



Universiteit
Leiden
The Netherlands

Introduction to Research Handbook on Warfare and Artificial Intelligence

Geiß, R.; Lahmann, H.C.

Citation

Geiß, R., & Lahmann, H. C. (2024). Introduction to Research Handbook on Warfare and Artificial Intelligence. In *Research Handbook on Warfare and Artificial Intelligence* (pp. 1-8). Cheltenham: Edward Elgar.
doi:10.4337/9781800377400.00005

Version: Publisher's Version
License: [Leiden University Non-exclusive license](#)
Downloaded from: <https://hdl.handle.net/1887/4139130>

Note: To cite this publication please use the final published version (if applicable).

Introduction to *Research Handbook on Warfare and Artificial Intelligence*

*Robin Geiß and Henning Lahmann*¹

I.1 THE USE OF AI IN WARFARE: SETTING THE SCENE

The development of artificial intelligence (AI) technologies over the past two decades has already begun to fundamentally change human society in a variety of ways and in different fields. Not least, this holds true for the realm of the military and the conduct of warfare, where the technological progress naturally influences how militaries operate and what future wars will look like.

Unsurprisingly, the prospect of an increased use of AI-enabled applications in the military domain has sparked a fierce academic and political debate. Initially, the discourse was almost exclusively focused on the question of whether machines should be ‘allowed’ to make potentially lethal targeting decisions in weapon systems, a matter that involves a string of intricate and interconnected legal, ethical, technical and political issues. Whereas a number of states, scholars and activists advocate for an outright ban of such a use of autonomous systems, others have been working on proposals for more permissive regulation. The United Nations (UN) Secretary-General Antonio Guterres, in the New Agenda for Peace, has called on states to ‘conclude, by 2026, a legally binding instrument to prohibit lethal autonomous weapons systems that function without human control or oversight, and which cannot be used in compliance with international humanitarian law, and to regulate all other types of autonomous weapons systems’.²

The emergence of lethal autonomous weapons systems (LAWS), however, merely marks the tip of the iceberg when it comes to the potentially transformative impact that these technologies will have on armed conflicts and military operations at large. AI-supported and AI-augmented decision-making in defence and in military operations already exists, and its use is fast increasing across all military domains and theatres of conflict (that is, land, water, air, outer space and cyberspace). Indeed, it seems safe to assume that at some point, to analyse the vast and ever-growing amounts of data, collected by an ever-increasing number of battlefield sensors, and to turn them into actionable intelligence for military operations, will be beyond human capacity and will thus increasingly depend on the help of AI. The technological progress and further proliferation of AI carries promise for strategic and tactical decision-making and greater efficiency in military operations; but it also entails significant risks.

¹ The views expressed in this book are the sole responsibility of the individual authors. They do not necessarily reflect the views or opinions of the United Nations, the United Nations Institute for Disarmament Research (UNIDIR), its staff members or sponsors.

² New Agenda for Peace, July 2023, p. 27, <https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf>.

In many instances, when working as intended, AI enables better-informed and more precise decisions as well as vastly accelerated processing of relevant factors. At the same time, AI developers and programmers make choices and trade-offs that will affect outcomes. Through programming, they embed – intentionally as well as unintentionally – ethical and other preferences within the technology that will be multiplied at the implementation stage. In this context and with regard to military operations in particular, it is essential to ensure that AI-supported military decision-making processes are at all times in conformity with international humanitarian law (IHL) and other applicable rules, and that they are grounded in sound ethical considerations to avoid unlawful and morally invalid results. The overarching challenge in this regard is to ensure that AI remains a technology that serves human actors by extending and enhancing their own ability to decide without hampering or replacing it.

More recently, the debate has started to open up to include a broader range of topics beyond the question of targeting that require thorough and critical examination from various scholarly perspectives. Major points of discussion comprise the transfer of civilian-security applications of AI to situations of armed conflict, such as questions surrounding privacy and data protection.

More overarching problems remain. Will the increased deployment of AI systems by technologically advanced states have a destabilizing effect on international peace and security? What due diligence obligations for states that use AI applications for military purposes can be derived from customary international law? How do we ensure that accountability and legal responsibility can always be established when AI systems violate rules of international law and cause damage?

At the same time, the progress of developing ever more sophisticated autonomous systems for use in the armed forces has moved on, and a new arms race is looming large. While the urge to prevent future disadvantage in relation to other advanced militaries is rational, this must not happen at the expense of international law and the establishment of legal guidelines that bind all involved actors at the international plane. However, and concerning, arms control and related norm development processes at the multilateral level have stalled despite discussions for years on end.

In view of this, it is high time to take stock of the complex and multifaceted discourse in order to highlight possible paths forward. To that end, the *Research Handbook on Warfare and Artificial Intelligence* has brought together scholarship from some of the world's leading experts in the fields of AI and security. The book tackles a string of the most pressing issues in a critical, comprehensive and engaging manner in order to shape future discussions and to achieve the crucial impact on international decision-makers, while at the same time aiming to push the debate forward.

I.2 PRELIMINARY CONSIDERATIONS

So as not to forestall the nuanced deliberations of this book's contributions, the following few general considerations shall suffice as an introductory note.³ Importantly, as an all-purpose

³ These considerations were first laid down in Robin Geiß and Henning Lahmann, 'The Use of AI in Military Contexts: Opportunities and Regulatory Challenges' (2021) *The Military Law and the Law of War Review* 59, 165.

technology, AI is by no means confined to particular military devices such as weapon systems. Quite the contrary: most experts expect that in the coming decades, AI will become the defining tool for military operations on both a tactical and a strategic level. After proposing some useful categorizations to structure the legal, ethical and political discourse surrounding AI in the military, the following sections set out some already existing as well as possible future uses of the technology in and beyond armed conflict situations, leading up to a description of the emergence of an all-encompassing ‘digital battlefield’ that combines the potentials of AI with other transformative technologies, most significantly sensors and robotics.

One way to categorize AI in warfare is by the broader application or method the technology is intended to support: (1) physical military hardware; (2) cyber technologies; and (3) military decision-making in relation to armed conflict.⁴ First, AI can be used to control physical hardware in military settings, such as unmanned robotic systems. Since 2010, we have seen a gradual increase of autonomy in functions that range from targeting systems (including, but not limited to, lethal autonomous weapon systems) to navigation, driving, flight, or surveillance.⁵ Second, a more recent development that has been receiving significant attention of late is the use of ‘intelligent’ software for purposes of cyber warfare, both to strengthen military networks’ defences and for offensive operations.⁶ For example, AI has been found to be particularly expedient in detecting and autonomously exploiting software vulnerabilities, a capability that is useful for both defensive and offensive goals. Furthermore, AI is increasingly considered for bolstering information warfare such as disinformation campaigns. There exist a number of potential applications in this field, the most well-known being the creation of so-called ‘deep fakes’ of audio-visual content.⁷ A more recent development in this context is the expected increase in volume and sophistication of false and misleading textual content online that has been automatically generated by the latest generation of large language models such as GPT-4. And third, more generally, AI has a range of potential uses in military decision-making writ large. Algorithms can parse large quantities of data sources related to armed conflict and analyse them, for example to identify persons or objects, recognize and assess behavioural patterns, or survey historical data in order to come up with recommendations for military strategy or tactical aspects of individual operations. Such large-scale analysis of mostly unstructured data is where the predictive capabilities of strong machine-learning algorithms, through pattern recognition that can relate to language, images or behaviour, arguably come to the fore, effectively in support of or even as a replacement of traditional military intelligence and reconnaissance activities.⁸ In this way, an AI-supported system has

⁴ International Committee of the Red Cross (ICRC), ‘Artificial Intelligence and Machine Learning in Armed Conflict: A Human-Centred Approach’ (2019) 3–4.

⁵ Fraunhofer, ‘Rise of Artificial Intelligence in Military Weapon Systems’ (2020), <https://www.fraunhofer.de/content/dam/zv/de/forschungsthemen/schutz-sicherheit/rise-ofintelligent-systems-in-military-weapon-systems-position-paper-fraunhofer-vvs.pdf>.

⁶ Wilson da Silva, ‘War of the Bots: Artificial Intelligence in Cyber Warfare’ (2019), <https://medium.com/illumination-curated/spy-vs-spy-cyber-warfare-gets-automated-aba60ece738c>.

⁷ Ian Sample, ‘What Are Deepfakes – And How Can You Spot Them?’ (13 January 2020) *The Guardian*, <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>.

⁸ Niklas Masuhr, ‘AI in Military Enabling Applications’ (October 2019) CSS Analyses in Security Policy, <https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse251-EN.pdf>.

been envisioned as potentially providing essential information to human operators and military decision-makers on who or what to attack in a given instance, who to detain and for how long, where to deploy troops, or how to engage the adversary in the short, mid and long term.

Beside an application-focused perspective, another pertinent way to categorize AI in the military is based on the critical aspect regarding the degree of human control over the system. Four classes of control are typically distinguished. While direct control implies that the person operating the system must constantly intervene, on site or remotely, shared control means that some functions of the systems are run autonomously whereas others require a human operator's active participation. An example is an armed drone that is capable of flying and navigating without external control, but that cannot engage a target without the human operator's intervention. Systems under supervisory control are able to perform tasks entirely without human intervention. The operator merely takes on monitoring functions in order to be able to assume control if it becomes necessary or feed the system with further instructions. This variant is often used in scenarios where shared control is not feasible, for example due to communication delays, for instance when the system is deployed under the sea or in space.⁹ Finally, a system can be fully autonomous, which implies that natural persons are completely out of the loop: the AI takes up all controlling functions and commences to fulfil the task without any outside intervention. While this is still rare in robotics, that is, in scenarios where the system directly interacts with the natural environment, algorithms that essentially run a fully autonomous application are increasingly common in purely virtual settings such as online marketplaces.

Finally, AI systems may be categorized according to their criticality in relation to societal values and the rights of human individuals who might be affected by the decisions made by autonomously operating algorithms. Such a framework is especially apt as regards the legal and ethical implications of AI. For instance, an AI system that takes on targeting functions in armed conflict, with potentially lethal consequences for civilians in harm's way, is clearly more critical than an AI system that predicts the maintenance cycles of military equipment, and should therefore receive a great deal of more careful and thorough scrutiny before and during deployment. Systems that make decisions on detaining enemy combatants, create propaganda materials, embark on surveillance missions, or autonomously navigate through conflict environments, arguably fall somewhere in-between these ends of the spectrum.

I.3 OVERVIEW OF CONTRIBUTIONS

The *Research Handbook* is split into four parts, and aims to capture a wide spectrum of views on the role of artificial intelligence in warfare. The views expressed are the views of the individual authors and do not necessarily reflect the views or opinions of the editors. Part I offers an overview of the most important general questions surrounding the use of AI in the armed forces. A historical survey of technological developments for the conduct of warfare is followed by an in-depth examination of some technical and geopolitical issues pertaining to AI. Part II lays out a number of urgent legal and ethical problems of artificial intelligence and warfare, including, but not limited to, the question of implementing meaningful human control of AI-supported military applications during armed conflict. Part III addresses a selection of

⁹ ICRC (n 4) 8–9.

specific applications of AI in warfare, such as cyber warfare, espionage and humanitarian services. Finally, Part IV takes a look ahead, dealing with ongoing efforts to establish disarmament and arms control regimes and to regulate the employment of AI in warfare; a comparative perspective to other fields that have recently been uprooted by the introduction of AI provides further guidance for the issue at hand. Two chapters on overarching matters concerning the progressing ‘de-humanization’ of military conduct and an outlook into remaining open questions concerning the future of warfare conclude the book.

Part I commences with Beatrice Heuser’s survey in Chapter 1 of the evolution of novel military technologies and their impact on strategy and tactics from the Middle Ages onwards. With a view toward increasing autonomy in weapons systems, Heuser asks when and under what conditions technological progress can be considered to have engendered a ‘military revolution’, and what the implications of such designation are. On the basis of her analysis, she expects that the employment of artificial intelligence in the military belongs to this category, deeply transforming many aspects of modern warfare at an increasingly fast pace.

In Chapter 2, Giacomo Persi Paoli and Sarah Grand-Clément explore the disruptive force of AI by examining in detail the whole range of potential applications of the technology, and by providing an overview of their expected benefits and challenges. Further, they compare how the major global powers have approached the development and use of AI. The chapter concludes with a consideration on what AI could mean for our understanding of ‘warfare’.

Chapter 3, by Larry Lewis and Diane Vavrichek, takes a deep dive into the critical functionalities that AI technologies can provide for the conduct of military operations. Explaining the different aspects of the technology, the authors appraise the ways in which it can be useful for armed forces, but also what risks stem from the prospect of proliferation, and how to mitigate those.

Zooming out a little, Elisabeth Hoffberger-Pippan and Anja Dahlmann in Chapter 4 deal with the broader concept of the emerging ‘digital battlefield’, explaining its technological particularities and making a first attempt at appraising its prospects. While pointing out expected advantages for armed forces, the authors also provide a first overview of the salient legal and ethical concerns, in particular in relation to autonomy in targeting decisions.

In Chapter 5, Vincent Boulanin takes up this thread and delves deeper into the practical concerns surrounding the military decision-making process assisted and enabled by AI technologies. After providing a nuanced appraisal of the evolution of such enhanced decision-making, the author maps possible pathways toward realizing responsible AI governance, with an optimistic outlook on the ongoing proceedings within the framework of the Convention on Certain Conventional Weapons (CCW Convention).

Part I concludes with Tobias Vestner’s birds-eye view of the development of military strategy under the shifting paradigms engendered by AI. From the preparation to the conduct of military operations, his Chapter 6 explores the multilayered implications of AI in the context of strategy, doctrine, plans, rules of engagement, and orders. The chapter focuses on broadly considered avenues to advance further reflection, research and prudent policy-making on the appropriate integration, management and use of AI in the military domain.

Part II opens with Chapter 7 and Dominik Steiger’s urgent deliberations concerning a possible future use of AI-enabled systems in high-level decisions to resort to the use of force in the first place. What is the impact of such prospect on the *jus contra bellum* as enshrined in the UN Charter? Could algorithmic systems be capable of taking account of the legal regime’s intricacies, and if so, what conditions must exist for them to do so?

Chapter 8 is the first of two to provide a critical and detailed engagement with the concept of ‘(meaningful) human control’ over AI systems in the military domain. Exploring the nuances of contemporary targeting processes and military command and control, Merel A.C. Ekelhof sets out to reappraise expectations towards this concept in light of the realities of modern warfare. What is the actual place and role of the human within these structures, and where are the actual issues once AI-supported systems come into play?

The second chapter to tackle ‘(meaningful) human control’ approaches the concept through a legal lens. In Chapter 9, Bérénice Boutin and Taylor Woodcock inquire what the law requires the concept to look like, and advance the notion of ‘compliance by design’ as a means to address the complex reality of situations in which military decision-making processes are mediated by AI-enabled technologies. This, the authors propose, constitutes one feasible pathway toward the operationalization of control by and through human operators.

That the principles of machine learning require the collection of vast amounts of personal and non-personal data is a truism, yet the issue has not always been adequately analysed in the military domain. Chapter 10 fills this gap. Asaf Lubin shows one way to apply the rights to privacy and data protection, so crucial when conceiving of AI systems that are safe to use in other contexts, to the acquisition and analysis of big data for military applications.

In Chapter 11, Binxin Zhang deals with the crucial and hotly debated questions surrounding accountability and responsibility for AI-enabled conduct, tackling the entire range of individual criminal responsibility, civil liability and state responsibility. Engaging with frequent claims of an emerging ‘accountability gap’, Zhang holds that the inadequacies of current positive legal regimes do not call for implausible novel interpretations of the law, but for the enactment of new, specific regulations on the development and use of AI-enabled systems.

Antonio Coco and Talita Dias then analyse in Chapter 12 the possible existence of due diligence obligations with regard to the employment of AI technologies in the military domain. Given the many risks inevitably associated with the use of such systems, turning to norms that impose positive duties on states as found in customary international law to avoid or mitigate harm caused by AI-enabled applications is a prudent way to complete the accountability and responsibility framework. The authors hold that states must exercise due care and caution in the design, development and deployment of AI-enabled technologies not only as a matter of ethics and common sense, but also as a matter of international law.

Subsequently, Part III zooms in on some of the more specific use cases of AI-enabled systems in the military domain. To this effect, in Chapter 13, Heather A. Harrison Dinniss delves deep into the topical issue of the utilization of AI-enabled technology – in particular those based on the prevalent principle of machine learning – for conducting cyber operations. She observes that the greater the role of the cyber domain in interstate conflict becomes, the more states will resort to AI both defensively and offensively. After exploring what existing ethical and legal frameworks may guide such state behaviour, the author closes by inquiring whether additional rules are needed.

In Chapter 14, Russell Buchan sheds light on one of the naturally most secretive but also most intriguing use cases of AI-enabled technologies: the use of such systems for the purpose of intelligence gathering and analysis in the context of armed conflict. The chapter explores the role and function of military intelligence in the prosecution of armed conflict, and examines the extent to which AI technology is being used to bolster such efforts, before finally assessing how international humanitarian law applies when parties to armed conflicts use AI-enabled platforms to gather information.

Naturally, AI-enabled systems are useful not only for the conduct of military operations. To broaden the scope of analysis, in Chapter 15 Ana Beduschi tackles the questions raised by the employment of such technology with the aim of improving humanitarian action in times of conflict and crisis. Listing the many benefits AI technologies can bring for humanitarian organizations, the author cautions that their deployment may also expose affected populations to additional harms and present significant risks to the protection of their rights. In response to such challenges, the chapter offers a set of essential guidelines for deploying AI in the humanitarian context.

Finally, Part IV aims to look ahead by first taking stock of some of the long-running efforts to regulate or prohibit AI-enabled technologies in the context of warfare. In Chapter 16, Gugu Dube and Mary Wareham provide a critical overview of state positions on lethal autonomous weapons systems. Drawing on official statements, commentaries, and other sources, the chapter traces the evolution of the debate on LAWS and shows how states have come to converge on the elements seen as necessary for a new multilateral treaty to establish the concept of meaningful human control over any use of force.

Chapter 17 tackles the question of what the future of arms control might look like in the age of AI-enabled warfare. Picking up on the significant risks and challenges in connection with the use of AI as discussed in earlier chapters, Jürgen Altmann discusses the role of arms control regimes for containing escalation risks between states in a world in which such technologies reduce reaction time to mere seconds. On the other hand, the author also assesses whether AI could even be used to improve existing arms control mechanisms.

Of course, the debates surrounding the increasing and future use of AI-enabled technologies in the military domain do not happen in a vacuum, but are embedded in a wider discourse on the adoption of machine learning systems in ever more everyday contexts. Appraising efforts to regulate the technology for the myriad of civilian use cases, in Chapter 18, Silja Vöneky and Thorsten Schmidt inquire what lessons have been learned from such attempts, and how and whether we can derive valuable insights from non-military AI governance that are applicable to the specific realm of the military and armed conflict.

The book closes with an important intervention from Eugenio V. Garcia in Chapter 19. He addresses the major strategic implications for the Global South in relation to the increasing use of AI-enabled technologies in the military. To begin with, it is not difficult to see how the development of warfare toward more and more automation will first of all directly or indirectly affect populations in the Global South, turning them into testing grounds for the use of such systems, while widening already prevalent technological asymmetries. Subsequently, the author assesses possible counterstrategies for these states to explore in order either to catch up, or to prevent unmitigated proliferation by way of international regulation.

I.4 OUTLOOK

Continuous efforts within the UN CCW framework regarding the regulation of LAWS notwithstanding, the obvious fact that the development of AI-enabled technologies in the realm of military affairs concerns a much broader range of equally momentous possible applications has begun to be acknowledged as a critical policy issue that states must tackle proactively. As a first response, in February 2023 the governments of the Netherlands and the Republic of Korea invited state representatives and other stakeholders from industry and civil society to

The Hague for the first global Summit on Responsible Artificial Intelligence in the Military Domain (REAIM 2023).¹⁰ The two-day meeting led to the adoption of a Joint Call to Action, which stresses ‘the paramount importance of the responsible use of AI in the military domain, employed in full accordance with international legal obligations and in a way that does not undermine international security, stability and accountability’.¹¹ The document, which was endorsed by over 60 states and other participants, continues with a list of broad acknowledgments and recommendations that may serve as a platform for subsequent efforts to establish consensus on the permitted and prohibited, ethical and unethical uses of AI-enabled systems for the conduct of military activities. Notably, on the same occasion the United States launched a political declaration on ‘Responsible Military Use of Artificial Intelligence and Autonomy’.¹²

At the same time, as the chapters in this *Research Handbook* make unambiguously clear, a number of states are moving ahead to deploy AI-enabled systems in military operations and on the battlefield. There can be no doubt that it is high time for clarifications that as many actors as possible can agree on and are willing to abide by.

¹⁰ See <https://www.government.nl/ministries/ministry-of-foreign-affairs/activiteiten/ream>.

¹¹ The text of the call is available at <https://www.government.nl/ministries/ministry-of-foreign-affairs/news/2023/02/16/ream-2023-call-to-action>.

¹² See <https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/>.