with a degree of subjectivity. This indirectly supports the argument that AI systems could have moral agency or legal personality, which is highly debatable.[75] In the context of broader public debates on retaining human control and human agency over AI technologies,[76] the argument should rather be to make sure to preserve the role of humans, also as part of the framework of state responsibility.

In conclusion, concepts of state responsibility and rules of attribution of conduct are amenable to the challenges of allocation of responsibility in relation to military AI. There is no 'responsibility gap'[77] as far as state responsibility is concerned, and wrongful conduct occurring in the use of military AI systems can be attributed to the state. Nonetheless, the characteristics of AI reconfigure the operation of attribution of conduct, as relevant human conduct may be relocated from operators to decision-makers.

4. State responsibility at the stage of development or acquisition of military artificial intelligence: Compliance-by-design

In addition to state responsibility for the actual use of military AI that results in violations of international law, responsibility can also be analysed at the earlier stages of the design, development, or acquisition of military AI. Prior to deployment and actual harm, states can incur responsibility if they develop or acquire AI technologies in violation of their international obligations.

[72]Whether our society can and should seek to develop AI systems able to act and develop in full autonomy is a critical debate. Russel warned that overly intelligent AI would pose existential threats to humanity (S. Russell, *Human Compatible: AI and the Problem of Control* (2020)). Bryson strongly advocated against assigning moral agency to technical artefacts (Bryson, *supra* note 21, at 15).

[73]See also Castel and Castel, *supra* note 5, at 9.

[74]See ARSIWA, *supra* note 4, at Art. 8.

[75]J. Bryson, M. E. Diamantis and T. D. Grant, 'Of, for, and by the People: The Legal Lacuna of Synthetic Persons', (2017) 25 *Artificial Intelligence and Law* 273; Bryson, *supra* note 21, at 15.

[76]EU High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy Artificial Intelligence* (2019), at 14; Russell, *supra* note 72.

[77]Matthias, *supra* note 1, at 175; Bo, *supra* note 53, at 275.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

Pursuant to the second constitutive element of an internationally wrongful act – breach of an international obligation – state responsibility can only arise in case of a violation of an international obligation binding on the state, either stemming from treaty obligations or based on customary international law. While the breach of an applicable international obligation is always required, actual damage is not necessary.[78] Responsibility can thus arise even in the absence of any injury suffered. By contrast, damage that is caused by the state without involving a violation of international law does not engage international responsibility. The framework of state responsibility is therefore geared towards a return to legality and the preservation of the international legal order.

The constitutive nature of the element of breach thus means that responsibility can arise prior to the harmful use of military AI. Indeed, as international obligations apply during the development of military AI, responsibility can be engaged if such development involves a breach of international law. In this regard, a number of positive obligations are particularly relevant. In contrast to negative obligations not to engage in certain conduct (e.g., the obligation not to target civilians) that are more relevant at the stage of deployment, positive obligations prescribe a duty to actively take steps to secure certain rights and ensure compliance with the law.

The act of developing or purchasing AI technologies can itself qualify as an act of the state, and attribution of conduct at the stage of development or acquisition would not be subject to significant hurdles. When state organs or state-controlled entities engage in AI research and develop military AI technologies, this conduct is attributable to the state pursuant to Articles 4, 5, or 8 ARSIWA.[79] At the stage of development, state responsibility for an internationally wrongful act can therefore arise as soon as there is a breach of an applicable international obligation. The same applies to the acquisition from third parties of military technologies by state organs and entities, which also constitutes conduct attributable to the state.

Existing international legal standards apply to all activities of the state, and the novelty or complexity of military AI or other emerging technologies does not displace the applicability of international norms to state conduct involving AI. The ICJ clearly affirmed that 'the established principles and rules of humanitarian law ... applies to all forms of warfare and to all kinds of weapons, those of the past, those of the present and those of the future'.[80] The particular characteristics of AI technologies such as autonomy and opacity might require efforts to interpret how existing rules operate and can be implemented in the sphere of new technologies, but it is broadly agreed that the law remains applicable,[81] and its violations will give rise to state responsibility. In that sense, the idea that AI technologies are unregulated and that law is not able to catch up with the fast development of new technologies is misleading. Rather than seeking to adopt new legal frameworks to address technological advancements, it should first be explored how existing norms can be applied to emerging technologies.[82]

Responsibility can therefore arise in relation to the way a state undertakes the development of military AI. Obligations of diligence applicable at the stage of development of military AI include

[78]See ARSIWA commentaries, *supra* note 4, at 36 (Commentary to Art. 2, para. 9).

[79]See also, arguing that private actors who supply AI systems for government decision-making should qualify as state actors under the domestic 'state action doctrine': K. Crawford and J. Schultz, 'AI Systems as State Actors', (2019) 119 *Columbia Law Review* 32.

[80]*Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion of 8 July 1996, [1996] ICJ Rep. 226, at 259, para. 86.

[81]Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System, UN Doc. CCW/MSP/2019 (2019), at Annex III, Principle (a); Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Report of the 2014 Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), UN Doc. CCW/MSP/2014/3 (2014), paras. 27–28.

[82]Similarly, see the *Tallinn Manual on the International Law Applicable to Cyber Warfare* (2013), on how existing international norms apply in the cyberspace.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

the duty to respect and ensure respect for IHL,[83] and positive obligations to take active steps to secure human rights within a state's jurisdiction.[84] The internal dimension of Common Article 1 notably refers to the training of armed forces to ensure they know and abide by IHL,[85] and involves a duty 'to take appropriate measures to prevent violations from happening in the first place'.[86] Applied to the development of military AI technologies, it implies a duty to ensure that AI can comply with IHL, to design and train algorithms in line with IHL standards, and to refrain from developing and adopting technology when it cannot be IHL-compliant.[87] In the military context, Article 36 of the Additional Protocol I to the Geneva Conventions furthermore provides for a specific obligation to determine, when considering the development or acquisition of a new weapons, means or method of warfare, whether its employment would, in some or all circumstances, be prohibited by any applicable rule of international law.[88] In the context of military AI, the scope of Article 36 arguably encompasses applications of AI that are not directly falling under the category of 'weapons', as such military AI systems can qualify as 'means or method of warfare'.[89] Equally relevant to the development phase of AI are obligations to protect and ensure respect for human rights,[90] which fully apply in the pre-deployment phase, and can address dual-use AI technologies that are not initially developed for a military purpose.

The applicability of existing obligations at the stage of development of AI technologies therefore means that AI must be designed in full compliance with applicable international norms, and that failure to do so can engage the responsibility of the state. In other words, this article argues that states have the duty to integrate international obligations from the outset and throughout the process of developing, training, and testing military AI. In line with the responsible innovation and ethics-by-design approaches, which seek to identify and integrate ethical values in the design of technology,[91] but applied – beyond ethics – to binding legal standards,[92] design choices must reflect and incorporate the state's international obligations. Upholding state responsibility at the stage of AI development, prior to potentially harmful use, leads to a compliance-by-design

[83]Common Art. 1 to the Geneva Conventions. See A. Berkes, 'The Standard of "Due Diligence" as a Result of Interchange between the Law of Armed Conflict and General International Law', (2018) 23 *Journal of Conflict and Security Law* 433.

[84]E.g., Art. 2 of the International Covenant on Civil and Political Rights (ICCPR); Art.1 of the European Convention on Human Rights (ECHR). See D. Shelton and A. Gould, 'Positive and Negative Obligations', in D. Shelton (ed.) *The Oxford Handbook of International Human Rights Law* (2013), 562, at 564–70.

[85]ICRC, *Commentary on the First Geneva Convention* (2016), para. 146.

[86]Ibid., para. 145.

[87]See also H. Nasu, 'Artificial Intelligence and the Obligation to Respect and to Ensure Respect for International Humanitarian Law', in E. Massingham and A. McConnachie (eds.), *Ensuring Respect for International Humanitarian Law* (2020), 132; E. Massingham, 'Weapons and the Obligation to Ensure Respect for IHL', in Massingham and McConnachie, ibid., at 115.

[88]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, Art. 36. See T. Vestner and A. Rossi, 'Legal Reviews of War Algorithms', (2021) 97 *International Law Studies* 509.

[89]K. Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare', (2022) 23 *Yearbook of International Humanitarian Law* 123.

[90]L. McGregor, D. Murray and V. Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability', (2019) 68 *International & Comparative Law Quarterly* 309, at 327–8.

[91]J. van den Hoven, in Owen, Bessant and Heintz, *supra* note 22, at 75; Floridi et al., 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations', (2018) 28 *Minds and Machines* 689; Institute of Electrical and Electronics Engineers (IEEE), *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, 2018, available at www.ethicsinaction.ieee.org; EU High Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI* (2019).

[92]P. Nemitz, 'Constitutional Democracy and Technology in the Age of Artificial Intelligence', (2018) 376(2133) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*; E. Aizenberg and J. van den Hoven, 'Designing for Human Rights in AI', (2020) 7(2) *Big Data & Society*, available at doi.org/10.1177/2053951720949566.

approach which fulfils a preventive aim.[93] In this perspective, state responsibility proves to be a particularly useful doctrine towards ensuring that military AI is developed and adopted in full compliance with the applicable international norms.

Integrating international law norms in the design of AI systems is not without difficulties, as certain principles such as proportionality and distinction may not be reducible to code.[94] Nonetheless, seeking to ensure compliance-by-design remains the applicable benchmark. The obligation to ensure compliance with international law by embedding norms in AI design can serve to identify the boundaries of whether and how technologies should be developed. Indeed, if it is found that a system cannot technically be made to comply with certain principles, then its development should be either halted or reframed in order to ensure sufficient human involvement to achieve compliance with these principles. Continuing to develop a certain technological system when it has been shown that it cannot comply with certain norms would result in a failure of diligence engaging state responsibility.[95]

At the stage of procurement from third parties, there is similarly a duty to verify that AI technology has been designed and developed in line with the obligations of the state. Again, the complexity of technologies or issues of secrecy do not displace international obligations to ensure that new technologies can and will comply with the law. Wilful blindness is not an excuse for non-compliance, and the state has a duty to diligently seek information from the private or public actors from which it acquires technologies.[96] This also means that the state has an obligation to test systems acquired from third parties for compliance with IHL and other obligations.

In practice, the diligent development or acquisition and subsequent deployment of military AI in line with international obligations would involve processes of risk assessment, testing, auditing, certification, and continued compliance-monitoring mechanisms.[97] In order to operationalize frameworks of international responsibility in relation to military AI, further interdisciplinary research is needed to develop policy guidance and technical protocols for testing and certifying the compliance of AI systems with international law.

In view of the inherent unpredictability of certain AI technologies, as well as the risks of malfunction, a question that can be raised is whether a negligence model of responsibility is sufficient in the context of military AI, or whether the development and use of military AI should be subject to strict liability.[98] Under a strict liability model, actors can be held responsible even if some diligence was exercised, on the basis of the inherent high risk of certain otherwise lawful but hazardous activities.[99] The opportunity for a regime of strict liability for military AI should be discussed and debated amongst states and other parties, for instance in the context of the

[93]See also N. Stürchler and M. Siegrist, 'A "Compliance-Based" Approach to Autonomous Weapon Systems', *EJIL:Talk!*, 1 December 2017, available at www.ejiltalk.org/a-compliance-based-approach-to-autonomous-weapon-systems; 'Towards a "Compliance-Based" Approach to LAWS', (2016) *Informal Working Paper submitted by Switzerland, Informal Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), Geneva*.

[94]M. Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified', (2014) 90 *International Law Studies* 308, at 313; A. Deeks, 'Coding the Law of Armed Conflict: First Steps', *SSRN*, 28 May 2020, available at www.ssrn.com/abstract=3612329.

[95]The legal consequences of state responsibility entail an obligation of cessation and non-repetition (see ARSIWA, *supra* note 4, Art. 30).

[96]NATO JAPCC, *Future Unmanned System Technologies: Legal and Ethical Implications of Increasing Automation* (2016), at 19; ICRC, Commentary on the Additional Protocols (1987), at 426.

[97]See NATO JAPCC, ibid., at 30.

[98]N. Bhuta and S. E. Pantazopoulos, 'Autonomy and Uncertainty: Increasingly Autonomous Weapons Systems and the International Legal Regulation of Risk', in Bhuta et al., *supra* note 59, at 284; Crootof, *supra* note 5, at 1394; Geiß, *supra* note 51, at 448–9.

[99]Such regime has been developed in the context of environmental damage resulting from hazardous activity: Draft Articles on Prevention of Transboundary Harm from Hazardous Activities, with Commentaries, 2001 YILC, Vol. 2 (Part Two), at 146; Draft Principles on the Allocation of Loss in the Case of Transboundary Harm Arising out of Hazardous Activities, with Commentaries, 2006 YILC, Vol. 2 (Part Two), at 106.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

annual meetings of the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW), and would require the adoption of new rules.

5. State responsibility in relation to the conduct of other states and private actors

The third dimension of state responsibility analysed in this article concerns responsibility arising in relation to the conduct of other actors. Next to responsibility for a state's own conduct – either at the stage of deployment or development – explored in the previous two sections, state responsibility can further arise in relation to the conduct of other actors, namely other states or private actors. In such situations of derived responsibility, a state is not directly responsible for the conduct of others, but it can bear responsibility for its failure to abide by its own obligations that relate to the conduct of other actors.

First, with regards to responsibility in relation to the conduct of other states, there exist general and specific obligations not to blindly facilitate or knowingly foster violations of international law by other states.[100] In the ARSIWA, Article 16 provides for an obligation not to knowingly aid or assist another state in the commission of a wrongful act.[101] Similarly, the external dimension of the duty to ensure respect for IHL embedded in Common Article 1 to the Geneva Conventions arguably includes an obligation not to aid or assist other states in violations of IHL, as well as a positive obligation to seek to ensure that other parties comply with IHL.[102]

These obligations are particularly relevant in the context of multinational operations where several states engage together in military operations in which AI technologies might be used by some of the coalition partners. For instance, if a state provides to another state AI-based targeting acquisition support in the form of potential targets that need to be further verified and approved, and the other state engages in targeting on this basis without ensuring adequate human approval, the assisting state can bear derived responsibility in relation to wrongful targeting. Given that Article 16 ARSIWA and Common Article 1 are subject to a criterion of knowledge, respectively actual and constructive,[103] the responsibility of the state providing targeting support arises if it becomes aware of the wrongful conduct of the supported state, for instance due to recurrent targeting mistakes. Another illustrative example would be a situation where one state repeatedly deploys an AI system that results in biased outcomes, with other coalition states blindly allowing this conduct to continue despite becoming aware of the malperformance.

The prohibition of aid or assistance and the duty to ensure respect for IHL are also relevant in the context of the export of AI technologies by one state to another. A state not engaged in armed conflict could engage its derived responsibility if it knowingly transfers AI technologies, such as weapons, target acquisition software, or surveillance tools, to another state which uses them for international law violations.[104] In this respect, military AI technologies that can be used in weapons systems are also arguably subject to the specific rules of the Arms Trade Treaty (ATT), which apply to weapons as well as their parts and components.[105] Article 6 ATT provides for a negative obligation not to authorize any transfer of weapons, parts, or components if the state has knowledge that the items would be used in the commission of war crimes. For exports not prohibited as such under Article 6, Article 7 ATT imposes a positive obligation to assess the

[100]For a detailed analysis see B. Boutin, 'Responsibility in Connection with the Conduct of Military Partners', (2018) 56 *Military Law and the Law of War Review* 57.

[101]On the primary dimension of Art. 16 ARSIWA see Crawford, *supra* note 52, at 339.

[102]ICRC, Commentary on the First Geneva Convention (2016), paras. 158, 164; J. M. Henckaerts and L. Doswald-Beck (eds.), *Customary International Humanitarian Law, Volume 1: Rules* (2005), at 509, Rule 144; H. P. Aust, *Complicity and the Law of State Responsibility* (2011), at 388.

[103]ICRC, Commentary on the First Geneva Convention (2016), para. 160.

[104]Ibid., para. 167.

[105]2013 Arms Trade Treaty, 3013 UNTS, Art. 4.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

potential that the transferred items could be used to commit or facilitate serious violations of IHL or human rights.

Second, with regard to the conduct of private actors, a state can bear indirect responsibility if it fails to ensure that private actors within its jurisdiction operate with respect for international law. Under general international law, every state has a negative 'obligation not to allow knowingly its territory to be used for acts contrary to the rights of other States'.[106] In the field of human rights, the International Covenant on Civil and Political Rights (ICCPR) explicitly provides a positive obligation to 'take the necessary steps' towards ensuring respect for human rights.[107] Again, the state is not directly responsible for the conduct of private actors, but 'may be responsible for the effects of the conduct of private parties, if it failed to take necessary measures to prevent those effects'.[108] The obligation to promote and protect human rights is one of due diligence, which 'requires States to take measures designed to ensure that individuals within their jurisdiction are not subjected to' human rights violations, and 'to take reasonable steps to avoid a risk of ill-treatment' by third parties.[109]

The obligations of states to ensure respect for human rights by private actors is particularly relevant in the context of military AI, as one cannot disregard the major role of private companies in developing and selling AI susceptible to leading to international law violations.[110] It is also important to take into account that many AI technologies with potential military applications are of a dual-use nature,[111] so that efforts to protect human rights with regard to military AI must also address companies not directly involved in defence and security applications.[112]

One of the main tools for states to ensure that private actors respect human rights is through domestic regulation.[113] When it comes to new technologies such as AI, states must therefore put in place new regulations, if and when necessary, to ensure that technologies developed by private actors do not lead to human rights infringements. A recent report on 'Responsibility and AI' from the Council of Europe made clear that 'states are obliged under the ECHR to introduce national legislation and other policies necessary to ensure that ECHR rights are duly respected, including protection against interference by *others* (including tech firms)', and stated that:

> [the ECHR framework] offers solid foundations for imposing legally enforceable and effective mechanisms to ensure accountability for human rights violations, well beyond those that the contemporary rhetoric of "AI ethics" in the form of voluntary self-regulation by the tech industry can realistically be expected to deliver.[114]

[106] *Corfu Channel Case (United Kingdom v. Albania)*, Merits, Judgment of 9 April 1949, [1949] ICJ Rep. 4, at 22.

[107] 1966 International Covenant on Civil and Political Rights, 999 UNTS 171, Art. 2.

[108] See ARSIWA commentaries, *supra* note 4, at 39 (Commentary to Chapter II of Part I, para. 4). See also UN Committee on Economic, Social and Cultural Rights (CESCR), General comment No. 24 (2017) on State obligations under the International Covenant on Economic, Social and Cultural Rights in the Context of Business Activities, UN Doc. E/C.12/GC/24 (2017), para. 32.

[109] *El-Masri v. the Former Yugoslav Republic of Macedonia*, Judgment of 13 December 2012, App No. 39630/09, [2012] Reports of Judgments and Decisions, para. 198.

[110] S. De Spiegeleire, M. Maas and T. Sweijs, 'Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers', (2017) *HCSS Report*, at 47–50.

[111] Persi Paoli et al., *supra* note 28, at 6.

[112] See, for instance, the concerns raised by the use of the controversial and commercially-developed Clearview's facial recognition system in the Ukraine war: D. Meacham and M. Gak, 'Ukraine Military Gets Face Recognition AI. We're Worried', *Open Democracy*, 30 March 2022, available at www.opendemocracy.net/en/technology-and-democracy/facial-recognition-ukraine-clearview-military-ai.

[113] United Nations Guiding Principles on Business and Human Rights (UNGP), HR/PUB/11/04 (2011), at 4–6.

[114] K. Yeung, 'Responsibility and AI: A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework', (2019) *Council of Europe Study DGI (2019)05*, at 67.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

In the same vein, the 2018 Toronto Declaration, whose signatories include Amnesty International and Human Rights Watch, provides that: 'States should put in place regulation compliant with human rights law for oversight of the use of machine learning by the private sector in contexts that present risk of discriminatory or other rights-harming outcomes.'[115]

In practice, states must develop clear legal standards for the private sector, which translate general human rights duties to the context of AI, and apply at the stage of design and development.[116] In order to operationalize AI regulation in relation to human rights and international law violations, technical standards and processes will likely be needed. For instance, testing and monitoring schemes could allow the screening and certification of AI systems developed by private actors.

In view of the transnational nature of many companies involved in developing military AI, there is a risk of avoidance and buck-passing between states claiming to not be able to regulate such companies. However, regulatory obligations stemming from human rights law are limited to actors within the jurisdiction of the state (for instance, based in the territory of that state), and subject to a standard of reasonable diligence. The transnational nature of technology companies is then not an excuse to avoid regulating their conduct. In order to more effectively implement regulatory obligations in the technology sector, co-ordinated efforts of states at the EU or UN level might nonetheless be useful.

6. Concluding remarks

This contribution has provided a comprehensive overview of the situations in which state responsibility in relation to military AI can arise. State responsibility can be engaged for the wrongful use of AI-enabled technologies in the battlefield, negligent development or procurement of AI technologies (including in case of failure to integrate international norms in AI design and development), and failure to ensure respect for international law by other actors developing or deploying AI. At the stage of deployment, the article analysed the human dimension of attribution of conduct, arguing that, due to the characteristics of AI, the actions and omissions of direct human operators do not always provide ground for attribution. Nonetheless, attribution of conduct involving AI systems can be grounded in human conduct and human decision-making by other organs and agents (i.e., developers, political and military decision-makers). At the stage of development, it argued that existing obligations prescribe a duty to ensure compliance-by-design, that is, an obligation to seek to integrate applicable norms in the design of AI systems. This obligation also comes into play at the stage of procurement, where states must verify compliance. Regarding derived responsibility, the article discussed some implications of using AI for responsibility in multinational military operations, and analysed state obligations to ensure respect for international law in terms of the regulation of private actors.

The article demonstrated that, overall, the framework of state responsibility appears amenable to the specific challenges posed by AI technologies. Further, it argued that state responsibility has a useful role to play for AI regulation and accountability. In order to bridge potential responsibility gaps in relation to military AI, state responsibility under international law has a complementary function next to other responsibility frameworks which, together, have the potential to comprehensively ensure accountability at all levels. Seeking to hold states accountable in relation to AI further presents unique advantages in view of the primary role of states with regard to the

[115]Toronto Declaration on Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems (2018), para. 40.

[116]Comparably, the US issued a set of guidelines for private contractors, which provides detailed guidance on how to ensure that ethical principles are integrated and implemented in the development of AI systems. US Defense Innovation Unit (DIU), 'Responsible AI Guidelines in Practice', available at www.diu.mil/responsible-ai-guidelines; W. D. Heaven, 'The Department of Defense is Issuing AI Ethics Guidelines for Tech Contractors', (2021) *MIT Technology Review*, available at www.technologyreview.com/2021/11/16/1040190/department-of-defense-government-ai-ethics-military-project-maven/.

https://doi.org/10.1017/S0922156522000607 Published online by Cambridge University Press

deployment of military technologies, and the regulation of private actors. In particular, state responsibility proves to be a useful doctrine in order to ensure that military AI is developed and adopted in full compliance of applicable international norms. By building on approaches found in the field of ethics of technology, such as value-sensitive design and human-cantered innovation, the article also contributed to bridging ethical, technical, and legal approaches to AI.

In order to operationalize the framework of responsibility outlined in this article, further interdisciplinary research is however needed on engineering methods that would strengthen the capacity of states and corporations to ensure international legal compliance in the design and deployment of AI systems, and on governance approaches and policy options in relation to military AI. This will allow for the translation of legal principles into military and policy guidance as well as technical standards.

Cite this article: Boutin B (2023). State responsibility in relation to military applications of artificial intelligence. *Leiden Journal of International Law* **36**, 133–150. https://doi.org/10.1017/S0922156522000607