Ethics and Dynamics of AI System Use in Contemporary Armed Conflicts

Summary:

Contemporary armed conflicts are creating opportunities for the first genuine and widespread deployment of artificial intelligence (AI) systems in the military domain, while also serving as a highly dynamic and realistic “laboratory” not only for their further development and modification, but also for innovative ways of their hybrid use and direct application in combat.

The ethical and moral aspects of AI deployment in these conflicts pose a major challenge not only for the evaluation of existing ethical frameworks and recommendations, but also for a fundamental review of methods and the search for mechanisms to ensure their implementation and adherence in future conflicts. Future AI systems and military deployment strategies represent a particular aspect of this challenge.

Keywords:

artificial intelligence, lethal autonomous weapons (LAWs), surveillance, AI ethics

Author:

ThLic. Ing. Peter Šantavý, PhD.
Comenius University Bratislava
Faculty of Roman Catholic Theology of Cyril and Methodius
E-mail: peter.santavy@uniba.sk

APVV-23-0509

1. Introduction

The idea of intelligent machines has accompanied humanity for several centuries. With the development of computer technology and the gradual rise of the Third Industrial Revolution, this idea is taking on real contours.

Although general topics that remain relevant today (machine processing of human speech, neural networks, machine learning, abstract concepts and thinking, creativity, etc.) were defined at the dawn of the phenomenon *of artificial intelligence* (AI)[1] , representatives of intelligence services and the military also recognized the potential of AI very early on. They believed in the future capabilities and potential applications of artificial intelligence systems to such an extent that they were able to fund its development even during the so-called "AI winters," when the entire concept was in decline.[2]

Given the predicted capabilities of artificial intelligence since the days of the Dartmouth Conference, the development of military technology and military deployment stood as one of the first in line for practical application. And once the latest AI boom evolved into the practically continuous development of intelligent machines, artificial intelligence became one of the primary technologies not only for modern weapon systems but for military applications across their full breadth and complexity.[3]

Over the past two decades, a great deal has happened in the field of research and development of artificial intelligence systems. New insights, algorithms, the vast amount of available data, powerful systems, and the emergence of generative systems – all of this has contributed and continues to contribute to the unprecedented development of AI systems and the countless possibilities for their civilian applications.

AI is enjoying no less popularity during this period in intelligence services and the military. Particularly with the development of modern analytical, surveillance, and weapon systems, AI technologies have become indispensable, and mirroring the boom in the civilian sector, they have been given the green light across the entire spectrum of military and intelligence operations. Furthermore, there is a noticeable blurring of boundaries, or rather an overlap, between military and intelligence applications of artificial intelligence – some areas of deployment are identical, differing only in the intensity of the conflict (the cold war, or its peaceful equivalent of hybrid warfare, versus an actual hot military conflict).

---

1 The Dartmouth Seminar – a two-month research project on artificial intelligence held in the summer of 1956 at Dartmouth College in the United States – is considered the true birth of artificial intelligence.
Essentially, it was a workshop organized by the young mathematician John McCarthy in collaboration with Marvin Minsky (a former classmate who shared his fascination with intelligent computers and later became the founder of the Artificial Intelligence Laboratory at MIT), Claude Shannon (a colleague from Bell Labs/IBM and the inventor of information theory), and Nathaniel Rochester (a pioneer in electronics and the architect of several IBM computers).
ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master?*, 34.

2 During the so-called "winter periods" of artificial intelligence – that is, periods of significant decline in the field of AI – development was essentially limited to basic research at various university centers and progress in related fields (e.g., robotics and cybernetics, computing technology, data and computer science, within industry and military development).
DARPA (Defense Advanced Research Projects Agency) – the U.S. Department of Defense agency responsible for the research and development of new military technologies – has long played and continues to play a major role in this field.
ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master?*, 114.

3 Ibid.

Currently, the military and intelligence services are no longer merely dabbling in artificial intelligence, nor are they attempting an uncertain coexistence on a “floury book”[4] , but are forming a full-fledged “marital” relationship in which the phrase “until death do us part” takes on a whole new meaning.

2. Introduction to Artificial Intelligence

The military has long been familiar with and utilized various weapon systems characterized by a certain degree of automation, though this typically involves only the automated execution of tasks with little to no human intervention.[5]

True artificial intelligence systems must possess two fundamental characteristics – *autonomy and adaptability.* Autonomous systems are capable of acting independently. This refers to a system’s ability to perform tasks in a complex environment without constant guidance from a user. Adaptive systems demonstrate the ability to adapt, i.e., the ability to improve their performance (and capabilities) by learning from experience.[6]

Complex human intelligence encompasses a full spectrum of manifestations and characteristics that functionalist artificial intelligence lacks. Nevertheless, we can still discuss machine intelligence in terms of a continuum (one object is more intelligent than another) and multiple dimensions (intelligence across various domains). From this perspective, we distinguish *between narrow and general* artificial intelligence.[7]

Narrowly specialized artificial intelligence systems (ANI – Artificial Narrow Intelligence) are adaptive and autonomous only in a specific domain, i.e., they are capable of solving certain tasks “intelligently,” while failing in other areas. These are therefore highly specialized systems optimized to handle a specific task or set of tasks.

General artificial intelligence systems (AGI – Artificial General Intelligence) are capable of handling any intellectual task. Essentially, this refers to artificial intelligence that is on par with humans.

Similarly, we categorize artificial intelligence as *strong and weak* based on the distinction between truly intelligent systems and systems that, while acting intelligently, merely simulate (mimic) intelligence.[8]

Strong artificial intelligence (strong AI) is truly intelligent (and potentially even “conscious”), i.e., it actually understands what it is solving and executing. Truly intelligent AI is also general, because it has the ability to generalize – that is, to apply and transfer or adapt learned abilities to other tasks (which, incidentally, is one of the foundations of human thinking).

Weak artificial intelligence exhibits intelligent behavior based on models, applied methods, and the data on which it is trained. These are therefore systems focused on solving specific tasks and

---

4 In the 1950s, this was the term used in the former Czechoslovakia to describe couples who lived together but were not married.

5 Examples of advanced automation include, for instance, the older radar-guided CIWS Phalanx systems used in naval applications since the 1970s, and in tank systems, for example, the Russian Arena system, the Israeli Trophy, and the German AMAP-ADS.
ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master?*, 122.

6 ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master?*, 38.

7 Cf. Ibid.

8 Cf. ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master?*, 39.

dependent on human input and configuration.[9] Weak AI represents the systems we currently have, i.e., computer systems that exhibit intelligent behavior.

Currently, the acronym ANI refers to narrow and weak artificial intelligence, while AGI is generally understood to mean general and strong artificial intelligence.[10]

All the methods and systems of artificial intelligence we use today fall into the category of specialized AI systems (ANI), and current developments in this field are advancing by leaps and bounds, so many experts refer to the best current systems as BAGI (it should be noted that an approximately equal number of experts disagree with this view).[11]

Given our limited understanding of intelligence and the resulting inability to grasp the creation of true artificial intelligence, instead of a clear path for AI development, progress is achieved through such a diversity and complexity of AI methods that we can speak of anarchy.[12]

Despite this anarchy, however, we can at least in principle (philosophically) divide this broad set of methods into two underlying approaches – symbolic AI and subsymbolic AI.[13]

Symbolic artificial intelligence follows the path of creating artificial intelligence based on human thinking, i.e., concepts, words, phrases (= symbols), and the relationships between them. Symbolic systems, based on defined rules and procedures ("if this, then that"), can process individual symbols and perform assigned tasks. For symbolic AI, the meaning of individual symbols depends on how they are combined, their mutual relationships, and the operations that can be performed on them.[14] Since logic is a necessary condition for our understanding of general intelligence, symbolic systems attempt to solve diverse tasks requiring some degree of intelligence through logical operations. The strict formality of logical reasoning allows it to be algorithmized and translated into machine form.

Subsymbolic artificial intelligence and its development were inspired by advances in neuroscience. The subsymbolic approach to AI seeks to capture our thought processes, which we might sometimes call unconscious or automatic, and which form the basis of so-called fast perception – a process we use, for example, in face recognition or the identification of spoken words. Subsymbolic artificial intelligence programs thus do not contain a set of precise software procedures at the level of logical thinking, but consist only of a stack of equations – to the uninitiated, merely a jumble of hard-to-interpret numerical operations. Subsymbolic systems are designed to learn to perform tasks based on data.[15]

The symbolic approach – utilizing strict logic and relying on clearly defined characteristics of the environment and relationships between objects of AI systems' activities – has proven successful only

---

9 Creating a highly functional AI system requires true experts capable of applying the right methods and properly configuring and parameterizing the system. To the uninitiated, it may look like magic or voodoo :-)
Cf. MITCHELL, *Artificial Intelligence*, 98.

10 Moreover – in the context of current advances in generative AI systems – there is increasing discussion of various levels of AGI, with a more or less open effort to identify the best language or multimodal models at the boundary between ANI and AGI. The distinction between ANI and AGI is insufficient, and therefore AGI is typically divided into several levels, for example: BAGI (below-human), HAGI/HLAGI (human-level AGI), MAGI (moderately-superhuman AGI), SAGI/ASI (superintelligent AGI).
Cf. GABE. *Four Phases of AGI*.

11 On Platform X, there is a well-known, sharp yet scholarly rejection of classifying current generative systems into the BAGI category by the renowned Professor Gary Marcus (https://x.com/garymarcus), a cognitive scientist and one of the authors of advanced tests of AI systems' intellectual capabilities.

12 Cf. LEHMAN, CLUNE, RISI, *An Anarchy of Methods: Current Trends*, 56–62.

13 Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master?*, 40–44.

14 Cf. MITCHELL, *Artificial Intelligence*, 21–23.

15 Cf. MITCHELL, *Artificial Intelligence*, 24.

in solving specific types of problems based on concrete environmental characteristics, selected interactions of AI systems with that environment, and a given specific set of goals. Everything else that could and should be the subject of any intelligence, however, was off-limits to these systems – symbolic systems failed and continue to fail when solving problems that cannot be precisely described, and in real-world environments that cannot be deterministically grasped and are full of ambiguous information. Therefore, most modern implementations of artificial intelligence systems (which include machine learning algorithms in general and neural networks in particular) are based on a subsymbolic approach that attempts to address these problems and, to a certain extent, successfully solve them – whether through classical methods, such as regression and statistical methods, or by emulating brain activity at the neuronal level via so-called neural networks.[16]

On November 30, 2022, OpenAI introduced ChatGPT – a large language model that generated desired responses based on text inputs. Technologically, it was based on the 3rd generation of the GPT (Generative Pre-trained Transformer) model, which was a representative of the promisingly developing field of so-called *generative artificial intelligence* (genAI). As the name suggests, the algorithms of generative AI systems are capable of generating diverse content – text, images, audio, video, and more.[17]

At the core of large language models are so-called GPT modules, which operate by analyzing the input sequence and using complex mathematics to predict the most likely output. As an application of deep learning technology in artificial intelligence, GPT models can process prompts in natural language and generate relevant, human-like text responses. Extensive training datasets enable GPT to mimic the ability to understand human language.[18]

The capabilities of GPT models stem from two key aspects. These are generative pre-training, which teaches the model to recognize patterns in unlabeled data and subsequently apply these patterns to new inputs, and the transformer architecture, which allows the model to process all parts of the input sequence in parallel and generate outputs.

Generative systems – whether language, image, or video-generating models – can be remarkable at processing input, generating responses, and producing outputs, often delivering better results faster than humans. As the development of generative AI continues, we are encountering various advanced applications and modifications of these systems. In particular, we can mention so-called *reasoning models*, which are capable of mimicking advanced modes of reasoning when generating responses, and *AI agents* that can independently perform complex tasks based on given commands to achieve a specified goal. Their applications are truly wide-ranging and can be successful in understanding risks and consistently applying principles of proper use and control.

---

16 ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master?*, 43–44.

17 ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, 79.

18 Based on statistical analysis of a vast amount of text, models learn syntactic patterns and relationships that are captured in neural weights (not in explicit grammatical rules).
The emergent result is so-called semantic representations, which are not the formal semantic analysis of traditional linguistics, but rather the learning of meanings and relationships between words from context.
Large language models, or genAI in general, do not perform syntactic and semantic analysis – they "perceive" it, or rather, realize it through learned patterns. Therefore, "understanding" of language is not the result of formal processing of grammar and meaning, but is the fruit of probabilistic learning on a large amount of training data.
ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, 80.

3. Risk Factors of AI Systems and Their Consequences

The use of AI systems is becoming so commonplace that we don't even realize we're using them regularly; however, we accept them and – as we grow accustomed to the benefits of their deployment – we come to demand them, since they provide us with added value in many ways that we don't want to give up. In general, the focus on the convenience and benefits of information technology is often much stronger than caution and safe behavior when it comes to protecting against cyber threats, data leaks, and the misuse of personal data, and so on. A similar lack of caution accompanies society even in the use of artificial intelligence systems. The current level of acceptance of simple AI systems and the way they are used suggests that the use of sophisticated AI systems may entail risks whose scope we cannot even imagine.

In this chapter, we will at least briefly address the issues we are aware of and must take into account when designing, developing, and using artificial intelligence systems. First and foremost, these involve *technological failures, limitations, and risks of AI systems*. Their logical consequence is *the risks associated with the use of AI systems in real-world deployment*, which include not only the misuse of AI systems by humans but also the complex impacts and effects of AI system use on humans and society.[19]

We are still operating in the realm of ANI, i.e., narrow artificial intelligence systems (narrow AI) that are optimized to handle a specific task or set of tasks. These are also systems of weak AI, which exhibit intelligent behavior based on models, applied methods, and training data. Even with the latest sophisticated AI technologies – which undoubtedly include advanced reasoning models and genAI agents – we are still talking about systems that are focused on solving specific (sets of) tasks and are dependent on human input and configuration.[20]

One of the fundamental risks stems from the fact that *we have virtually no or small understanding of how deep neural networks make their decisions*. We know how to design a neural network for a specific application. We know how to train it and, within the limits of what is possible, how to test it. However, since a neural network does not contain a set of precise software procedures at the level of logical reasoning, but is composed solely of a stack of equations – a dense network of hard-to-interpret numerical operations that function based on the correct setting of weights, constants, and threshold values, we essentially do not know exactly what the neural network has learned and how reliably it can apply it not only in normal operation but especially in borderline situations under extreme input conditions or during system operation. This degree of uncertainty increases with the complexity of the neural network, or rather the entire sophisticated AI system. The *black-box* problem is quite complex; it also encompasses phenomena such as the "alchemy" of hyperparameters (the basic design and tuning of the system), "movements" in models (generating different outputs based on nuances in the input), the concealment of reasoning (the inability to force reasoning models to disclose the essential steps of their reasoning), etc...[21]

---

[19] ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, 139.

[20] Since AGI systems – i.e., systems of strong and general artificial intelligence – do not currently exist, the focus on ANI is natural. Moreover – given the current absolute preference for neural networks and machine learning – we will focus primarily on their current manifestations in deep neural networks and deep learning.
Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master? Second, expanded edition*, 138–139.

[21] A sophisticated AI system appears as a black box that performs a task, but how and why it does so is not entirely clear.
Cf. ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, 141–144.

Whether we are talking about the black box, alchemy, multidimensional movement, or concealment, current AI systems are accompanied by legitimate and serious concerns that *if we do not understand how these systems work, we cannot truly trust them and will struggle to predict the circumstances under which these systems will fail*.

The answer to the black box problem lies in so-called *explainable and interpretable AI systems*. This involves an effort to develop systems that are explainable (we can describe how they work) and interpretable (we can explain what the outputs mean). Although this is a rapidly evolving area of AI technology development, it must be acknowledged that the following still holds true: the more sophisticated the AI system, the more limited – or problematic – its explainability and interpretability become.[22]

Over several decades of development in neural networks and machine learning systems, many *risk factors and vulnerabilities in AI systems* have been gradually identified. The interdisciplinary FutureTech group at MIT maintains a continuously updated comprehensive database of AI risks, categorized by cause and risk area. It is based on 43 different AI frameworks originating from research, government, and industrial organizations. The AI Risk Repository is truly "live"; in 2025, approximately 100 new risks were added monthly, and currently (March 2026), it catalogs more than 1,700![23]

Let's at least mention the most significant ones (illustrative and occurring across a wide spectrum of systems).[24]

Small training dataset. The success of most current artificial intelligence systems is extremely dependent on extensive and high-quality sets of properly labeled training data, or training iterations. Without them, current machine learning systems cannot be trained, and their absence leads, at best, to poor-quality results and, at worst, to incorrect results and catastrophic failures. Moreover, in the context of other types of vulnerabilities and risk factors in AI systems, the issue of training data is much broader...

An improperly selected or low-quality training dataset and biases. Based on an improperly selected or low-quality training dataset, an AI system learns to draw incorrect conclusions or produce "biased" results. In many cases, there is a risk that AI systems trained on biased data may amplify these biases and cause real harm. Another aspect of "biased" AI systems is the reduced effectiveness of their operations due to these biases.

Overfitting to training data. This is an undesirable behavior of a machine learning system that occurs when a machine learning model provides accurate answers for the data on which it was trained, but not for any other inputs. An overfitted model may provide inaccurate predictions or learn to distinguish something different from what it was supposed to learn from the training data.

Long-tail effect. In the field of artificial intelligence, this term refers to the wide range of possible unexpected situations that an AI system might encounter. In the real world, we simply cannot describe everything and present it to machine learning systems for training.

---

22 Explainability and interpretability lead to greater transparency in AI systems. The term eXplainable AI (XAI) is commonly used, encompassing the entire spectrum of methods contributing to increased trust in AI models. Cf. ALI, ABUHMED, EL-SAPPAGH, et al., *Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence*.

23 *AI Risk Repository.*

24 Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master? Second, expanded edition*, 147–159.

Fooling deep neural networks and vulnerability to hacking. Neural networks are prone to failure when presented with adversarial examples. The result is a whole spectrum of very simple ways to fool deep neural networks. Unfortunately, many of the possible attacks are surprisingly robust, capable of effectively fooling various and diametrically different sophisticated machine learning systems.

Superstition. We usually refer to superstition as the mistaken belief that a certain action or act can help bring about a good or bad outcome. In the field of artificial intelligence, this is primarily an issue within reinforcement learning algorithms, where the training of the model (agent) is carried out through interaction with the environment using a trial-and-error method. In the context of training an AI system, a superstition arises when the system mistakenly learns to perform an unnecessary, or even dangerous, action or activity to achieve the desired goal.

Despite the undeniable success of the development and deployment of current artificial intelligence systems across a wide range of academic and real-world environments, *we must always bear in mind that these systems can fail in a wide variety of often unexpected ways* due to the inability to prepare a sufficiently large set of training data, or the selection of data that is of insufficient quality or biased, overfitting to training data, the long-tail effect, risks arising from deep network deception, their vulnerabilities, and biases – all of which are compounded by a lack of the technical expertise required to design and tune hyperparameters when developing a functional and successful solution. Upon closer examination of these issues, we may also be confronted by their further consequences – not only the risks of direct failure, but also the reality of results that can be difficult to interpret correctly (what the network has actually learned, what the output from the given input data actually means) and the inability to predict when individual failures will manifest (under what conditions, through what combination of circumstances, as a result of what dynamics of internal development, or the operation of the AI system).[25]

Generative AI systems also bring other potential and real problems, e.g., hallucinations (making up answers that the system presents as relevant and correct), biases and biased outputs (the risk of half-truths and incorrect answers), unclear data handling practices (which may result in leaks of confidential data, breaches of personal data protection, and copyright issues), issues related to the loss of skills due to the replacement of certain job positions, consequences for society and democracy (propaganda and manipulation), misuse as a tool for cybercrime, loss of critical thinking and sense of reality, risk of digital dementia and brain rot, digital division, disruption of psychological development, erosion of identity, deepfakes, algorithmic modeling of history, and more...[26]

Although it appears that genAI systems “think” similarly to humans, we must realize that these are statistical models and dynamic systems. *The modus operandi of these systems is significantly different from the way the human brain works*.[27]

Therefore, regarding generative AI systems, we caution that t*hey cannot be viewed as factual, reliable, or ethical sources, and that they require a human who knows how to use them correctly and verify their results*.

The risk factors mentioned so far imply serious consequences:[28]

---

25 ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, 160.

26 Cf. ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, 190–210.

27 For example, generative systems, by their very nature, do not know whether an answer is correct or not, since they offer only the statistically most probable outputs.

AI systems *"think" completely differently than humans*. Their failures differ from human ones, are difficult to predict, sometimes easy to execute, and often surprisingly robust.

AI systems do not actually think – *they merely mimic intelligence*. AI systems do not understand meaning; they do not possess true intelligence, they only simulate it.

The operation of AI systems carries serious ethical implications. *They are incapable of distinguishing between moral good and evil! They do not understand meaning or consequences!*

4. General Ethical Principles in the Field of AI[29]

The fundamental principle for any artificial intelligence system is a focus on human well-being, that is, the well-known and widely accepted *principle of human-centered and beneficial artificial intelligence.*

This fundamental principle should be understood in the spirit of Christian anthropology, building upon biological and cultural anthropology, upholding human dignity and promoting the integral development of the human person and society, embracing every human being and discriminating against no one, keeping in mind the good of humanity and society, while protecting and respecting the good of every human being, and be characterized by care for our "common and shared home," that is, the entire created world.

This principled stance of prioritizing the human good thus becomes equivalent to the issue *of AI trustworthiness*, whereby conditions must be established without which the deployment of AI systems into the real world – where they interact with humans and influence society – should not be permitted.

Based on the Ethical Guidelines for Trustworthy Artificial Intelligence from the EU Expert Group on Artificial Intelligence, as well as more recent EU and UNESCO regulations and guidelines[30] , we formulate *the basic requirements for trustworthy AI systems*, which must be:

– *functional and useful* – designed and implemented to perform the intended task.

– *lawful* – compliant with required standards, laws, and regulations, and meeting all applicable laws and regulations.

– *ethical* – respecting ethical principles and values.

– *resilient and robust* – meeting the necessary standards of safety and reliability not only from a technological perspective but also taking into account the social environment and impacts on society.[31]

The ethical requirements, proposed by the author based on recommendations from the EU, IEEE, and UNESCO, are as follows:

---

[28] Cf. ŠANTAVÝ, KUBICOVÁ, *Artificial intelligence systems do not possess true intelligence; they merely simulate it. They do not distinguish between moral good and evil*.

[29] Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master? Second, expanded edition*, 280–295.

[30] *Rome Call for AI Ethics.*
Ethics guidelines for trustworthy AI.
The global landscape of AI ethics guidelines.
Artificial Intelligence Act.
UNESCO Recommendation on the Ethics of AI.

[31] We are talking about technological security and societal safety.

1. In the development, production, deployment, provision, and use of artificial intelligence systems, the protection of the freedom, dignity, and safety of every human person and of society as a whole must be guaranteed.

2. Artificial intelligence technologies must be fully under human control and controllable by humans.

3. The algorithms and results of AI systems must be understandable and reviewable by humans.

4. Any deployment of AI technologies must be beneficial to people and society.[32]

5. Artificial intelligence systems must not be a tool of digital division.

6. Artificial intelligence technologies must not harm our common home and should contribute to social and environmental well-being.

These requirements further emphasize the aforementioned framework for trustworthy AI systems, as they cannot be fully realized without meeting the criteria of functionality, legality, and resilience.

5. Specific Ethical Principles and Recommendations[33]

In the areas of military applications, intelligence services, and algocracy[34] , in addition to the general principles presented so far, further recommendations and necessary conditions for the trustworthy use of AI systems must be emphasized.

Given the specific nature and scope of AI technology deployment *in the areas of advanced state governance, intelligence services, and widespread surveillance* – which impact human rights, the protection of democracy and freedoms, we believe that this area should, in addition to a technological framework, be fundamentally covered by basic legislative mechanisms and public oversight within a democratic society.

Without rigorously implemented control mechanisms at the level of laws and the constitution, it will likely not be possible to effectively and sustainably ensure the implementation of criteria for trustworthy AI systems. Moreover, considering that under current legislation, military and intelligence AI systems are typically exempted and granted exceptions.

We also propose that the export of artificial intelligence products and technologies that could be misused in the areas of advanced state governance, intelligence, and mass surveillance be subject to international regulation to prevent their export to high-risk countries. However, sanctions should not be a tool for (geo)political struggles, but rather for genuine, responsible engagement in the field of the ethical use of AI systems.

The area of military development, deployment, and use is more complex.

The implementation of AI systems in this area does not occur in a legal vacuum, but within existing legal frameworks, within which *fundamental principles must be upheld*: the principle of the legal regulation of war, the principle of humanity, the principle of distinction (military/civilian targets), the principle of proportionality (appropriate to the objective pursued), and the principle of necessity.

---

[32] They must minimize toxic psychological and social consequences.

[33] Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master? Second, expanded edition*, 296–305.

[34] This concerns the use of AI systems in the field of advanced state governance. It encompasses a wide range of areas and levels of application, from algorithmic support for the activities of government agencies, courts, and decision-making processes, to algorithm-based governance, an algorithmic legal system, algorithmic management of society, and so on.

It is also necessary to distinguish between legal frameworks outside (e.g., ius ad bellum – the right to wage war) and during armed conflicts (ius in bello – the law applicable during war).

In times of peace, many state actors would like to apply the Latin maxim "*si vis pacem, para bellum*" to the military deployment of AI technologies in weapon systems and their potential to strengthen their position. For others – looking toward the horizon of possibilities for autonomous weapon systems – their introduction into military arsenals could resemble the deterrent potential of nuclear weapons. However, if we look beyond the current capabilities of military AI systems, we see technologies whose risk potential may even exceed the danger posed by current nuclear arsenals.

The field of military deployment, however, is under great pressure from potential technological advantages – or disadvantages.

Despite all the risks, the capabilities of AI-powered autonomous weapon systems and cyberweapons are leading to increasing pressure to fund and deploy offensive cyberweapons. Individual countries are unable to give up such a tempting technological advantage and are firmly determined to implement artificial intelligence technologies across the full range of possible meaningful applications. The issue of restricting these offensive systems is further complicated by the blurring of lines between defensive and offensive deployment in nearly all areas of military application of artificial intelligence technologies.[35]

Since any restriction of artificial intelligence technologies in the military sphere can be perceived as a security risk and a reduction in the combat effectiveness of a modern army, unilaterally adopted regulations may not be effective – not only because they are difficult for one side to accept (though for a values-oriented society, this should be an obligation), but also because of the slim chance of their extraterritorial expansion and acceptance.

Furthermore, weapon systems powered by artificial intelligence may raise new questions that current legal frameworks did not anticipate. However, the Martens Clause of humanitarian law should always apply as a minimum.[36]

Many military systems utilizing AI technologies can operate autonomously. In principle, we distinguish three dimensions of autonomy based on the relationship between human and machine (human-in-the-loop weapons, human-on-the-loop, human-out-of-the-loop weapons), the complexity of the machine (automatic, automated, autonomous), and the type of automated decision (what task is to be performed, the ability to perform it, and how to perform it).

Given the degree of autonomy in the decision-making of combat systems, *fully automated lethal weapon systems (LAWs) pose the greatest threat from an ethical perspective*. However, in the following text, we will show that *other systems can also be deadly dangerous, in which, for various reasons, we transfer control and decision-making authority to machines*.

In general, one can agree with the principles of the document *"AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense" from the U.S. Defense Innovation Board*[37], which includes key conclusions such as accountability (decision-making

---

35 There is no issue with the humanitarian, medical, communication, and, in most cases, defense applications of AI systems in the military.

36 The Martens Clause serves as a "safety lock" and stipulates that in cases not expressly regulated by applicable international conventions, civilians and combatants remain protected by the principles of international law arising from customs established among civilized nations, humanitarian principles, and the dictates of public conscience. Cf. GEFFERT, *The Martens Clause in International Law of Armed Conflict*.

37 Cf. *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense.*

authority), caution in preparing test data and system design, traceability, reliability, and controllability.

In the case of any human-controlled autonomous weapon systems, the following principles must apply:

– A necessary condition for the operation of any AI system that may pose a risk to any human being is the ability and opportunity for a human to take control of the system at any time, or the right and opportunity to verify and review the results of its operations.

– The limits, regulation, and restrictions on LAWs should constitute an ethical framework based on the moral values of human society, not on relativistic so-called "follow-up regulation."

In the case of fully automated lethal weapon systems (LAWs), a clear position must be taken: *Artificial intelligence-driven lethal autonomous weapon systems, automatic targeting and target-selection systems, and autonomous systems capable of deciding on a lethal response of any kind (from a drone strike to the unleashing of a nuclear conflict) without human intervention must be banned*.

In the context of ongoing military conflicts in which various AI systems are being used, it is worth highlighting the aforementioned relativistic "follow-up (ex post) regulation". This refers to the stance of certain military and political circles that reject limits, regulations, and restrictions on LAWs based on the moral values of human society. "Follow-up regulation" should involve a wait-and-see approach, with regulation evolving in response to new advances in the development and deployment capabilities of LAWs.

Essentially, this involves speculating on the evolution of ethics and the value framework, thereby adapting to the state of development of AI-based weapon systems. Legal scholars such as Kenneth Anderson and Matthew Waxman, who advocate this approach, argue that regulation will have to evolve alongside the technology and believe that ethics and the boundaries of moral justifiability will develop in tandem with technological progress.[38] Supporters of "follow-up regulation" can thus easily slip into the realm of value and moral relativism, attempting to reevaluate arguments regarding the irreplaceability of human conscience and moral judgment.

Despite reduced competitiveness and a slim chance of acceptance by other states, unilaterally adopted ethical and legal regulations should be a duty for a value-oriented society. We call for this step for several reasons:

– No state, if it renounces ethical principles and moral standards, has the right to justify its participation in a military conflict and emerge victorious.

– With the deployment of modern weapon systems with area-of-effect capabilities and technologies that, in hybrid wars and fourth-generation wars, affect virtually the entire population of states, the concept of a just war becomes unacceptable.

– Only on the basis of specific commitments can the global discourse among powers, public pressure, the engagement of various segments of society in different regions of the world, and the efforts of responsible parties into the gradual adoption of global rules and commitments regarding the development and deployment of high-risk military systems equipped with artificial intelligence technologies.

---

[38] Cf. ANDERSON, WAXMAN, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can*.

To conclude this chapter, let us summarize the specific ethical requirements for the military deployment of AI technologies, which *are based on the fundamental framework of trustworthy AI systems*.

The principle *of ius in bello* is important, encompassing proportionality, distinction, accountability, and responsibility, as well as the prohibition of unnecessary suffering.

Conditions for the military deployment of AI – the fundamental requirement to keep AI systems under control:

– responsibility (decision-making authority) during deployment;

– caution when preparing training data (unintended bias);

– traceability of every step in the system's operation;

– reliability in deployment;

– controllability during every step of operation.

Limitations of military AI deployment:

– A necessary condition for operating an AI system that may pose a risk to any human being: the ability and opportunity for a human to take control of the system at any time, or the right and opportunity to verify and review the results of its operations.

– A principled stance on LAWs: AI-powered lethal autonomous weapon systems, automatic targeting and target-selection systems, and autonomous systems capable of deciding on lethal responses of any kind without human intervention must be prohibited.

– An ethical framework for the limits, regulations, and restrictions on LAWs: it must be based on the moral values of human society, not on relativistic so-called "follow-up regulation."

6. Military Use of AI in General[39]

The use of AI technologies in the military is not limited to countries where one might expect it given their scientific, technological, or military capabilities. With the availability of modern AI-based systems and their commercialization, these systems are attractive to virtually anyone.

The implementation and use of AI systems in the military sector is advancing primarily in the following countries:

– superpowers (the U.S., China, Russia, etc.), which possess both sufficient scientific and technological potential and large military forces;

– highly technologically advanced states (Israel, Japan, some EU countries, etc.), whose technological portfolio almost naturally includes military applications;

– highly militarized countries (India, Turkey, etc.), which are building modern military forces and investing in state-of-the-art technologies in this field;

– states considered problematic from NATO's perspective (Iran, North Korea, etc.), which, for their defensive, ideological, and political goals, emphasize the development of military capabilities but, within their means, focus on technologies that are both accessible and effective for them, including AI systems.

---

[39] Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master? Second, expanded edition*, 205–235.

These countries are striving to adopt the use of artificial intelligence in the military domain in a relatively comprehensive manner. Currently, however, not only they but nearly every country seeking to develop or build modern military forces utilizes AI systems in at least some of the following areas:

– military intelligence,

– modeling of technologies, conflicts, and operations,

– command support,

– trainers, simulators, and training,

– autonomous weapon systems,

– group control of combat assets and autonomous systems,

– cyber warfare,

– enhancement of manpower (skin-in and skin-out),

– protection and rescue of personnel,

– damage mitigation and ensuring basic life support.

The above list indicates that the use of artificial intelligence technologies in the military is not limited to weapon systems and the management of combat operations. *The deployment of AI systems can have a positive impact* on the protection of civilians and military personnel, in rescue operations, in damage mitigation and the provision of basic necessities, in replacing dangerous training areas with trainers and simulators, etc.

Several current conflicts, their individual phases, and their course can be classified as fourth-generation wars, which are characterized by the blurring of boundaries between war and politics[40] , between the military and the civilian population, decentralized warfare, guerrilla tactics and elements of terrorism, disinformation and propaganda, attacks on culture, and psychological methods to weaken the opponent.[41]

Fourth-generation wars are characterized by the massive adoption of artificial intelligence, so that many AI-powered military systems (particularly those in the category of cyber warfare) can be classified as so-called fourth-generation warfare capabilities.

7. AI in Current Armed Conflicts

Recent years on the international stage have been and continue to be very turbulent. Several serious disputes have escalated into open military conflicts that go beyond the “normal” scope of smaller local conflicts. These include, in particular, the war in Ukraine, the Israeli-Palestinian conflict in Gaza, and the military attack by Israel and the U.S. on Iran. The actors in all three conflicts are armies that possess modern autonomous weapon systems and AI technologies. These conflicts, which are a

---

[40] It should also be noted that the actors need not be limited to states or state-led groups. These may include not only proxy organizations of certain states but also any other non-governmental actors.

[41] An interesting concrete example of the complexity of the fourth generation is a study by the prestigious American think tank RAND Corporation on the possibilities of destabilizing and economically exhausting Russia. Cf. DOBBINS, COHEN, CHANDLER et al., *Overextending and Unbalancing Russia: Assessing the Impact of Cost-Imposing Options*.

tragedy for humanity and a source of great suffering, have simultaneously become a vast laboratory for military technologies and strategies. Artificial intelligence plays a major role in this laboratory.

7.1. Drones and Robotic Systems on the Front Lines of the Conflict in Ukraine

The war in Ukraine has brought about a clash of technologies between NATO and Russian forces. This technological clash is not static but extremely dynamic; specific deployment strategies and operational procedures change on a monthly, and sometimes even weekly, basis. Not only are the procedures changing, but the technologies themselves are evolving as well.

One of the key phenomena of military operations is the massive use of drones. Both the Russian and Ukrainian armies have specialized units (e.g., Russia's Rubikon) that resemble highly specialized technology teams more than combat deployment units. In addition to this specialization, operational procedures and doctrines are changing, so that the reconnaissance and combat use of drones is also finding its way into regular units.

The effective deployment of drones is typically dependent on the actions of operators who remotely control them. Remote control thus becomes the Achilles' heel of their successful operational deployment.

Consequently, during a conflict, not only are methods for protecting equipment and destroying attacking drones developing extremely rapidly, but even more so are (radio-)electronic warfare (REW/EW) capabilities capable of effectively jamming the remote control of these machines. Radio-controlled drones thus become unusable in certain areas of the front line.

The response to sophisticated and high-quality EW is the development of alternative control methods. For example, Russian drones controlled via a special optical fiber have achieved great success. Another technology involves the development of control units powered by artificial intelligence, enabling drones to operate autonomously in the so-called "last mile" (front line, heavy EW).

Most drones operating on the front line fall into the category of so-called FPV drones, which target live targets. Given the current state of AI system development and size and weight constraints, the implementation of AI control units in these devices is problematic. *The AI in these drones is unable to analyze a potential target in detail, i.e., to determine whether it is a civilian or a legitimate military target*. If it is a soldier, whether they are attacking or surrendering, or whether they are wounded.

Another problem is the analytical use of AI systems in combination with advanced satellite imaging and monitoring.[42] The immediate and deep combat zone thus becomes nearly transparent. Therefore, the strategy focuses on detecting enemy forces while simultaneously deceiving their surveillance systems. As a result, the front line between the two forces, extending approximately 40 km on both sides, is now a highly lethal zone through which it is difficult to fight one's way to victory.[43]

For this reason, Ukraine and Russia are gradually upgrading their current semi-autonomous air, land, and naval systems using artificial intelligence. As a result, these robotic systems will be far less vulnerable to electronic warfare jamming (local automation via AI algorithms) and will be able to autonomously identify enemy targets and (independently) launch attacks. The issues regarding the aforementioned risks and shortcomings of AI drone control units also apply to these systems.

---

[42] On the Ukraine/NATO side, systems such as Palantir Maven, Palantir technologies integrated with large language models from Anthropix and OpenAI, Arta GIS, and others are used.

[43] Cf. LAYTON, *Artificial intelligence at war.*

7.2. Lavender: Military Analytics in Lethal Deployment in Gaza[44]

The Israeli Defense Forces (IDF) have been attempting to implement and utilize AI systems in weapon systems for quite some time. The IDF even dubbed its 11-day campaign in Gaza in May 2021 the "first war of artificial intelligence." In the current offensive in Gaza, Israel is using AI capabilities in three categories:

– lethal autonomous weapon systems (LAWS) and semi-autonomous weapons (semi-LAWS), e.g., drones, robotic gun turrets, etc.;

– facial recognition systems and biometric surveillance of Gaza's residents to create an extensive database of residents' biometric data;

– automated target generation systems: Lavender, Where is Daddy?, and Gospel.

Even before the conflict in Gaza, Israeli intelligence services were using the Lavender analytical system, which used artificial intelligence to evaluate data from the Palestinian environment and was able to identify Hamas leaders and generate potential individual human targets. Gospel generates infrastructure targets, and Where is Daddy? focuses on tracking and targeting suspected militants when they are at home with their families.

In its original mode, the results of the Lavender system were subject to human supervision, i.e., the review and validation of potential targets. The system also focused specifically on Hamas leaders.

Since the outbreak of the conflict in Gaza, the use of the Lavender system has changed – the system is now used to identify any Hamas member, and the process of reviewing the generated results has also changed: instead of thoroughly verifying the outputs, the supervision time has been gradually reduced to a few tens to a few seconds! *In effect, the human approval of target confirmation has been delegated to machines*. Instead of human oversight and decision-making, an AI system now decides on matters of life and death.[45]

In addition to delegating decision-making to machines, the Lavender system is associated with other problems:

– *the data on which the system was trained was flawed*, as it also contained information about non-military employees of the Hamas government in Gaza, leading the Lavender system to mistakenly identify as targets individuals with communication or behavioral patterns similar to those of known Hamas militants. Among them were police officers and civil defense workers, relatives of militants, and even people who simply shared the same name as Hamas members.

– despite the fact that *the Lavender system has an error rate of approximately 10%* when identifying whether an individual belongs to Hamas, the IDF obtained general approval to automatically accept its lists of individuals designated for elimination "as if it were a human decision." Soldiers were not required to thoroughly or independently verify the accuracy of the Lavender system's outputs or its intelligence sources; the only mandatory check before approving a bombing was to ensure that the marked target was a man, which took approximately "20 seconds."

---

[44] Cf. FATAFTA, LEUFER, *Artificial Genocidal Intelligence: How Israel Is Automating Human Rights Abuses and War Crimes*.
Cf. ABRAHAM, *"Lavender": The AI machine directing Israel's bombing spree in Gaza*.
Cf. LAYTON, *Artificial Intelligence at War.*

[45] The so-called "kill chain" – i.e., the sequence from target identification to the actual attack – is being significantly shortened.

There is no ethical or humane way to use systems like Lavender or Where is Daddy?, as they are based on the fundamental dehumanization of people. Since the use of these systems began even before the outbreak of open conflict, the entire "peacetime" surveillance infrastructure, biometric databases, and other tools that enable the deployment of such systems in war zones must be eliminated even in peacetime.

7.3. Generative AI Systems Used in Attacks on Iran

According to available information, analytical and generative artificial intelligence systems play an important role in several phases of U.S. and Israeli military operations against Iran:

– Palantir's Mosaic analytical AI system likely contributed to the miscalculation of the risk of Iran developing nuclear weapons within a matter of weeks.[46]

– Anthropic's Claude model was used in intelligence operations and planning for an attack on Iran. The model was intended to assist in processing intelligence data, supporting target selection, and simulating possible scenarios. In this deployment, the system does not decide on a specific target or the bomb's impact. It is an analytical tool capable of processing vast amounts of data – including satellite imagery, intercepted communications, and logistical movements – within seconds to generate risk assessments or recommendations.[47] Interestingly, this use occurred only after a federal ban on the use of Anthropic's AI systems by U.S. government agencies. The official reason for the ban is security, but according to Anthropic, it was due to unauthorized use, as its models are not intended to be used directly for military interventions and operations. The Pentagon, however, argues the opposite: if AI is critical to national security, the technology company should not, in its view, determine the limits of its use.

– Iran is using Chinese artificial intelligence technologies to precisely target U.S. bases in the Middle East. This is a commercial solution from the Chinese company MizarVision, which, based on object recognition, data analysis, and long-term digital footprints, can identify aircraft, radars, fuel depots, and troop concentrations and prepare intelligence for missile and drone strikes. The plan for a specific attack itself is also the subject of planning by a generative AI system. What was previously the subject of a lengthy manual analysis is now a matter of a few minutes. Using AI technologies, Iran is attempting to asymmetrically confront a technologically superior adversary.[48]

Several important facts emerge from the above.

The potential of generative systems can be a major asset for the analytical processing of vast amounts of diverse data, the creation of risk assessments and recommendations, the modeling of conflict and operational scenarios, and for comprehensive command support. The flip side of genAI technologies is the reality of their risks and limitations. *Without the application of comprehensive risk management frameworks, the reliable deployment of generative systems is questionable and, in classified environments with lethal consequences and where decision-making is critical, downright dangerous.* The use of genAI technologies in recent conflicts highlights the ethical problem of sophisticated AI systems, which fundamentally fall into the category of non-lethal lethal autonomous

---

46 SANTOS, *Tulsi Gabbard now says Iran could produce a nuclear weapon 'within weeks'*.

47 Cf. STOJKOVSKI, "*US forces used Claude in Iran strikes for intelligence, targeting even after Trump's ban*."

48 Cf. SERVAES, *Iran Uses Chinese AI Satellite Imagery to Target U.S. Military Bases and Equipment in the Middle East*.
See SINHA, *Iran using AI-enhanced satellite images from China to hit US bases in the Middle East*.

weapon systems. Reality shows that *they are capable of operating and being deployed even in modes that could be ethically unacceptable*.

The Pentagon's argument that "if AI is critical to national security, a technology company should not, in its view, determine the limits of its use" is very close to the philosophy of "follow-up regulation," which we have already mentioned in the section on Specific Ethical Principles and Recommendations. By disregarding ethical constraints, we can easily slip into the realm of value and moral relativism, attempting to reevaluate arguments regarding the irreplaceability of human conscience and moral judgment.[49]

In 2023, I wrote: "Regarding the ethical-legal discourse, much of which takes place within U.S. military circles, it must be said that the current environment is conducive to the implementation of a clear ethical framework and rules. In 2019, the government's *Defense Innovation Board*, in the document *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense*, proposed principles for the use of artificial intelligence systems in the U.S. military, with the essential tenet of these principles being that decision-making authority remains with humans, particularly in missions involving the use of lethal force."[50] Unfortunately, the relatively brief experience of a few hot conflicts has been enough to show that the reality of applying this ethical framework looks entirely different.

7.4. A Vision of the Near Future

Examples from conflicts in Ukraine, Gaza, and Iran highlight the immense dynamism of the development and deployment of AI systems in current military conflicts (changes occur practically on a monthly basis). Equally dynamic is the procedural aspect of the matter, which is subject to current strategic and political objectives. *This dynamic is virtually devoid of any ethical assessment or regulatory oversight. We are thus witnessing the uncontrolled implementation of AI technologies in current conflicts.*

This unflattering state of affairs leads us to reflect on what the vision for the use of AI systems might look like in the near future. The following realities certainly belong to this vision:[51]

– technological synergy, i.e., an escalating transition from standalone systems to a complex ecosystem encompassing comprehensive AI systems for command and control (analytical "unlimited" Palantir, surveillance Flock, genAI,...) and group control of combat assets and autonomous systems (e.g., "motherships");

– intensive deployment in hybrid operations, where the distinction between attack and defense, active (e.g., hacking) and passive (e.g., surveillance) operations, is blurred; modern forms of propaganda are extensively used, and public opinion is manipulated (e.g., sentiment prediction and modeling...);

– the use of the latest AI technologies (e.g., genAI) will grow despite insufficient ethical and regulatory frameworks;

– the application of so-called agentic AI and the emergent potential of multi-agent systems in achieving objectives may pose new and complex risks, particularly in the areas of delegating decision-

---

49 Cf. SHANKLIN, "*The Pentagon's Pressure Campaign*."

50 ŠANTAVÝ, *Artificial intelligence – a good servant or a bad master?*, 138–139.

51 Cf. LAYTON, *Artificial Intelligence at War.*
Cf. ŠANTAVÝ, *Artificial Intelligence – A Good Servant or a Bad Master? Second, expanded edition*.

making to machines, the inability to control and monitor key steps in the processes occurring within these systems, and the blind growth of trust in the results of these agents' activities;

– efforts to achieve AGI (artificial general intelligence) will intensify. The arrival of AGI would represent an extreme shift in the capabilities of AI systems. Society is currently unable to absorb the paradigm shift that the arrival of AGI would entail, and these systems would be so out of control that we would not be able to use them safely at this time!

In this context, there is ongoing debate about *how to win a war that is controlled and influenced by artificial intelligence*. Such a war would be characterized by:

– a constant effort to utilize more sophisticated and better-trained AI systems;

– the ability to properly deploy and combine AI systems with conventional combat elements;

– the ability to operate AI systems at a high level of efficiency over the long term;

– the ability to integrate them operationally into operational and tactical plans;

– the ability to quickly and effectively deploy innovations that the adversary does not yet possess or cannot effectively counter;

– applying the "find-and-fool" method (deceiving AI systems and, thanks to AI, not being outmaneuvered).

In light of the above, the concept of "hyperwar" is also being redefined.[52] In anticipation of future conflicts driven and influenced by artificial intelligence, this means:

– a conflict driven by artificial intelligence and controlled by machines;

– unmatched speed enabled by automated decision-making and concurrent actions;

– human decision-making is almost entirely absent from the Observe-Orient-Decide-Act (OODA) loop.

The speed of AI systems, which human capabilities cannot match, the marked tendency to trust the results of artificial intelligence, and the extreme effort to hand over decision-making powers to machines pose a great danger to humanity and to the armed conflicts that will continue to accompany our civilization for a long time to come.

Despite the fact that these technologies are imperfect, prone to error, and merely simulate intelligence, with unmanaged consequences for humans and highly problematic systems in wartime deployment, we cannot stop their rise. However, this does not mean we should abandon our efforts to address these issues. The struggle for ethical principles and regulatory frameworks governing the use of AI technologies in military environments and armed conflicts is not over – and may, in fact, be just beginning...

---

[52] Cf. ALLEN, HUSAIN, *On Hyperwar.*

Bibliography

Alain SERVAES, *Iran Uses Chinese AI Satellite Imagery to Target U.S. Military Bases and Equipment in the Middle East,* Army Recognition Group, 2026. https://www.armyrecognition.com/news/army-news/2026/iran-uses-chinese-ai-satellite-imagery-to-target-u-s-bases-in-middle-east (April 8, 2026)

Bojan STOJKOVSKI, *US forces used Claude in Iran strikes for intelligence, targeting even after Trump's ban, Interesting Engineering*, 2026. https://interestingengineering.com/military/us-forces-used-claude-iran-strikes (March 11, 2026).

James DOBBINS, Raphael S. COHEN, Nathan CHANDLER et al., *Overextending and Unbalancing Russia: Assessing the Impact of Cost-Imposing Options,* Santa Monica, CA: RAND Corporation, 2019.

John R. ALLEN, Amir HUSAIN, *On Hyperwar,* in: Proceedings 143 (2017), U.S. Naval Institute.

Joel LEHMAN, Jeff CLUNE, Sebastian RISI, *An Anarchy of Methods: Current Trends*, in: *How Intelligence Is Abstracted in AI,* IEEE Intelligent Systems 29 (2014). pp. 56–62.

Kenneth ANDERSON, Matthew WAXMAN, *Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can,* Stanford University: Hoover Institution Press, 2013.

Melanie MITCHELL, *Artificial Intelligence,* Farrar, Straus and Giroux, 2019. ISBN: 978-0-374-71523-6.

Peter LAYTON, *Artificial Intelligence at War,* Australian Strategic Policy Institute, 2024. https://www.aspistrategist.org.au/artificial-intelligence-at-war/ (March 31, 2026).

Peter ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master?*, Bratislava: RKCMBF UK, 2023. ISBN 978-80-88696-91-9.

Peter ŠANTAVÝ, *Artificial Intelligence - Good Servant and Bad Master? Second, expanded edition*, Bratislava: RKCMBF UK, 2026. ISBN 978-80-88696-96-4.

Peter ŠANTAVÝ, Júlia KUBICOVÁ, *Artificial Intelligence Systems Do Not Possess True Intelligence, They Only Simulate It. They Do Not Distinguish Between Moral Good and Evil*, Bratislava: Nové mesto, 2023. ISSN: 2729-9597.

Sajid ALI, Tamer ABUHMED, Shaker EL-SAPPAGH, et al., *Explainable Artificial Intelligence (XAI): What We Know and What Remains to Be Done to Achieve Trustworthy Artificial Intelligence*, Information Fusion 99 (2023). ISSN: 1566-2535. https://doi.org/10.1016/j.inffus.2023.101805 (October 12, 2025).

Slavomír GEFFERT, *The Martens Clause in International Law of Armed Conflict,* in: *Selected Issues in Private and Public International Law,* BRATISLAVA: PF UK, 2021. ISBN: 978-80-7160-584-3

Sujita SINHA, *Iran using AI-enhanced satellite images from China to hit US bases in the Middle East,* Interesting Engineering, 2026. https://interestingengineering.com/military/iran-china-satellite-images-target-us-bases (April 8, 2026)

Will SHANKLIN, *“Pentagon’s pressure campaign,”* Engadget, 2026. https://www.engadget.com/ai/anthropic-weakens-its-safety-pledge-in-the-wake-of-the-pentagons-pressure-campaign-183436413.html (March 31, 2026).

Information from the Internet

Artificial Intelligence Act. https://www.consilium.europa.eu/en/policies/artificial-intelligence-act/ (March 30, 2026).

AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense. https://admin.govexec.com/media/dib_ai_principles_-_supporting_document_-_embargoed_copy_(oct_2019).pdf (March 30, 2026).

AI Risk Repository. https://airisk.mit.edu/ (March 20, 2026).

Anna JOBIN, Marcello IENCA, Effy VAYENA, *The global landscape of AI ethics guidelines.* https://doi.org/10.1038/s42256-019-0088-2 (March 30, 2026).

Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (March 30, 2026).

Marwa FATAFTA, Daniel LEUFER, *Artificial Genocidal Intelligence: how Israel is automating human rights abuses and war crimes*. https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/ (March 31, 2026).

Michael GABE, *Four Phases of AGI.* https://www.lesswrong.com/posts/qeJomTN2yp5tQG4rL/four-phases-of-agi (December 22, 2025).

Rome Call for AI Ethics. https://www.romecall.org/wp-content/uploads/2022/03/RomeCall_Paper_web.pdf (March 30, 2026).

UNESCO Recommendation on the Ethics of AI. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics (March 30, 2026).

Yuval ABRAHAM, *“Lavender”: The AI system directing Israel’s bombing campaign in Gaza*. https://www.972mag.com/lavender-ai-israeli-army-gaza/ (March 31, 2026).