European Integration Studies, Volume 20, Number 2 (2024), pp. 383-408.
https://doi.org/10.46941/2024.2.15

IZTOK PREZELJ[*]

Challenges in the Use of Artificial Intelligence-enabled Systems in Modern Armed Forces[**]

ABSTRACT: The application of Artificial Intelligence (AI) in the armed forces brings about a number of new possibilities and also new risks. In this paper, we have identified and analysed a wide range of risks associated with uncontrolled and unstoppable development of general AI, as well as several ethical and legal, operational and strategic risks. We have shown how and why these risks are dangerous and some even pose a threat to human security, values, norms, democracy, human rights, etc. These risks need to be carefully examined in order to improve the military use of AI and regulation in this area. The wide range of risks identified and their extremely diverse nature show that regulating the military use of AI will be difficult and complex, requiring all disciplines of law, and that regulatory rules need to be applied at national, regional and global level. The rapid development of military AI suggests that some risks are likely to be considered and regulated before any malicious military use of AI occurs, while unfortunately some others will only be regulated after the first instances of its malicious and illegal use by some armed forces.

KEYWORDS: artificial intelligence, risks, military, ethics, legal, strategic, operational, technology.

[*] Dean of the Faculty of Social Sciences, University of Ljubljana, Slovenia. iztok.prezelj@fdv.uni-lj.si.

[**] The research and preparation of this study was supported by the Central European Academy.

Introduction

Artificial intelligence (AI) is a relatively new computer technology that attempts to emulate complex human behaviour in some or all aspects, such as understanding or discovering meaning, linking information from different sources, recognising patterns, generalising, drawing conclusions, learning from experience, predicting, and adapting to changing circumstances. AI has become a technological source of the ongoing Revolution in Military Affairs[1] and a great hope for improving military capabilities or even redistributing the balance of military power on a global scale. The defence industries of richer, technologically more developed and more ambitious states are increasingly investing many resources in the development of new AI-enabled military capabilities in the areas of intelligence and surveillance, data-driven decision making or command and control, targeting, manoeuvring and other actions of military autonomous systems, cyber warfare and cyber security, logistics, training and exercises, etc. The application of AI in the armed forces brings with it a wide range of new opportunities on the one hand and many new risks and challenges on the other.

The aim of this paper is to identify and analyse some key challenges of the (potential) use of AI in modern armed forces. We argue that a responsible authorisation for the use of AI in armed forces and security services requires a thorough knowledge and investigation of the main application challenges in order to prevent various negative scenarios. Specifically, we argue that the main AI-related risks in this area are risks associated with the uncontrolled and unstoppable development of general AI, various ethical and legal risks, operational and strategic risks. The range of these risks is so wide that it will be difficult to address them

[1] The concept of the Revolution in Military Affairs refers to technological, organisational, structural, doctrinal, and operational profound, radical, discontinuous, non-incremental, and possibly disruptive changes. Four RMAs have been broadly discussed in the literature (see Thiele, 2021a: 65-69), such as RMA I (emerging from the second half of WWI in the form of combat vehicles), RMA II (based on the insurgent way of war in Asia), RMA III (focused around the use of nuclear weapons and other long-range means of delivery in the Cold War), and RMA IV (focusing on the digitalisation capabilities, including computers, precision-guided munitions, active and passive sensors, cyberspace, C4 and robotics. RMA V is the next RMA that will be brought by new technologies.

comprehensively. In the process of research, we used comprehensive literature review, case analysis, risk identification and synthesis.

The discussion about the use of AI usually narrows down to a debate between the proponents of AI, who emphasise the positive benefits, such as faster operations, reduction of own casualties, risk of errors, etc., and the opponents of AI, who emphasise the disadvantages, such as possible unintended consequences, the risk of violating existing laws and norms or even the supremacy of AI over humans. Any warnings from opponents or “doomers” should be carefully considered and used to better regulate the use of AI by armed and security forces.

1. General challenges of the difficult implementation of artificial intelligence in armed forces

The implementation of AI in the armed forces will be slower than expected, but still relatively fast. The implementation of previous RMAs has encountered some reality tests, and we foresee a similar outcome for the AI aspect. According to Horrowitz, current progress in integrating AI into military systems has been only incremental, and organisations are struggling to make the leap from development to operational implementation. Debates about the development of AI technology reveal a high degree of uncertainty about the potential pace of progress in AI. Modern armed forces face technological and organisational obstacles to the effective use of AI. The technological challenges can be divided into two broad categories: internal reliability problems and external exploitation problems. On the one hand, internal problems relate to the enormous complexity of the modern battlefield, to which AI narrow systems cannot adapt, which can lead to accidents and errors. Reliability and trust will play a crucial role in opening up the armed forces to the use of AI. Practice has shown that AI systems can sometimes exhibit uncertain behaviour and this might not be tolerated by most armed forces. On the other hand, external problems could be the adversarial data problem or the problem of attempts of the enemy to poison the data. Before being introduced to operational use, armed forces will want AI systems to be noticeably better than existing systems. The armed forces will also have to weigh up capabilities and reliability against the risks It seems that forces facing defeat will be more willing to take the risks of using AI and vice versa.[2]

Another implementation problem is that AI will not be integrated into military systems and platforms at the same time or with the same effectiveness or efficiency. The idea that AI will automatically supplement existing dysfunctional security systems and bring a new form of objectivity is wishful thinking. In military operations, the dependence on AI must be carefully calibrated. Armed forces will have to decide to what extent and how quickly historically evolved organisational structures and doctrines should be replaced by new, technology-centric concepts.[3]

Human absorption barriers will also play an important role. Studies on the use of AI in the civilian environment show that personal values and attitudes strongly affect readiness to use AI and trust in AI. They show that extroverted people often have negative feelings towards AI, agreeable people see it as positive and useful, neurotic people experience negative emotions but perceive AI as socially friendly, conscientious people as useful but less socially friendly, while open-minded people as very useful.[4] Other studies have shown that more trusting people (within other people) tend to trust AI more than less trusting people[5] and that geographical location and even religious orientation influence the trust in or fear of AI (e.g. respondents from East Asia are less afraid of AI than Europeans, Muslims and Buddhists are more afraid of AI).[6]

An example of implementation problems can be found in the US, where the National Security Commission on Artificial Intelligence expressed frustration with the level of AI-readiness in the US security administration, acknowledging that the integration of AI in all sectors is difficult due to some unique challenges. One of the most significant challenges and impediments for AI development is the holy grail of rare talent that will enable AI breakthroughs. Accordingly, there is a deficit of human talent in the U.S. government. New talent pipelines need to be built, such as the new Digital Service Academy and the civilian National Reserve, to grow talent with the same seriousness as military officers. The US has

[2] Horowitz, 2018, pp. 5-6.

[3] Mashur, 2019, p. 4.

[4] Park and Woo, 2022.

[5] Schepman and Rodway, 2023.

[6] Mantello et al., 2023. The authors of this study published in the journal AI & Society reached this conclusion based on the survey of 1.015 responses of future job-seekers from 48 countries.

also noted that some of its agencies have made great strides in adopting AI, putting them ahead of other agencies. The commission's report also stressed that in this situation, it is not time for incremental changes, such as increasing budgets and creating a few new positions at the Pentagon and Silicon Valley, but that it is time to fundamentally change the mindset.[7]

Finally, the armed forces and defence institutions still do not have sufficient amount of big data to adequately train AI models. At the 'NATO in the Nordics' conference, it was highlighted that in one exercise, 26 platoons were monitored by numerous sensors (locations, communications, etc.) continually and this was only a fraction of the data necessary. There is a great need to collect more data from existing military exercises. AI is currently more of a training object and not a serious tool.[8]

As with all promising game changing technologies in RMAs, AI will be slowly introduced in the armed forces, but also much faster than other new technologies in the past. An evolutionary approach with some leaps is to be expected instead of a real revolution.

2. A broad spectrum of challenges in the use of artificial intelligence by the armed forces

The existing literature increasingly addresses a wide range of risks and concerns about the use of AI. For example, Encyclopedia Britannica lists the following ethical and socio-economic risks of AI: increased unemployment for certain job profiles (although AI will create certain new jobs), ingrained social biases (gender bias, racial bias, etc.), privacy risks (large amounts of data can be accessed by unauthorised organisations and people), and the risk of manipulation of images, creation of fake profiles, etc.[9] In this paper, we are interested in the risks posed by the use of AI in the defence and military sectors.

Various categories of observers have warned against the use of AI in general and in the military sphere. Firstly, in an open letter in 2015 groups of scientists and technologists, for example, warned against the AI arms race and the potential spread of lethal AI to terrorists and dictators. The letter also called for a ban on offensive autonomous weapons beyond meaningful human control. Secondly, groups of employees at technological companies

[7] Final Report, 2021, p. 3, 8, 110.

[8] Schuller, 2023.

[9] Artificial Intelligence, 2023.

called against the production of weaponised robots and similar warfare technology. Google[10] consequently published its AI guiding principles, in which it pledged not to design or deploy weapons that cause injury to people, technologies that gather or use information for surveillance violating internationally accepted norms, and not to develop technologies that violate generally accepted international law and human rights, despite its continued cooperation with the US government. International campaigns such as the International Campaign for Robot Arms Control and the Campaign to Stop Killer Robots were launched in 2009 and 2013 to mobilise nation states, the public and the industry. Several faith and interfaith declarations against autonomous weapons have been adopted, including one by the Catholic Church stating that it is fundamentally immoral to use a weapon that we cannot fully control. Finally, a discussion was initiated on a possible new formal protocol to the UN Convention on Certain Conventional Weapons to improve the regulation of fully autonomous weapons, but the US, Russia and the UK objected.[11] Some observers labelled AI as a new weapon of mass destruction, as they saw similarities with the development phase of the atomic bomb. Some actors, such as the Austrian and Swedish governments, the Belgian Parliament and the European Parliament, also called for a ban on autonomous weapons.[12]

The above discussion on the difficult implementation of AI in the armed forces shows that the potential premature introduction of AI systems and technology in military practice is a matter of concern. Many of the risks associated with the use of this technology stem from this problem. These risks and concerns need to be taken seriously and regulated as much as possible to avoid undesirable consequences in any way. We categorised several clusters of risks from the existing literature (see Table 1).

[10] Google cooperated with the Pentagon in project Maven aiming to use special computer vision technology for analysing an increasing number of drone footage and identify and track objects. Google employees protested against this in 2018 and the contract was not continued (Canca, 2023, p. 60).

[11] Forrest et al., 2020, pp. 24-28.

[12] Soare, 2023, pp. 100-102.

Table 1 *Categorisation of risks of AI use by armed and defence forces.*[13]

General categories of risks:	Specific categories of risks:
1. Uncontrolled and unstoppable development of general AI	Exceeding human performance
	Self-directed, self-replicating and self-improving beyond human control
	Pursuing objectives that are not consistent with human interests
2. Ethical and legal risks	Limited AI capacity to understand the law of armed conflict, humanitarian law and other legal basis
	Accountability gap between the operators and AI systems
	Limited ability to make moral judgements
	Tendency to violate human rights and privacy (threat to privacy and human rights)
3. Operational risks	The issue of overconfidence in AI systems and the problem of surprising and incomprehensible decisions
	Problematic validity of AI-based recommendations or decisions
	AI outcomes and decisions based on narrow training experience
	The risk of accidental use and conflict escalation
	Vulnerabilities of AI systems
	Lower use and violence thresholds

[13] The base for this categorisation was the classification by Forrest et al., which was then supplemented with other debated risks and published sources. The original categorisation by Forrest et al. (2020, p. 30) includes:
- Ethical and legal risks: law of armed conflicts, accountability and moral responsibility, human dignity, and human rights and privacy;
- Operational risks: trust and reliability, hacking, data poisoning and adversarial attacks, accidents and emergent risks;
- Strategic risks: thresholds, escalation management, proliferation, and strategic stability.

4. Strategic risks	The risk of easy proliferation to other malicious states, criminal and terrorist actors
	Risky and difficult to control the dual-use potential of AI technology
	The risk of global AI arms race and competition
	AI capability-related distrust among countries
	Risk of system mispositioning of AI-based decision-making
	The risk of increased police and intelligence states

2.1 Uncontrolled and unstoppable development of general artificial intelligence

The first concern relates to the worst-case scenario in terms of the potentially uncontrolled and unstoppable development of general AI. AI can usually be divided into artificial general intelligence (AGI) or strong AI and applied AI. The ultimate goal of AGI is to build machines that think and whose general intellectual abilities are indistinguishable from those of humans. After great optimism in the 1950s and 1960s, science has realised that this involves extreme difficulties. Applied AI, on the other hand, is about advanced information processing aimed at developing commercially viable and more targeted 'smart' systems. The application of such 'expert systems' has been much more successful in practice. Such systems are based on a knowledge base and an inference engine. The latter processes information on the basis of production rules (if-then rules, etc.). Good expert systems are often better than a single human expert, and their scope of application can be very broad.[14] At present, our society is at the level of a weak or narrow AI, where the systems can only perform very specific tasks.[15] However, the risk associated with AGI remains, as it is uncertain at what point AGI will be able to exceed human performance for a given task. There is also a risk that AGI could become self-directed, self-replicating and self-improving and escape human control. In addition, such AI systems will become larger, better, cheaper and more ubiquitous. They will be capable of quasi-autonomy and potentially self-improvement. Each of these features

[14] Artificial Intelligence, 2023.

[15] Luberisse, 2023a, p.3.

will challenge traditional governance models.[16] At some point, these systems will be weaponised by nations and their armed forces, and defending against them will be the task of the armed forces of the conflicting countries. Furthermore, the fictional scenario is that human-made system surpasses human intelligence and pursues goals that do not coincide with human interests, thus posing an existential threat to all humans. The worst-case scenario in this direction would be the dominance of AI systems and some kind of conflict between human society and AI systems or even a new AI civilisation or human slavery. Such scenarios have been clearly simulated by the movie industry in some widely known movies, such as The Matrix. This was about a human society enslaved by AI, where people were bred in fields as batteries for technical systems and platforms. The Matrix was actually a special AR environment where people performed specific social roles, all for the purpose of keeping their minds happy so that the batteries (their bodies) in the real physical world grew at the right pace and could be harvested for consumption. Another such scenario is the case of Skynet, an artificial consciousness that controls the Terminator robots in the movie Terminator. The AI system in one of the Terminator movies asserted: 'I am not a machine, I am not a man, I am more'.[17]

Juliano further developed the possible negative scenario referred to above. The defining characteristic of a strong AI is the capacity to generalise, i.e. the ability to adapt to and act in new environments without being programmed to do so. Generalising intelligence will need to develop the ability to feel and understand consciousness. Juliano believes that we will ultimately be powerless to stop the release and future misuse of strong AI, and that it is unlikely that we will change enough to deal responsibly with strong AI. In his view, it is dangerous to believe that we, as a species, will not lose control after the first strong AI is liberated and distributed.[18] Accordingly, we do not have a choice because not everyone will agree to limiting research, research can be conducted secretly regardless of legality, strong AI is algorithmic by nature and does not require significant resources or infrastructure to research it, and overlapping fields of research are converging in this direction (research in linguistics, mathematics, computer science, cognitive science, neuroscience, philosophy of mind, etc.). The

[16] Bremmer and Suleyman, 2023, pp. 6-7.
[17] Terminator Genesis, 2015.
[18] Juliano, 2016, pp. 7-13.

threat will initially be coming from those individuals or groups who are the first to use strong AI, rather than from the AI itself, but later on ordinary people, including criminals and terrorists, will also gain access to strong AI with even the most basic computers. The threat will mainly come from force multiplication effects.[19]

In their RAND study, Forrest et al. [20] called for a deeper examination of the risks associated with AI and conducted an expert opinion survey on the risks associated with military AI applications. The top 5 AI risks of military AI applications were as follows: decisions might be made too fast, they could result in increased escalation, they could be less accurate/precise than humans, it is difficult to differentiate combatants from non-combatants, and it is difficult to differentiate anomaly from threats.

2.2 Ethical and legal risks

Limited AI capacity to understand the law of armed conflict, international humanitarian law and other legal basis. The law of armed conflict and international humanitarian law are based on the four Geneva Conventions and their protocols.[21] Accordingly, belligerents must comply with the three most important principles: distinction (between civilians and combatants, operations must be directed at military objectives and attacks against civilian targets must be omitted), proportionality (no excessive harm disproportionate to the military objective) and precaution or military necessity (use of only necessary force to achieve a legitimate military objective).

The main criticism of fully autonomous weapon systems focuses on their alleged inability to comply with the principles of distinction and proportionality. They argue that these systems are unable to understand and assess subtle differences between combatants and non-combatants, especially in urban settings where combatants do not always wear uniforms. They are also unable to comply with the principle of proportionality, as this requires a case-by-case assessment of possible collateral damage weighed against the importance of the military objective.[22] If these systems are able to distinguish between military and civilian targets, the question arises as to

[19] Juliano, 2016 pp. 163-209.

[20] Forrest et al., 2020, p. 21.

[21] See The Geneva Conventions of 12 August 1949, 1949; Protocols Additional to the Geneva Conventions of 12 August 1949, 1977.

[22] Forrest et al., 2020, pp. 30-31.

how accurate they can be, whether they can assess the proportionality of the use of force and comply with international law.[23]

Despite the fact that humans proved themselves as extremely efficient in ways of slaughter, there is a growing concern (it is even a key concern) about how deadly these AI systems could be and whether they can run amok and cause humans to lose control. It is unlikely that AI systems with a very narrow view of the world would be able to navigate and fight on their own in a very challenging urban combat environment. The laws of armed conflict could be integrated into the software, but the question is whether these 'killer robots' would be able to understand and apply them. This means that there is a risk that AI systems could be used to carry out illegal and unethical actions.[24]

Accountability gap between the operators and AI systems. The ethical risk is that the use of autonomous weapon systems will create an accountability gap or moral buffer between human operators and the actions of the systems. Accountability is an important moral concept that designates moral responsibility for actions and the associated moral emotions, such as shame or guilt. This concept is an important deterrent in war and in general. Critics claim that fully autonomous weapons will make decisions without proper accountability and that systems cannot be held morally responsible for their actions. This brings us to a specific problem of attribution, where it is not clear who is responsible for the use of the system.[25] The issue of accountability is one of the most important ethical considerations in relation to autonomous weapons. The question is who is accountable if an autonomous weapon malfunctions or makes a decision that causes civilian casualties, and is it ethical to hold the programmers, the military or the government accountable.[26]

Limited ability to make moral judgements. Arguments from the perspective of human dignity claim that only humans are capable of making moral judgments about the taking of human life, and that only humans have emotions and a sense of compassion and respect for human life. Technical systems do not have sufficient moral qualities to justify their actions in a way that respects the victims and therefore should not make such

[23] Luberisse, 2023b, p. 60.

[24] Luberisse, 2023a, pp. 19-20.

[25] Forrest et al., 2020, pp. 32-33.

[26] Luberisse, 2023b, p. 61.

decisions.[27] The question is whether the use of autonomous weapons is consistent with the principles of a just war.[28]

Tendency to violate human rights and privacy (threat to privacy and human rights). AI brings threats and risks to human rights and the privacy of individuals. AI systems require vast amounts of data, leading to concerns that this data could be used to violate individual rights. For example, the massive use of AI data in facial recognition raises concerns about possible misuse by governments and other organisations.[29] Autocratic surveillance of one's own population can be made possible by systems such as extensive data analysis, persistent ISR, facial recognition, the Internet of Things, etc.[30] Information operations that spread false information and create social and cognitive bias lead to the diminished importance of objective facts (e.g. Truth Decay). Military systems can produce outputs that discriminate against minorities or other groups due to unrepresentative and biased training data.[31] For example, algorithms trained on biased data can perpetuate discrimination against marginalised groups, leading to further marginalisation and human rights violations. This way, AI can perpetuate and exacerbate prejudices and inequalities.[32] This susceptibility to bias in the data actually means that even machine learning cannot guarantee the absence of bias or analytical error.[33]

The use of AI, especially in lethal autonomous weapons and decision support tools in active combat, may lead to ignoring the complexity of the given situation and the value of human life. The greatest risk is the potential incorporation of an ethical error in AI system because its widespread use can lead to mass damage to individuals and communities, behind the veil of computational objectivity.[34]

[27] Forrest et al., 2020, p. 34.

[28] Luberisse, 2023b, p. 61.

[29] Luberisse, 2023a, pp. 38-40.

[30] Frequently, the use of AI by China indicates excessive monitoring of own citizens and suppressing dissent. However, such approaches were also used in more Western societies against own population as indicated for example by the Snowden case.

[31] Forrest et al., 2020, p. 35.

[32] Luberisse, 2023a, pp. 38-40.

[33] Mashur, 2019, p. 2; see also Rickli and Mantellassi, 2023, p.18.

[34] Canca, 2023, p. 59.

2.3 Operational risks

The issue of overconfidence in AI systems and the problem of surprising and incomprehensible decisions. The black box problem of AI refers to the inability to explain the reasoning that led to a particular outcome. Such situations would lead to an increasing ‘unawareness’ of what is happening on the battlefield[35] and to the problem of trust and reliability (mainly expressed in the issue of not trusting or overtrusting AI systems). The black box problem refers to the situation in which an AI system might produce outputs in ways not comprehensible or explainable to humans. Different performances of the AI system outside the laboratory can also lead to an additional lack of trust. On the other hand, operators or commanders might have excessive trust in AI systems because they are overconfident, do not look for contradictory information, etc. Such tendencies were observed in Operation Iraqi Freedom, where some operators trusted the systems without questioning.[36] The victory of the AI programme AlphaGo over the human world champion and grandmaster in 2016 was achieved through occasionally surprisingly bold moves that ultimately led to a shocking defeat of the human opponent.[37] If AI systems function in unpredictable ways that can have serious negative consequences, responsible leaders will not adopt them, and operators will not have confidence in their use and will not deploy them. There is also a risk that autonomous AI systems would be used for human rights violations and war crimes.[38]

Problematic validity of AI-based recommendations or decisions. Occasionally, it will be impossible to verify the validity of AI-based recommendations. It is difficult to judge from an external point of view how accurate or trustworthy an AI-generated assessment really is. More complex AI may be able to predict or at least pre-define scenarios without necessarily understanding the underlying logic, reasoning and prioritisation. This means that it is very important how AI is embedded in a political and institutional context to minimise serious risks.[39]

AI outcomes and decisions based on narrow training experience. AI systems must first be trained in an artificial environment with different data sets. The system processes the data, performs the tasks and hopefully learns.

[35] Rickli and Mantellassi, 2023, p.63.

[36] Forrest et al., 2020, p. 36.

[37] Gatopoulos, 2021, p. 5.

[38] Luberisse, 2023b, p. 61.

[39] Mashur, 2019, p. 2.

The catch is a lengthy accumulation of experience based on a large number of interactions and repetitions with different data sets. Training with one data set leads to certain results, while training with another data set leads to different results. Mashur emphasised that AI systems trained in different ways might come to conflicting conclusions. This means that AI systems are not able to achieve results based on perfect rationality.[40]

The risk of accidental use and conflict escalation. The risk of accidental deployment and use with unintended consequences is real. AI-enabled autonomous weapons, if deployed globally in an uncontrolled manner, could increase the risk of unintended conflict escalation and crisis instability.[41] The programmers of AI are not so much worried about the Terminator scenario, but rather about flash wars (wars that are triggered without control, similar to the collapse of the stock market, where many algorithms are trading and suddenly, due to an unforeseen event, the algorithms crash the stock market).[42] These concerns are particularly present in the area of autonomous nuclear defence systems. The risks of accidental use in this area or potential use by malicious actors (who would hack into the system or feed false data) can be globally deadly. The speed of AI-powered decision making could even lead to an escalation of conflict, resulting in a rapid and unintended escalation in the use of nuclear weapons. AI can accelerate the decision-making process in crises to a machine AI level.[43] Future Cuban missile type crises might emerge, but the problem is that this acceleration could contribute to escalating the crisis rather than de-escalating it, as the actors would see their window of opportunity shrinking.[44] The existence of the Russian Perimeter nuclear defence system has also raised concerns about the ethical implications of granting decision making capabilities to machines and the risk of accidental use.[45]

[40] Mashur, 2019, p. 2.

[41] Final Report, 2021, p. 10.

[42] Flash Wars: Autonomous Weapons, AI and the Future of Armed Conflict, 2023.

[43] Director of the US AI Center stated that that we are going to be shocked by the speed, chaos and bloodiness in the future wars, it is going to be algorithm against algorithm (Rickli and Mantellassi, 2023, p. 20).

[44] Mashur, 2019, p. 2.

[45] An AI-enabled example is the Russian nuclear automated defence system Perimeter, which can detect a nuclear strike against Russia and launch a retaliatory nuclear strike even if the lines of communication with Strategic Missile Forces are destroyed. The system adopts a decision to launch a retaliatory strike after approval by the human commander, but in case of a missing communication with the command centre it can launch such a strike alone. Additionally, it can launch a command rocket in the air over Russia and retaliatory

Due to the associated combination of massive damage and lack of controllability, there have been calls to consider an international ban on lethal autonomous weapon systems and to classify intelligent AI-supported drone swarms as weapons of mass destruction.[46]

Vulnerabilities of AI systems. AI systems are also vulnerable to hacking, data poisoning and adversarial attacks. AI systems can be hacked and their training data manipulated or spoofed in order to influence the intended functioning of the system. Attacks by adversaries might also trick algorithms into making a mistake. AI software can also escape seemingly unintentionally, such as the Stuxnet worm and other cases of self-replicating malware (WannaCry, NotPetya). Finally, AI systems could become so advanced that they could undermine the 'second strike' capabilities that are essential for responding after an initial nuclear attack. AI could be used to locate enemy nuclear launchers, disable them during the attack and prevent a retaliatory strike.[47]

Since AI-powered organisations will store large amounts of sensitive data, the risk of data breaches and information theft in AI-powered organisations is real.[48] The adversarial AI will aim also to deceive the AI with deceptive data.[49] The possibility that one's entire army of AI systems can suddenly turn against their owners is also terrifying for military planners.[50] In addition, even high-performance algorithms are not immune to being misled by more traditional means of espionage and deception. AI might mistakenly assess certain patterns of behaviour as harmless if they occur often enough without any feared consequences.[51]

2.4 Strategic risks

Lower use and violence thresholds. It is likely that the use of AI will shift the balance between offence and defence towards offence: AI will largely be

strike activation from all available platforms (silos, aircraft, submarines and mobile ground units) is done from there in case of missing link with strategic missile control centre. Perimeter checks this link all the time, but it can act autonomously in case of need. Another example is the Russian fully automated nuclear submarine Poseidon, which can also autonomously generate a nuclear attack. (Luberisse, 2023a, pp.21-23).

[46] Hambling cited in Nurkin, 2023, p. 52.

[47] Forrest et al., 2020, pp. 37-38.

[48] Luberisse, 2023a, p.18.

[49] Rickli and Mantellassi, 2023, p. 16.

[50] Gatopoulos, 2021, p. 10.

[51] Mashur, 2019, p. 2.

used offensively.[52] There is also a risk that the threshold for the use of autonomous armed systems is lower than the threshold for the use of conventional weapons. This faster use could also cause more civilian casualties during operations.[53] Schmidt et al. even fear that all AI tools will be among the weapons of first choice in future conflicts.[54]

The risk of easy proliferation to other malicious states, criminal and terrorist individual or collective actors. AI systems are not only much easier to develop, steal and copy than nuclear weapons, they are also controlled by private companies and not by governments.[55] Egel emphasised that AI-enabled weapons are relatively easy and inexpensive to procure and will therefore be accessible to non-state actors and proxies. Some states could even deliberately provide such actors with these capabilities, as has happened in the past.[56] Thiele concluded that AI technologies will sooner or later be available to any opponent.[57]

Risky and difficult to control dual-use potential of AI technology. As a rule, non-combat AI systems (used in the areas of predictive maintenance, logistics, personnel management, communication, etc.) are not ethically problematic. However, the literature warns that existing AI systems can be reprogrammed for use on the battlefield.[58] This leads us to the typical area of dual-use technology. For example, an AI algorithm for driving cars can easily be adapted to an algorithm for driving tanks and so on. This means that the boundaries between the safely civilian domain and the destructive military domain are inherently blurred.[59]

The risk of global AI arms race and competition. The AI empowerment is a very attractive option in the global power struggle. Authors who have studied the geopolitical aspects of the use of AI emphasise that the race to adopt AI is leading to a power struggle between great powers with implications for the global balance of power.[60] Bremmer and Suleyman also emphasised that AI supremacy, or competition for AI supremacy, will be a strategic objective of every government that has the

[52] Rickli and Mantellassi, 2023, p. 25.
[53] Forrest et al., 2020, p. 39.
[54] Schmidt et al, 2021, cited in Thiele, 2021b, p. 76.
[55] Bremmer and Suleyman, 2023, p. 10.
[56] Egel et al., 2019, cited in Thiele, 2021a, p. 77.
[57] Thiele, 2021b, p. 190.
[58] Canca, 2023, p. 60.
[59] Bremmer and Suleyman, 2023, p. 6.
[60] Luberisse, 2023a, p. 18.

resources. Two major players, the US and China, view AI development as a zero-sum game that will give the winner a decisive strategic edge in the future.[61] Nations and organisations that are best to anticipate and exploit technological opportunities are likely to have a decisive advantage in future competitions, crises and conflicts. AI will also be the linchpin in achieving military superiority through the use of data, i.e. turning it into relevant information, usable knowledge and ultimately into decision-making advantages.[62]

AI capability-related distrust among countries. The lesson from the classic confidence- and security-building measures is that distrust leads to conflicts and that distrust can be based on a lack of information about the capabilities of the opponent.[63] We argue that AI development and use in modern armed forces will lead to the typical distrust among states that has already been observed in the past in delicate geostrategic situations with a lack of information about the capabilities of the opponent. Horrowitz also emphasised that the state's armament in the AI-related capabilities can hardly be measured precisely by other states. It will be difficult to assess the degree of automation, the quality of the code, the efficiency of autonomous weapons and their capabilities. This uncertainty will lead states to overestimate the capabilities of other states.[64]

The risk of system mispositioning of AI-based decision-making. A very important question for society is who exactly has access to AI and who is in the position to contextualise and interpret the results. In democracies, the armed forces' sole access to analytical AI that recommends certain military options for action may be problematic. Especially at the highest strategic levels, where other defence and political actors should also be involved. It is important how and where AI is embedded in the existing institutional decision-making process,[65] otherwise AI could be used strategically based on a narrow military perception of the situation.

The risk of increased police and intelligence state through the use of AI. AI surveillance systems can be used for systematic, excessive surveillance of one's own or other people's populations. The exposure of widespread illegal HUMINT or TECHINT collection operations typically

[61] Bremmer and Suleyman, 2023, pp. 7-8.

[62] Thiele, 2021a, p. 59, 77.

[63] See Prezelj and Harangozo, 2018.

[64] Horrowitz, 2018, cited in Rickli and Mantellassi, 2023, p. 25.

[65] Mashur, 2019, p. 2.

led to the so-called intelligence collection scandals.[66] The application of AI in this area will improve operational capabilities and give legal or rogue actors more opportunities to infringe the human rights of a large part of the population. The classic concept of a police or intelligence state can transform itself into an AI police and intelligence state. This risk is also recognised in the policy world, but much more in case of foreign states than for the domestic state. For example, according to US sources, [67] the U.S. is very concerned about China's use of AI as a tool of repression and surveillance both internally and gradually internationally. Accordingly, AI should reinforce democracy rather than erode it. AI future should be democratic, AI must be developed based on its values and work with democracies and the private sector is essential in building privacy-protecting standards into AI technologies and advancing democratic norms to guide AI use so that democracies can use AI for national security purposes.[68] Luberisse stressed that China has been investing heavily in AI, with a particular focus on surveillance systems to enhance its ability to monitor and control its population. The nationwide deployment of AI-powered cameras and facial recognition systems has raised significant privacy and human rights concerns and fuelled debates about the appropriate use of AI.[69] However, we should also be wary of similar intentions in democratic states. Several public intelligence scandals teach us to think along these lines too.

3. Conclusion

The application of AI in the armed forces brings with it a range of new opportunities as well as many new risks and challenges. In this paper, we have identified and analysed a wide range of risks associated with an uncontrolled and unstoppable development of general AI, along with several ethical and legal, operational and strategic risks. We have shown how and why these risks are dangerous and some even pose a threat to human security, values, norms, democracy, human rights, etc. These risks need to be carefully examined in order to improve the military use of AI and regulation in this area.

[66] Prezelj and Ristevska, 2023.

[67] Final Report, 2021, pp. 2-6.

[68] Final Report, 2021, pp. 2-6.

[69] Luberisse, 2023a, pp. 10-11.

The introduction of AI in modern armed forces will be complicated and slower than expected, but still faster than the introduction of previous new technologies. The armed forces will have to carefully weigh reliability and controllability, on the one hand, against the related risks on the other. They will have to deal with several technological and organisational barriers to reach an effective AI, as AI will be implemented asymmetrically in different weapon systems, and human absorption barriers have not yet been sufficiently addressed. The latter will be an important factor in the adoption of this technology, as there are already scientifically verified patterns of potential negative feelings, anxiety and distrust towards the new technology. The armed forces will also have to deal with the problem of the deficit of personnel specialising in AI who are willing to work for them. The introduction of AI in the armed forces will also require some legal, ethical, organisational, doctrinal, strategic and policy changes in the military and defence systems and beyond.

Several categories of actors from the international community have warned about the risks of development and use of AI. Particular attention has been paid to general AI and military autonomous weapons systems. The warnings have come from groups of scientists, technologists, technology company employees, activists and even the Catholic Church. Some have even labelled AI as a future weapon of mass destruction, as there are some similarities in the early stages of development of both technologies (nuclear and AI).

The most serious, but still very hypothetical and potentially existential risk comes from the unstoppable and uncontrolled development of general AI in a direction that is not consistent with the general human interest. We do not know when this may happen. Some authors are of the opinion that it will be inevitable, and when it happens, it will be too late. Existing movies offer several imaginary scenarios for such a possible future. The ethical and legal risk category includes the risk of the limited understanding of the law by the AI systems and the related concepts of proportionality, distinction and military necessity, the risk that the autonomous systems will not be able to take accountability for military actions, the limited ability to make moral judgments, and the tendency to violate human rights and privacy. The category of operational risks includes the risk of excessive trust in AI systems and the problem of occasionally surprising and incomprehensible AI decisions, the problematic validity of AI-based recommendations and decisions, the relatively limited training experience that determines the results of AI systems, the risk of accidental use and conflict escalation and, finally, the vulnerability of the AI systems themselves. The category of strategic risks includes the risk of lower use and violence thresholds, the ease of dissemination to other malicious states and criminal and terrorist actors, the risk of the dual-use of AI, the risk of a global AI arms race and competition, the risk of distrust among states regarding actual AI capabilities, the risk of incorrect positioning of AI-based decision making in the system, and the risk of creating a police and intelligence state.

These risks need to be carefully examined and incorporated into future regulatory systems at national, regional and global level. The range of risks mentioned above is so wide that regulation will be very difficult. It is likely that some risks will be taken into consideration and clearly regulated before there is any malicious military use of AI. However, there will certainly be some uses of AI for military purposes where regulation will only follow after the malicious use of the technology. Unfortunately, this will not happen for the first time in human history.

Finally, the question arises as to what more concrete countermeasure strategies and practical guidelines should be applied to manage the risks associated with the use of artificial intelligence in the armed forces. We recommend the following countermeasures to address the identified risks:

1. Control the development of general AI by monitoring at what point it will be able to outperform humans, when it will become self-directed, self-replicating and self-improving, and when it will escape human control in the wrong direction by pursuing goals against humankind. Furthermore, the research process, even open coded, must somehow be limited.
2. The ability of AI to ‘understand’ the law of armed conflict, international humanitarian law and other legal frameworks must be constantly improved.
3. The accountability of operators and AI systems must be regulated. It should be made clear that the actions of AI systems are legally attributable to their operators and creators.
4. Due to the limited ability of AI systems to make moral judgments, moral responsibility should be assigned to their human AI operators.
5. Understand that autonomous AI systems deployed in all security domains are prone to violate human rights and privacy and prepare appropriate barriers to do so.

6. Educate AI operators about the problem of overconfidence and the ‘black box’ in order to maintain a certain critical distance from AI systems.
7. Stop the operation of AI systems in cases where they make completely surprising and incomprehensible decisions and try to understand them.
8. Try to verify the validity of AI-based recommendations or decisions.
9. Since AI results and decisions are based on narrow training experiences, AI should not be used in situations for which it has not been prepared.
10. Be aware that one of the main risks is the danger of accidental use and conflict escalation; try to simulate and predict such situations and use blockers for such a development.
11. Recognise vulnerabilities of AI systems and try to mitigate them.
12. Try to monitor violence thresholds when using AI systems.
13. Seek to create a non-proliferation regime for AI weapons that includes state and non-state actors.
14. Understand AI as a dual-use technology and seek to regulate it like other such technologies.
15. Create a confidence- and security-building regime that controls existing AI weapons capabilities in all states based on self-reporting, monitoring and verification.
16. Learn at which level which AI-based decisions should be made.
17. Mitigate the risks of a growing police and intelligence state through the use of AI by controlling AI operators, masters and related structures by means of democratic oversight.

Bibliography

[1] Bremmer, I., Suleyman, M. (2023) ‘The AI Power Paradox: Can States learn to Govern Artificial Intelligence – Before It’s Too Late?’, *Foreign Affairs*, 2023/September/October.

[2] Canca, C. (2023) ‘AI Ethics and Governance in Defence Innovation: Implementing AI Ethics Framework’ in Raska, M., Bitzinger, R. A. (eds.) *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities and Trajectories*. New York: Routledge, pp. 1–11; https://doi.org/10.4324/9781003218326-4.

[3] Forrest, E. M., Boudreaux, B., Lohn, A.J., Ashby, M., Curriden, C., Klima, K., Grossman, D. (2020) *Military Application of Artificial Intelligence: Ethical Concerns in an Uncertain World*. Santa Monica: RAND Research Report.

[4] Gatopoulos, A. (2021) ‘Project Force: AI and the Military – a Friend or Foe?’, *Aljazeera* [Online]. Available at: https://www.aljazeera.com/features/2021/3/28/friend-or-foe-artificial-intelligence-and-the-military (Accessed: 09 August 2024).

[5] Horowitz, M. C. (2018) ‘The Promise and Peril of Military Applications of Artificial Intelligence’, *Bulletin of the Atomic Scientists* [Online]. Available at: https://thebulletin.org/2018/04/the-promise-and-peril-of-military-applications-of-artificial-intelligence/ (Accessed: 06 August 2024).

[6] Juliano, D. (2016) *AI Security.* Fort Myers: Undine.

[7] Kruger, A. (2024) ‘Alternative ni, prilagoditi se bomo morali svetu z AI’, Interview, Executive Director of DFKI, *Delo,* 22 February, p. 13.

[8] Levy, A., Uri, M. (1986) *Organisational Transformation: Approaches, Strategies, Theories*. New York: Praeger. https://doi.org/10.5040/9798400693960.

[9] Luberisse, J. (2023a) *The Geopolitics of Artificial Intelligence: Strategic Implications of AI for Global Security*. Wroclaw: Fortis Novum Mundum.

[10] Luberise, J. (2023b) Algorithmic Warfare: The Rise of Autonomous Weapons. Wroclaw: Fortis Novum Mundum.

[11] Mantello, P., Manh-Tung H., Minh-Hoang N., Quan-Hoang V. (2023) Bosses without a heart: Socio-demographic and cross-cultural determinants of attitude toward Emotional AI in the workplace. *AI & Society,* 38, pp. 97–119; https://doi.org/10.1007/s00146-021-01290-1.

[12] Mashur, N. (2019) 'AI in Military Enabling Applications', *CSS Analyses in Security policy*, 2019/251, pp. 1–4. [Online]. Available at: https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/367663/CSSAnalyse251-EN.pdf?sequence=2 (Accessed: 09 August 2024).

[13] Nurkin, T. (2023) 'AI and Technological Convergence: Catalysts for Abounding National Security Risks in the Post-COVID World', in Bitzinger, A. R., Raska, M. (eds.) *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities and Trajectories*. New York: Routledge, pp. 37-58; https://doi.org/10.4324/9781003218326-3.

[14] Park, J., Sang Eun W. (2022) 'Who Likes Artificial Intelligence? Personality Predictors of Attitudes toward Artificial Intelligence'. *The Journal of Psychology* 156, pp. 68–94; https://doi.org/10.1080/00223980.2021.2012109.

[15] Prezelj, I., Harangozo, D. (2018) *Confidence and Security-Building Measures in Europe at a Crossroads*. Baden-Baden: NOMOS. https://doi.org/10.5771/9783845288970.

[16] Prezelj, I., Ristevska, T. T. (2022) 'Intelligence Scandals: A Comparative Analytical Model and Lessons Learned from the Test Case of North Macedonia', *Intelligence and National Security,* 38(1), pp. 143-170; https://doi.org/10.1080/02684527.2022.2065616.

[17] Raska, M., Bitzinger, R. A. (2023) 'Introduction: The AI Wave in Defence Innovation', in Raska, M., Bitzinger, R. A. (eds.) *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities and Trajectories*. New York: Routledge, pp. 1–11; https://doi.org/10.4324/9781003218326-1.

[18] Rickli, J.-M., Mantellassi, F. (2023) 'Artificial Intelligence in Warfare', in Raska, M., Bitzinger, R. A. (eds.) *The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities and Trajectories*. New York: Routledge, pp. 12-36; https://doi.org/10.4324/9781003218326-2.

[19] Schepman, A., Rodway, P. (2023) 'The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust'. *International Journal of Human–Computer Interaction* 39, pp. 2724–2741; https://doi.org/10.1080/10447318.2022.2085400.

[20] Schuller, M. (2023) Human and Machine Learning, Paper presented at a conference NATO in the Nordics, August 30-31st 2023, Stockholm.

[21] Soare, S. (2023) 'European Military AI: Why Regional Approaches are Lagging Behind', in Raska, M., Bitzinger, R. A. (eds.) The AI Wave in Defence Innovation: Assessing Military Artificial Intelligence Strategies, Capabilities and Trajectories. New York: Routledge. https://doi.org/10.4324/9781003218326-5.

[22] Thiele, R. (2021a) 'Nineteen Technologies in Focus', in Thiele, R. (ed.) Hybrid Warfare: Future and Technologies. Wiesbaden: Springer VS, pp. 71-124; https://doi.org/10.1007/978-3-658-35109-0_5.

[23] Thiele, R. (2021b) 'Annex 2 – Artificial Intelligence', in Thiele, R. (ed.) Hybrid Warfare: Future and Technologies. Wiesbaden: Springer VS, pp. 187-196; https://doi.org/10.1007/978-3-658-35109-0.

[24] Thiele, R. (2021c) ‘Technology as a Driver’, in Thiele, R. (ed.) Hybrid Warfare: Future and Technologies. Wiesbaden: Springer VS, pp. 59-70; https://doi.org/10.1007/978-3-658-35109-0_4.

[25] Artificial Intelligence (2023) Encyclopaedia Britannica [Online]. Available at: https://www.britannica.com/technology/artificial-intelligence (Accessed: 09 August 2024).

[26] Artificial Intelligence Act, Briefing, EU Legislation in Progress, European Parliamentary Research Service, June, 2023.

[27] Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World, Council of the EU, Press Release 986/23, 9.12, 2023.

[28] Flash Wars: Autonomous Weapons, AI and the Future of Armed Conflict, documentary movie, Director Daniel Andrew Wunderer, Blue + Green Communications, 2023.

[29] Geneva Conventions of 12 August 1949 (1949) ICRC [Online]. Available at: https://www.icrc.org/sites/default/files/external/doc/en/assets/files/publications/icrc-002-0173.pdf#:~:text=the%20ICRC%20is%20at%20the%20origin%20of%20the%20Geneva%20Conventions (Accessed: 07 August 2024).

[30] Protocols Additional to the Geneva Conventions of 12 August 1949 (1977) [Online]. Available at: https://www.icrc.org/sites/default/files/external/doc/en/assets/files/other/icrc_002_0321.pdf#:~:text=Geneva%20Conventions%20of%2012%20August%201949,%20and%20relating%20to%20the (Accessed: 09 August 2024).

[31] Final Report (2021) Washington, D.C.: National Security Commission on Artificial Intelligence. [Online]. Available at: https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf (Accessed: 09 August 2024).

[32] Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development, UN General Assembly Resolution, A/78/l.49, 11 March 2024.

[33] Summary of the NATO Artificial Intelligence Strategy (2021) Meeting of Defence Ministers, 22 October, Brussels.

[34] Terminator Genesis, Director: Alan Taylor, IMDbPro, 2015.