Introduction

Niklas Schörnig and Thomas Reinhold

Abstract In 1987, Allan Din published the seminal book "Arms and Artificial Intelligence," in which he argued that the future military use of AI would be a double-edged sword. Warning about control failures and accidental war on one hand, Din also pointed out the potential of AI to enhance arms control. 35 years later, what was a niche technology in Din's day has since become one of the most influential technologies in both the civilian and military sectors. In addition, AI has evolved from sophisticated yet deterministic expert systems to machine learning algorithms. Today, AI is about to be introduced in almost every branch of the military, with a variety of implications for arms control. This book reflects the work of the individual authors and identifies common themes and areas where AI can be used for the greater good or where its use calls for particular vigilance. It offers an essential primer for interested readers, while also encouraging experts from the arms control community to dig more deeply into the issues.

1 The Use of AI as a Revolution in Military Affairs

"The envisaged uses of computer and IT techniques in weapon systems give rise to both skepticism and concern, for example because of the risk of control failures leading to crisis and accidental war. There are, however, also possible applications of these techniques within arms control which may have a more positive connotation" (Din, 1987, p. 8). When Allan Din wrote these words in his seminal edited volume "Arms and Artificial Intelligence" in 1987 they almost sounded like science fiction. However, when Din and his co-authors talked about AI they had an understanding of it that is completely different how it is perceived today. In the 1980s, AI often boiled

N. Schörnig (✉)
Peace Research Institute Frankfurt, Germany
e-mail: schoernig@hsfk.de

T. Reinhold
PEASEC, TU Darmstadt, Darmstadt, Germany
e-mail: reinhold@peasec.tu-darmstadt.de

T. Reinhold, N. Schörnig (eds.), *Armament, Arms Control and Artificial Intelligence*, Studies in Peace and Security, https://doi.org/10.1007/978-3-031-11043-6_1

down to so-called expert systems, that is, highly complex, deterministic systems supporting decision-making "based on the heuristic knowledge of domain experts in combination with decision rules" (Orhaug, 1987, p. 167), or problem solving by brute force, a rather simple trial-and-error approach limited by processing power.

Today, things look rather different: Not only has the processing power of CPUs increased by a factor of 1000[1]—following Moore's Law until very recently—new AI techniques such as machine learning have revolutionized the potential of AI applications. When Stanley, the autonomous Volkswagen Touareg, won the DARPA Grand Challenge in 2005 it was the first of five cars to cover the 213 km distance without an accident—in the open desert, without any other traffic. Even 17 years later, fully self-driving cars are still not on the market, but the assistance systems are still aiming at making the driver in the cockpit almost superfluous and already work quite well in well manageable situations.

Some experts claim, and rightly so, that development will continue at a rapid pace, especially as AI is, as Paul Scharre puts it, "not a discrete technology like a fighter jet or locomotive, but rather is a general-purpose enabling technology, like electricity, computers, or the internal combustion engine" (Scharre, 2019). Since civilian advances in AI have a high dual-use character, they have also advanced the military use of AI to a massive degree. Other experts are less optimistic and warn that AI is still best applied to very specific tasks and that the vision of a more universal, flexible and adaptable AI might turn out to be a dead end.

But so far AI seems to provide the technology for enhancing solutions for technical challenges and the latest technical advancements in computer processing power and size described above now allow powerful devices that can handle the processing of AI algorithms, making it increasingly clear that AI will permeate and transform all military domains, from reconnaissance and analysis to key decision-making processes on both the tactical as well as the strategic level, and, finally, the direct execution of a military strike or attack. There are numerous examples, mostly from the US military, that is (still) at the forefront of implementing AI in the military: The US Navy, for example, is "looking to leverage advanced technological capabilities in artificial intelligence (AI) and machine learning (ML) in the tactical and operational realm" (Munoz, 2022, p. 8). In the future, a new Joint-All-Domain Command-and-Control-Concept (JADC2) is expected to combine data from a multitude of sensors and apply AI-enhanced evaluation to the data in order to identify targets and recommend the optimal weapon (White, 2021, p. 21).

While military AI still consists of isolated islands in many places, no-one would doubt that it will have an even stronger impact in the years to come, when the so-called "Internet of Military Things" will "change the landscape of defense operations," as the trade magazine Jane's Defence Weekly predicts (Torruella, 2021, p. 3).

[1] In 1985, Intel's new 80,386 CPU combined 275,000 transistors on one chip, while the current generation of microprocessors squeeze over 3 billion transistors into a very small space. However, the number of transistors is of course not the only factor determining a CPU's performance.

In any case the steady rise in the use of military AI is leading to an ever-increasing acceleration of decision processes and cycles. It is no wonder that many experts, some of whom are also represented in this volume, have argued for some time that the introduction of AI will primarily lead to an acceleration of military actions and responses, shorter reaction times and higher alert levels. This assessment is also shared by the military. The German armed forces, for example, expect a "battle at machine speed" in the future, with decisions to be made in minutes or even seconds, rather than hours (Doll & Schiller, 2019, p. 4).

Thus, while military commanders recognize a tremendous advantage when they—and only they—have a significant advantage in the use of AI, arms controllers and other critics primarily see the dangers of an unhindered and unrestricted military "AI race."

2 The Purpose of the Book

However, such observations are neither new nor innovative, and debating only the military impact of AI on war in general would probably not warrant another book. Our book seeks to go a step further and look at the military use of AI from the perspective of arms control and verification. While we describe the idea behind arms control and verification in more detail later in the book, it is fair to say that all serious arms control needs verification to ensure compliance and be effective. Unfortunately, when it comes to arms control, AI causes potential problems unseen in the older days of physical weapon systems. Not only is arms control hindered by the fact that in contrast to hardware such as tanks, planes, or missiles, which can be physically inspected and counted, software code is notoriously hard to control and verify—if ever. Given the large capacity of modern memory hardware, even extremely complex programs can be stored on fingernail-size memory cards. Software can be updated or replaced in an instant—even if a specific military system passes an inspection, the chance is high that a software update will increase its performance and the dangers it poses tremendously. Consequently, AI will have a very sizable impact on arms control—for better or worse. Unfortunately, many arms control experts who are very familiar with the particular weapon category they work on, still shy away from dealing with AI as they fear that their knowledge in computer science is not sufficient. Our book thus aims at broadening understanding of the relevance of software, AI and ML in the military and arms control realm and seeks to encourage experts to look more deeply into the advantages and disadvantages of AI in their field. The book offers background knowledge about what AI and ML are, how they work, and what they can and cannot achieve, and provides both broader perspectives on the way AI will transform the military as well as insights into key players. It also offers an overview of the relevance of software, AI and ML within several weapon fields in the realm of nuclear, biological and chemical (NBC) weapons, conventional weapons and emerging technologies and examines how the respective fields are dealing with the increasing relevance of new AI-technologies.

For those who are more strongly focused on AI, the book introduces relevant theoretical concepts of arms control and verification, and the way different AI developments will impact arms control. While almost all chapters could easily be twice as long as they are now, all authors were asked to be brief, crisp, and understandable. Consequently, the chapters are not only usable for gaining knowledge but are also very suitable for classroom teaching.

3 The Structure of the Book

After this introduction, the book is divided into three sections. The first section contains theoretical reflections and looks at key actors in the field. The sections starts with a text by Peter Buxmann and Melanie Reuter-Oppermann. They provide an informed introduction to the topic of artificial intelligence and machine learning. Without drifting into formal or mathematical argumentation (which has been placed in the appendix), they first provide a short historical overview of artificial intelligence and machine learning followed by a more concrete introduction to different forms of machine learning algorithms and methods for measuring algorithm quality. This chapter is unrelated to the topics of armament and arms control and can be used as a general introduction to AI. Chapter three, written by Frank Sauer, describes the military rationale for the use of AI. Sauer starts with an understanding of AI encompassing automated tasks "which previously required the application of human intelligence" (p. 27). While this understanding is rather broad, it is also common within military circles. Sauer concludes that in the debate on the military use of AI and ML both are "simultaneously over- and underestimated" (p. 27), blurring clear-cut discourse on opportunities and threats. Frank also sees signs of a dynamic that is detached from actual military needs, arguing that many military officials are employing AI in the armed forces "because everyone else is doing it" (p. 28) and pointing to the pressure many militaries see themselves under. But he also concludes that AI has much to offer the military, at least at first sight and from a strictly military point of view where AI allows faster targeting cycles which eventually lead to superiority over the opponent. He concludes that "the hype is real" but cautions that "so are the risks" (p. 28), and argues that the widespread misunderstanding of AI's strengths and weaknesses is in large part responsible for making the introduction of AI in military applications fraught with risk due the hype surrounding it (p. 35).

Chapter four, written by Sophie-Charlotte Fischer, looks more closely at issues from the end of Sauer's more general chapter and introduces the key players regarded as responsible for the AI arms race. Fischer argues "that important clues to inform the nascent academic and policy discourse on the military and broader security effects of AI can be derived from analyzing and comparing how different countries pursue military AI—in what kind of applications they invest and in which selected areas they already deploy AI" (p. 40). After developing a framework for analysis, Fischer assesses the military capabilities of four countries, the United

States, China, France, and Israel. She concludes that all four countries view synergies between the commercial and military sector as critical to realizing their AI objectives and are in the process of implementing AI "across a wide range of areas including logistics and training, cyber and information operations, Intelligence, Surveillance and Reconnaissance (ISR), and (semi-)autonomous vehicles as well as command and control" (p. 52). However, each country differ significantly in their approach to the risks associated with the application of AI in the military.

In the fifth chapter, Niklas Schörnig looks at the brighter side of the use of AI in a military context and examines how AI can be used to foster arms control. After a general overview presenting the theoretical background of arms control, disarmament and non-proliferation from the specific perspective of verification, Schörnig systematizes the use of AI for arms control in several broader categories, including the use of AI for translation and analysis of text in arms control and verification contexts, the analysis of graphical data, other sensory data, and multimodal data. He concludes that while AI will not replace inspectors in the foreseeable future, it nevertheless offers very helpful support that facilitates the work of inspectors and should be used more in the future.

The second major section of the book, "Empirical Examples from Different Fields of Arms Control," starts with chapter six written by Alex Kelle and Jonathan E. Forman, both of whom have a background as former employees of the Organization for the Prohibition of Chemical Weapons (OPCW). They address the issue of "Verifying the Prohibition of Chemical Weapons in a Digitalized World." In order to understand how new technologies including AI fit into the elaborate verification mechanism of the OPCW, the text first offers a basic understanding of the different verification rules and procedures implemented by the OPCW. They also show that the use of state-of-the-art science and technology for verification purposes flows directly from the Chemical Weapons Convention itself. While AI can enhance verification, the authors also draw attention to the profound changes through which the chemical industry has gone in recent years due the adoption of AI as part of the so-called Industry 4.0. They conclude that this is no time to be afraid of the rapid changes in science and technology, but that scientific literacy is the key to keeping track of both beneficial and malicious use.

In chapter seven Filippa Lentzos looks at AI and biological weapons and highlights key impacts of machine learning and automation on biological research, medicine and healthcare. Lentzos argues that these developments could make the production of biological weapons easier and proliferation more likely. She continues that even though biological weapons are completely prohibited by the Biological Weapons Convention, artificial intelligence and other converging technologies are radically transforming the dual-use nature of biology and present significant challenges for the treaty. She discusses these challenges and presents a vision of how biological arms control can evolve in order to remain relevant in the Fourth Industrial Revolution.

Chapter eight, written by Jana Baldus, is the first of two chapters to look at AI and nuclear weapons. Baldus looks, first, at the connection between AI, nuclear weapons and autonomy and points out that during the Cold War earlier forms of AI were quite common in the nuclear domain. She argues that the use of AI and ML could lead to more reliable early warning and nuclear command systems, generally enhancing nuclear stability. She also points to the downsides, however, including, among others, biased datasets or even increased skepticism toward a high degree of technologization due to the excessive destructiveness of nuclear weapons. She also points out how "AI could help improve the cross-analysis of ISR data, for example to help control treaty declarations" or support the efforts against nuclear proliferation. Like others in this book, Baldus argues that experts in the weapon systems under consideration need to gain an even better understanding of what AI already exists and where and keep track of how these developments will impact nuclear strategy.

The next text, chapter nine, by Anna Heise, delves into an aspect Jana Baldus only touched on: The use of AI in nuclear testing, that is the simulation of nuclear explosions on powerful computer systems. Based on the little that is publicly known about the subject, Heise describes how AI has improved virtual testing and thus avoiding "live" tests with actual nuclear weapons. Heise stresses the human factor and argues that the results of tests "are only as good as the data and models you give them and the knowledge and experience of the person doing the calculations" (???). On this basis she concludes that the future use of AI in testing will "not only be dependent on the technology but on the emotional attitude of those in charge" (???). Heise than looks into the processes of detecting nuclear tests as it has been carried out by the Comprehensive Nuclear-Test-Ban Treaty Organisation (CTBTO) since its foundation in 1996. She explains, for example, how AI can be used to detect tests with seismic wave-form analysis or how AI can help estimate yields of nuclear explosions. Finally, Heise looks at the dangers, emphasized by some observers, such as the analysis of explosions potentially generating proliferation-relevant information on, for example, the design of warheads. She finally concludes that there is already relevant technology for both virtual testing and detection of real nuclear tests, but that these technologies are only being implemented tentatively. Obviously, there is still a lack of trust when it comes to the use of AI in such security-relevant contexts.

With chapter ten, written by Benjamin Schaller, the focus shifts from weapons of mass destruction to conventional aspects of arms control. Based on well-known theories of international relations, Schaller presents the need for conventional arms control and starts with a short overview of European conventional arms control. The European focus may surprise the casual reader, but in fact Europe is the only region in the world where, at least until recently, there was a comprehensive and established arms control architecture in place. Schaller first discusses whether the balance of power will be altered by the use of military AI. He argues that AI will make it even harder to come up with a "balance of power" as quantitative factors become less relevant in contrast to qualitative factors, which are harder to establish. Schaller also argues that at least within the OSCE, the Organization for Security and Cooperation in Europe, AI has played only a minor role, arguing that current differences have caused too many problems for the implementation of AI in fostering arms control to be considered. But he sees chances of fostering conventional arms control, for example by analyzing military information that has been exchanged in the context of confidence and security building measures, but also by enhancing more concrete verification measures. Schaller concludes by emphasizing what other authors have previously stressed: the importance of maintaining the “human factor” in arms control.

Leaving physical weapons altogether, chapter eleven, written by Thomas Reinhold and Christian Reuter, focuses on “cyber weapons and AI.” After an insightful introduction to the militarization of cyberspace, Reuter and Reinhold examine how the development of future cyber weapons will be influenced and driven by AI and ML. The authors argue that cyber and AI/ML are closely related and that all positive effects of AI and ML on developing software when transferred to the cyber sphere as well as current software architecture of course provides an ideal platform for having AI/ML components added to them. They argue that the problems normally associated with AI, namely the loss of human control due to ever-shorter reaction times, are particularly relevant in the cyber domain, “an environment that is marked by extremely low response times.” Reuter and Reinhold also draw attention to the fact that the black-box character of AI and ML systems could lead to new problems regarding attribution of attacks. But they also see a bright side, for example a time when AI-enhanced algorithms will be able to find slightly altered code instead of looking for exact matches or reveal hackers by identifying their particular “digital fingerprint.”

Many of the previous texts have described lethal autonomous weapons as a prime example of the future use of military AI. In chapter twelve, Anja Dahlmann finally looks at the two most prominent “emerging military technologies,” drones and lethal autonomous weapon systems (LAWS). Dahlmann describes remotely piloted military drones as a step toward autonomy. From a military perspective, future drone systems will probably involve more new functions be carried out autonomously, such as air-to-air combat or manned-unmanned teaming. More autonomy will also offset current shortcomings, such as latency problems or broken or jammed communication links. Dahlmann raises the point that all these autonomous functions will most probably be based on AI and ML, drawing a direct line between current drones and future LAWS. Dahlman continues to argue that this development will necessitate a new perspective on arms control, with a focus on the element of human control. In that context, Dahlmann also reminds us that many of the components of LAWS will be dual-use. She concludes that, due to the lack of concrete regulation of LAWS, it is only hypothetical whether AI could have a positive impact on arms control for LAWS—and whether only LAWS should be equipped with “some sort of ethical behavior” (???).

The third and last section of this book focuses on the question of “what should be done.” In chapter thirteen, Maaike Verbruggen focuses on the technical aspects of making ML-based AI reliable. Using the term “verification” in the strict technical sense of software engineering rather than in the sense of arms control, Verbruggen shows the great difficulties when applying time-proven concepts of engineering to software in general and self-learning software in particular. These problems are compounded by the fact that AI is often integrated modularly, leaving open questions of how the AI and the rest of the software interact. She proposes that integration of verification and validation measures should be structurally integrated into the design process of AI-based software from the very start, arguing for a “correct by construction” approach. While on the one hand Verbruggen stresses that these problems are already being examined by defense ministries around the world, she also fears that establishing international validation and verification standards will become a very difficult task.

In chapter fourteen, Kolja Brockmann discusses how current export control regimes are already applicable to AI and ML algorithms and how they should be improved to restrict the proliferation of malicious AI applications. Brockmann starts from the assumption that there is “lack of clarity about the extent to which export control instruments already cover dual-use goods and technologies used in AI and its military applications” (???). While examining existing export control regimes for dual-use goods, such as the Wassenaar Arrangement, in detail, Brockmann identifies both controls relevant to hardware (e.g., CPUs specifically designed for AI) as well as software, or even “technologies,” understood as specific information necessary for the development of AI tools. He then describes current review processes by, for example, the United States or the European Union and how these processes deal with emerging technologies. Going beyond existing regimes, Brockmann finally looks at challenges and opportunities in applying export controls to AI, weighing up the conflicting aims of export control and describing opportunities and benefits. He concludes that coordination and exchange between the major stakeholders will be the key to finding the right balance in the control of AI exports.

In the final chapter, chapter fifteen, Thomas Reinhold looks at a topic most people would consider a non-starter: the application of hard arms control measures to artificial intelligence and machine learning. While many observers would argue that conventional arms control instruments, such as verification and inspections, cannot be applied to software at all and that only weaker normative restrictions have a chance of being applied, Reinhold looks at best practices from the cyber realm as a source of innovative ideas. To achieve this he disaggregates the process of building an AI application into several independent elements, including training data, classifiers, the AI model and the effectors where the AI is finally applied, and discusses how specifically tailored arms control instruments could be applied separately. Reinhold himself points out that these considerations are currently only theoretical and that significantly more work is required in order to arrive at initial proofs of concept. Viewed optimistically, however, the chapter shows that the statement that hard arms control cannot be transferred to “soft” software needs to be reconsidered.

4 Conclusion

Looking at all fifteen chapters, several general conclusions can be drawn. As was to be expected, AI has an impact on almost all types of weapons. Even if individual weapons are not always optimized by AI, “mosaic warfare” (Torruella, 2021), that is, the enormous relevance of data and information exchange and analysis, has already arrived in many areas of the military. In more and more instances, humans are supported and assisted by AI, leaving the human as the slowest link in military decision-making. Developments are often driven by the AI race in the civilian sector. The states with a dynamic civilian technological AI base are also the states that want to reap the benefits for the military. Almost all authors fear that the use of military AI will lead to an increased speed of military operations and the need to act faster in times of crisis, leading to instability and hair-trigger alerts. The general unpredictability of current black-box AI algorithms must also be added to this, potentially worsening situations where human soldiers have to trust their computer. Thus, both finding ways to increase the reliability of AI as well as forms of control for AI are the imperatives of future research. But there are also positive developments: In many contexts, projects are exploring how AI can be used to enhance arms control in general and verification in particular. International institutions such as the IAEA are looking very closely at how they can harness AI for their own purposes (IAEA, 2020). While arms control is in its most severe crisis since its introduction in the 1960s, reliable AI might be a key to restarting arms control in a new and reliable fashion. However, there is also agreement that verification should not be outsourced to computers completely, but that AI should primarily aim at supporting human inspectors rather than replacing them.

Finally, we hope that this book will encourage experts from the arms control community who until now have shied away from the topic of artificial intelligence in their respective fields to dig more deeply into the issues. What is needed is genuine interdisciplinarity, something which is far too rarely seen. We hope that our book shows that interaction between the two professions is needed and possible.

References

Din, A. M. (Ed.). (1987). *Arms and artificial intelligence*. Oxford University Press.

Doll, T., & Schiller, T. (2019). *Artificial intelligence in land forces* (Position Paper). German army concepts and capabilities development Centre. Retrieved from Cologne: https://www.bundeswehr.de/resource/blob/156026/3f03afe6a20c35d07b0ff56aa8d04878/download-positionspapier-englische-version-data.pdf

IAEA. (2020). *Emerging technologies workshop. Insights and actionable ideas for key safeguard challenges. (Workshop Report STR-397)*. IAEA Safeguards.

Munoz, C. (2022). USN developing AI to drive tactical, operational focus. *Janes Defence Weekly, 59*(3), 8.

Orhaug, T. (1987). Computer applications in monitoring and verification technologies. In A. M. Din (Ed.), *Arms and artificial intelligence* (pp. 165–178). Oxford University Press.

Scharre, P. (2019). *Military applications of artificial intelligence: Potential risks to international peace and security*. Center for a New American Security. Retrieved from https://stanleycenter.org/wp-content/uploads/2020/05/MilitaryApplicationsofArtificialIntelligence-US.pdf. from Stanley Center https://stanleycenter.org/wp-content/uploads/2020/05/MilitaryApplicationsofArtificialIntelligence-US.pdf

Torruella, A. (2021). Mosaic warfare. *Janes Defence Weekly, 58*(40), 20–25.

White, A. (2021). Evolving connections. *Janes Defence Weekly, 58*(46), 20–25.