The Central Research institute
of the Armed Forces of Ukraine

Military Science. Vol. 3 No. 3 (2025)

Received: *August 25, 2025* | **Revisions Requested:** *September 9, 2025* | **Revision Received:** *October 3, 2025*
Accepted: *October 3, 2025* | **Published Online:** *December 1, 2025*

Licensed (C) by Creative Commons Attribution International License 4.0 (CC BY-NC-SA)

DOI: 10.62524/msj.2025.3.3.11

Risk-Based Approach to AI Implementation in Ukraine's Defense Sector: Conceptual Framework

Stanislav Shumlianskyi
PhD in Political Science
Senior Officer
Department of the State Special Transport Service Reforming and Development
Research Centre State Special Transport Service
02094, Yakov Hnizdovsky Street, 5, Kyiv, Ukraine
e-mail: stanislavshum@icloud.com
ORCID: 0009-0008-2739-7614

Yevhen Khomenko
PhD student in Archeology
Commander
Research Centre State Special Transport Service
02094, Yakov Hnizdovsky Street, 5, Kyiv, Ukraine
e-mail: centrd@dsst.gov.ua
ORCID: 0009-0006-7006-3439

Serhiy Popriyenko
Officer
Department for Armaments and Military Equipment Unmanned Systems and EW Research
Research Centre State Special Transport Service
02094, Yakov Hnizdovsky Street, 5, Kyiv, Ukraine
e-mail: centrd@dsst.gov.ua
ORCID: 0009-0006-8517-1045

Abstract. The article discusses the conceptual foundations of a risk-oriented approach to introducing artificial intelligence (AI) into Ukraine's defense sector. The study's relevance stems from the rapid development of AI technologies in military affairs and the need to counter the prevailing enemy. Thus, the article aims to identify and substantiate the fundamental principles of a risk-oriented approach to integrating AI into Ukraine's military. The primary method used in this study is analysis, which allows to consider non-technical internal risks, which are less obvious and potentially more dangerous than technical and external risks. The article identifies and categorizes key non-technical risks arising from the introduction of AI in the military. Specifically, it identifies the risks associated with underestimating threats, the uncertainty of AI development, the negative impact on humans (e.g., reduced cognitive skills and excessive trust), and problems related to automatic translation and decision-making. The article also considers external risks arising from the enemy's use of AI, such as increased intelligence effectiveness, cyberwarfare, propaganda, and data poisoning. The authors substantiate the necessity of creating a register and map of these risks for further management. The

article's materials are practically valuable for developing an AI policy for the military, creating a risk register and map, and for further work with them.

Keywords: artificial intelligence, AI, AI risks, AI strategy, risk identification, information technologies, risk-based approach, intellectualization of warfare, risk map, risk register.

Introduction

The modern world is experiencing rapid development and active implementation of artificial intelligence (AI) technologies in various spheres, and military affairs are no exception. AI systems are becoming an integral part of military operations, allowing for the analysis of intercepted signals, the identification of enemy positions, the disruption of communication networks, and the attainment of tactical advantages. They are capable of processing vast amounts of information, making autonomous decisions, recognizing threats, and performing other critical tasks.

The latest trend in warfare is the intensive development of robotic and autonomous air, land, and sea-based weapon systems. Leading military powers such as the US, Russia, China, and Israel have already developed strategic plans for the use of such systems. These technologies help the military to quickly make management decisions, analyze scenarios, optimize reconnaissance processes, detect and identify targets for the purpose of destroying the enemy, monitor threats, transport ammunition and provide medical assistance, as well as ensure cybersecurity.

Countries that have been developing and implementing AI technologies for at least the last 5-8 years are paying more attention not so much to promoting and developing this field, but rather to standardization, regulation, restrictions, and even outright bans. For example, in the US, back in 2020, Congress ordered the Federal Communications Commission to develop a program to remove and replace equipment from Chinese manufacturers (Huawei and ZTE Corporation) as the material basis for AI technologies, given the potential danger they could pose to the United States [1].

The Artificial intelligence act, adopted in May 2024 by the Council of the European Union, aims to harmonize AI regulations based on the logic of potential risks. It states that artificial intelligence systems that can be used in various applications are analyzed and classified according to the risk they pose to users. Different levels of risk mean more or less regulation, i.e., the higher the risk of harm to society, the stricter the rules. Artificial intelligence systems with unacceptable risk are systems that are considered a threat to people. The law mentions such unacceptable systems: cognitive-behavioral manipulation of people or certain vulnerable groups (e.g., voice-activated toys that encourage dangerous behavior in children); classification of people based on behavior, socioeconomic status, or personal characteristics; real-time and remote biometric identification systems, such as facial recognition, etc. [2].

The People's Republic of China considers AI to be a critical component of its national strategy, crucial to its economic development, technological superiority, and, in particular, to strengthening its defense capabilities [3]. This vision is reflected in policy documents, including *Made in China 2025* and the *New Generation Artificial Intelligence Development Plan*, which outline China's ambitious goal of becoming a global leader in AI by 2030 [4]. China's approach, known as “intellectualisation” as the basis for the development and re-equipment of the country and the defense sector in particular, involves the total implementation of AI technologies by the state without taking into account or even mentioning (in the public sphere) the risks, let alone neutralizing or minimizing them [5]. Thus, the existing strategic approaches to the implementation of AI by states can be defined as: 1) “AI skepticism” (the EU, other Western countries, and to a lesser extent the US) – it qualifies AI as a potentially dangerous technology and aims to prevent or minimize future risks and dangers from (or during) its implementation, i.e., it is risk-oriented; 2) “AI optimism” (primarily China, and in the future, possibly other countries of the “global South”) – focused on the speed of implementation, pace

of development, results and achievements, and, accordingly, a lack of adequate attention to potential risks or even disregard for them.

Ukraine is currently implementing AI technologies, having identified their development as one of the priorities for the army and the defense-industrial complex, primarily in view of the need to organize an effective defense against the large-scale invasion by the Russian Federation, which has been ongoing for more than three years. Previously, technical innovations and inventions were introduced in other countries and Ukraine imitated and borrowed them, but with the start of a large-scale invasion and the need to resist an enemy that is superior in everything, Ukraine has become a "pioneer" of military developments and inventions and, in particular, a "testing ground for AI warfare" [6]. Thus, the introduction of AI elements by Ukraine (or, more precisely, "in Ukraine," being more a collection of individual grassroots projects than a centralized systemic strategy from above) is proceeding at a rapid pace, and in some respects even ahead of other countries (at least, this is characteristic of the defense sector). AI policy in Ukraine is reactive in nature, i.e., its development at the national level lags significantly behind both the development of the technology itself and even its practical application in Ukraine, especially, as noted above, in the areas of weapons and military equipment, the execution of combat missions in the context of armed aggression by the Russian Federation, etc.

This state of regulation, or more precisely, the lack of regulation or insufficient regulation of the use of AI as the main trend in the development of technology and military art, is entirely justified in the extraordinary conditions of the "hot" phase of a large-scale war, but it carries risks and dangers. At this point, given that the introduction and spread of AI has not yet reached a significant scale and ubiquity, these risks and dangers have not yet materialized, i.e., they are potential and subject to identification, assessment, and minimization (prevention). Since our state's policy on AI is currently embodied in the form of concepts [7][8], i.e., it is in the early stages of formation, and sectoral (in particular, defense) strategies, let alone doctrines and standards for the use of AI, have not yet been developed or are still in the development stage, justifying the feasibility and critical importance of a risk-based approach can be a practical contribution to this process.

The purpose of this article is to define the basic principles of a risk-based approach to the implementation of AI by Ukraine in the military sphere. At the same time, based on the division of risks into internal and external, technical (risks inherent in the technologies themselves) and all others related to the use of AI by humans ("non-technical"), the object of study is internal non-technical risks, as they are less obvious, less predictable, much less frequently mentioned in scientific and expert works and, therefore, potentially more dangerous compared to technical and external risks (the use of AI by the enemy), which are fairly well described and identified.

Research objectives: 1) to outline the problem area and the main areas of non-technical risks arising from the introduction of AI in the military sphere; 2) to identify, as far as the current state of the subject of research allows, specific risks; 3) to describe the main components of a risk-based approach to developing a policy for the implementation of AI in the military sphere, its stages, and priority measures; 4) to offer general recommendations for working with risks and recommendations for preventing or minimizing specific identified risks; 5) to determine the main areas for further work with AI risks in the military sphere, in particular, their prediction (foresight).

In addition to its purely scientific value, this article also has **practical significance**. The proposed risk-based approach to AI implementation, which is used in the EU and, to some extent, in the US, together with a list of identified risks, can be applied in the development and further implementation of AI implementation policy (this is most urgent in all matters that are in one way or another related to combat operations, the use of weapons and, accordingly, the lives of personnel, civilians, and significant destruction), the creation of a register and map of risks, as well as further work with them, which will soon become mandatory and routine at all levels and in all spheres of public life and military activity, to which this article is devoted, in particular.

To our knowledge, a risk-based approach to AI implementation has not yet been proposed in Ukraine, and a register of risks from AI implementation – both in the military sphere and in general – as well as a risk map have not yet been compiled, and there is no information (at least in open sources) about work on them. Therefore, the justification of their feasibility presented in the article, as well as the list and description of risks themselves, can contribute to their creation, although the list itself is certainly not exhaustive and comprehensive, requires discussion, clarification, and correction, as well as constant refinement, given the rapid development of the subject itself and the application of AI in real combat conditions.

Methodological Framework

One of the most common definitions of risk is the potential possibility of adverse events occurring [9, p. 263]. The existence of risk presupposes the presence of certain mandatory conditions: the possibility of a random event with negative consequences; the presence of material or other damage, as well as activities related to this event [9, p. 264]. Risk is also defined as a derivative of challenge, danger, and threat [10, p. 20].

With regard to the military sphere, NATO's risk management manual defines risk as “an uncertain event or condition that, if it occurs, has a positive or negative impact on the objectives of the project” [11, p. 2]. According to this definition, the term “risk” covers both hazards and threats. In essence, the management of all risks (both hazards and threats) is carried out in a similar manner (although there are differences in the quantitative measurement of risk and response strategies). The same manual introduces a division of risks into internal and external.

Internal risk is defined as "a risk whose source falls within the scope of responsibility of the risk management process. The source of external risk, on the contrary, is outside the scope of responsibility of this process" [11, p. 9]. Dividing risks into internal and external simplifies the process of choosing ways to respond to risks at a certain level, but does not help to understand why and how to manage these risks. The responsibility for identifying all risks (even those that require the involvement of other levels/stakeholders) lies with each level, as does the responsibility for initiating a response at the appropriate level.

Let's consider the main approaches to risk management and their components.

According to Ukrainian ISO 31000:2018 standard (international standard “Risk Management – Guidelines”), the risk management process consists of the following stages:

- defining the scope, context, and criteria for identifying risks;
- risk assessment (consisting of identification, analysis, and, in fact, risk qualification);
- risk measures and risk management.

In addition, the components that support this process and/or are carried out at all its stages are information gathering and reporting, communication-consulting, and monitoring and evaluation [12, pp. 8-15]. In the NATO standard for risk management, this process is divided into risk identification, risk analysis, risk response planning, risk monitoring and control, and planning of all risk management activities [13, pp. 24-44].

Order of the Ministry of Defense of Ukraine No. 145 dated April 2, 2019 introduces such concepts as risk profile (“the result of assessing the level of risk materiality in an institution on a specific date before taking into account measures to minimize risks, as well as taking into account such measures, in terms of each risk”), risk register (“a document that includes a description of identified risks and their assessment”), a risk management plan consisting of risks, response measures, deadlines, responsible persons, and expected risk management results [14, p. 2]. In addition, the order provides a risk assessment matrix based on two criteria: consequences (impact) and probability of occurrence (Appendix 3).

Another document – an analytical report by the Ukraine's National Institute for Strategic Studies in 2020 – based on an analysis of national risk and threat assessment systems in various

countries, proposes the following algorithm for working with risks:

1) identifying the greatest threats to national security (screening);
2) conducting an in-depth analysis of possible consequences, developing scenario forecasts, and performing modeling;
3) assessing capabilities and identifying vulnerabilities;
4) conducting comprehensive mapping and geospatial support (geospatial support);
5) dissemination of risk assessment results, visualization;
6) monitoring and reassessment of risks, taking into account lessons learned [15, pp. 34-39].

One way to show the risks of using artificial intelligence – both in general and in the military affairs in particular – is with a risk map, which is a ”graphic and text description of a limited number of risks to an organization, arranged in a rectangular table, with one axis showing the impact or significance of the risk, and the other axis indicates the probability or frequency of its occurrence“ [16, p. 40]. At the same time, ”the ultimate value of a risk map ... is not in determining the exact impact or probability level of a specific threat, but in the relative location ... and their location in relation to the risk tolerance threshold“ [16, p. 45].

When compiling such a map, “each axis is divided into three or four sections, which intersect to form 10-20 areas,” each risk falls into the corresponding cell, after which the so-called “tolerance line” is determined on the map in such a way that some of the risks are above and some below this line [17, p. 104].

Less complex to create and usually preceding the map is the risk register. Tools for creating AI risk registers are developed based on general risk management principles, but they also have their own specific features, primarily due to the need to capture the dynamism, uncertainty, and autonomy inherent in AI risks [18].

Since AI risk registers are often internal documents of organizations, direct references to specific, fully completed registers may be rare due to confidentiality. However, there are public frameworks, templates, and recommendations that describe how to build such registers [19].

Results

Below, we will outline the most obvious and potentially dangerous risks posed by the introduction of AI in the defense sector, particularly those that are decisive and primary in relation to other, technical risks.

Underestimation of AI risks. Identification (detection) and subsequent assessment of each risk is a basic element of the internal control and risk management system. In the field of defense, “risk identification consists of determining events, circumstances, or a combination thereof that will affect an institution's ability to perform its tasks and functions, the targeted and effective management of budget funds, state property, and other resources, the functioning of information (automated), electronic communication and information and communication systems, the functioning of internal control, and achieving the defined goals (mission), strategic and other objectives of the institution's activities“ [20, p. 4].

As mentioned above, risk assessment is given considerable attention in policy development and implementation analysis, to the extent that risk is the basic criterion for evaluating a particular AI technology (as in documents from the EU, the United States, and NATO). Meanwhile, in Ukraine, as shown by our analysis of the only document on AI currently adopted at the legislative level (the 2020 Development Concept), as well as the Strategy developed by the institute of the National Academy of Sciences [21], no risk assessment has been carried out, and there is not even a list of possible risks and threats. Similarly, in another document, Ukrainian Scientific and Technical Foresight, which presents a summary assessment by leading Ukrainian scientists of strategic priority areas and prospects for the development of science and technology, and where, in the section “Digital Technologies, Artificial Intelligence, and Cybersecurity,” the risks, dangers, and threats of AI are

mentioned only in passing and without emphasizing the importance of this issue (the risks of implementation) [22, pp. 21-27].

Risk of uncertainty. Another fundamental risk lies in the inability to accurately predict the development of AI and, consequently, the risks and dangers that will arise. This is due to both the fact that the systems are self-developing and that user control over the technology is insufficient.

Negative impact on humans. The negative impact on humans is a whole set of factors, consisting of both the general impact of this technology on people (which must be taken into account) and the impact of AI on a specific person (or persons) in a specific situation and when performing specific tasks, and can lead to negative (or unforeseen) consequences.

Although the purpose of AI is to increase capabilities and ensure better results, its use can have a negative impact on performers. Research in the civilian sphere shows that the systematic use of AI tools can lead to a decline in cognitive skills, a reduction in critical thinking, and dependence [23, p. 13]. As experts point out, "our brain prefers to conserve energy," and the use of AI can lead to "intellectual laziness" when more work is transferred to technology, which can potentially reduce the skill level of even experienced professionals.

Decisions related to the implementation and use of AI can be influenced by **cognitive biases**, which are systematic errors in thinking that arise under the influence of psychological, emotional, and social factors and can distort the perception of reality [24]. These biases, namely the Dunning-Kruger effect, the gambler's fallacy, the planning fallacy, the illusion of truth effect, wishful thinking, common ground bias [25, p. 21], Peter's principle, etc. can interact with each other, significantly increasing their impact on the decision-making process and creating complex reinforcement effects, since the simultaneous manifestation of several biases can lead to even more significant distortions and irrational decisions than their separate influence, complicating the process of making informed and accurate decisions, contributing to the formation of a distorted perception of real facts, increasing the likelihood of serious, combined errors, and influencing management strategy.

One such bias, and at the same time psychological attitude, is excessive trust and reliance on authority, which in this case is AI. It is not difficult to predict that in an environment such as the army, especially when it comes to tasks involving risk and great responsibility for choosing alternatives, this can lead to (often unconscious, subconscious, which only exacerbates the risk) shifting responsibility to "omniscient AI." And although artificial intelligence models do indeed possess a vast amount of information and the ability to process it hundreds and thousands of times greater than that of an individual human being, all of this needs to be qualified with "in general," "in the vast majority of cases," and so on. AI also makes mistakes (as discussed below), and its mistakes can occur (and lead to extremely negative and even catastrophic consequences) in this particular situation.

Other factors that may cause negative consequences in the future (and over which there is no control) include:

1) possible errors in the program code, which the user may not be aware of and which could lead to fatal errors;

2) inertia (the task for which AI is used contains fundamentally new parameters compared to those that were originally set);

3) the inability of the system to independently change its parameters without human intervention, which can lead to inaccurate or false results. This means that the AI model always offers solutions based on information "as of...", although the user is largely unaware of this fact and, on the contrary, has a false perception that AI works with information in real time.

In addition, all modern large language models (with the possible exception of China's DeepSeek) are English-language tools. This is linked to a whole range of other factors (cultural, worldview, and so on) that are specific to the Western world, and above all to the United States. Citing numerous examples of AI being helpless when given tasks in other languages, and sometimes even giving obviously wrong answers (which can be proven with a "one-click" search), the researcher

claims that "harmful responses in these languages (all except English – S.S.) increase by 35%, and the accuracy of instruction execution decreases to 80%," warning that "for the non-Western world, all this makes AI only a novelty for experimentation, rather than a reliable partner in work or a reliable source of information" [26].

Thus, although users themselves do not notice it, their interaction with AI models is not direct but indirect, and there may be more than one translation operation. Therefore, there is a separate risk, or rather a group of risks, associated with automatic translation. While translators generally provide good and accurate translations from and into a language, an error or inaccuracy in a war situation or specific context can be very costly. In the context of AI translation, there is also a risk of data theft [27, p. 291].

AI in decision-making

The application of AI to specific tasks, including military ones, can vary in nature and depth. In general, three levels can be distinguished:

uncontrolled – developed programs or algorithms work and deliver results without direct human involvement;

auxiliary – artificial intelligence acts as a means (collecting, organizing, selecting, sorting data) or an "assistant" for human decision-making;

autonomous – AI is empowered to make decisions independently, human intervention may be selective and optional, in particular in the form of post-control [28, p. 194].

The same division into three levels can be made from another perspective – that of human authority. According to this approach, these levels are described as "human – part of the cycle," "human in the middle of the cycle," and "human outside the cycle" of working with artificial intelligence. Based on this division, all modern weapons are divided into: 1) semi-autonomous, 2) controlled autonomous, and 3) fully autonomous [28, p. 195].

At the same time, the greatest and least manageable (minimizable) risks are those associated with the use of AI in decision-making or when AI makes decisions independently. For the defense sector, particularly during Ukraine's repulsion of large-scale armed aggression by the Russian Federation, these risks are maximized when AI is given managerial powers. However, such use (without, as mentioned above, taking into account the possible risks) is provided for in the *Concept of Artificial Intelligence Development* (among other things, command and control, information analysis during combat operations, cognitive analysis of military units) [7] and in the foresight of the future of Ukrainian science, conducted within the framework of the National Academy of Sciences, where "the use of artificial intelligence to support decision-making and optimization in the field of defense" is indicated as one of the priorities [22, pp. 10, 15].

Granting AI the power to make decisions independently – i.e., its **agency** – is the most debated issue among its developers, human rights activists, politicians, and scientists. It is precisely in the agency of AI that the greatest risk lies, or more precisely, the source of future risks, only some of which (given the self-development of this technology) can be predicted. "AI models," according to Yuval Noah Harari, "are not just tools in our hands. They are agents of creating new reality" [29].

Given the diversity of their capabilities and, accordingly, their potential results, which are incomparable with AI tools, AI agents are being implemented in an increasing number of areas, with both the number and range of these areas growing exponentially. However, in Western European countries and the United States (mainly due to the legislative restrictions discussed above, precisely because of the existing and potential risks), as well as in Ukraine, AI agents currently only perform the role of assistants. An example of the use of an AI agent with a clear set of simple assistant functions is the latest initiative of the Ministry of Digital Transformation to introduce an AI agent in "Diia" with the functions of providing answers and creating references at the request of users [30].

However, in the People's Republic of China, where the implementation of AI is the basis of state policy, without any restrictions or possible risks, AI agents (only those developed in China and

fully controlled by it, such as DeepSeek) are already equipped with functions (so far only in an experimental, limited format) to make decisions that directly affect people's lives – for example, judges [31].

AI in general, and AI agents in particular, will eventually "know and be able to do" more and more, while making fewer and fewer mistakes. In particular, experiments in the field of weapons systems increasingly show the superiority of autonomous weapons over those controlled by humans. Therefore, giving them more and more "power" and reducing control may become a trend in the future. Many researchers are convinced that, in the words of Olha Kovalchuk, "AI will never grow up; decision-making must always remain with humans."

The question of giving AI the "final say" is a dilemma that will only become more relevant over time, with no clear solution in sight. But in the military sphere, especially in wartime and in specific combat situations, it cannot remain unanswered. "Will the machine be held responsible for a false target and, as a result, the death of personnel and/or civilians in real combat conditions?" asks the author of an analytical review of AI in future wars [32].

External risks: use of AI by an enemy

Artificial intelligence, being a dual-use technology, poses significant threats if used by an adversary. An enemy (another state, non-state actor, or terrorist group) can use AI to achieve its military goals, both offensive and defensive, as well as in the field of information and cyber warfare. This could radically change the nature of modern conflicts and increase their destructive power.

1. *Enhanced intelligence and surveillance capabilities.* The use of AI significantly expands the enemy's capabilities in gathering and analyzing intelligence. In particular, AI allows for much faster analysis of huge amounts of information (e.g., large arrays of satellite images) to identify military targets, troop movements, and changes in the landscape, providing a significant advantage in situational awareness. In addition, the use of swarms of drones or other autonomous robotic systems for reconnaissance reduces risks to personnel and allows for effective coverage of large areas [33, p. 192].

2. *Improved targeting and engagement.* AI can significantly accelerate and optimize target identification and engagement processes. AI systems are capable of processing data from various sensors (radars, thermal imagers, acoustic sensors) for real-time detection, offering optimal options for their engagement. The creation and use of autonomous strike drones or other systems that can independently find and destroy targets reduces decision-making time and increases strike potential [34].

3. *Cyber warfare and information operations.* Artificial intelligence significantly expands the possibilities for both defense and attack in cyberspace and the information environment. On the one hand, AI is used to improve cyber defense, predict and detect threats. On the other hand, AI also allows attackers to create more sophisticated and targeted attacks.

a) *Automated cyberattacks.* AI can automate the search for vulnerabilities in systems and the development and launch of malicious software, making cyberattacks faster and more sophisticated, thereby increasing their destructive potential [35, p. 124]. There is also a risk of confidential or sensitive information being extracted by analyzing large amounts of open data.

b) *Disinformation and propaganda.* AI allows generating realistic deepfakes (videos, audio, images) and writing convincing fake news that is practically indistinguishable from human-created content [36, p. 12]. Such materials can be distributed in large volumes with minimal human intervention, manipulating the public opinion of the enemy and one's own citizens. Detecting such content poses a serious challenge for existing automated detection systems [36, p. 29]. Targeted psychological operations, enhanced by AI, can analyze population profiles to create personalized messages that have the maximum impact on demoralization, social division, or the spread of panic. AI can also be used for recruitment and persuasion thanks to its ability to create flattering and personalized messages [37].

c) *Impersonation and fraud (scam).* Malicious actors can use deepfake technology to create fake voices, images, or writing styles to impersonate other people. This allows them to carry out fraudulent operations, such as imitating the voices of relatives to extort money, as happened in cases of "virtual kidnapping," or giving orders live on air, as at a meeting of the British construction corporation Arup, when a virtual (constructed) member of the board of directors (deepfake) gave an order to carry out 15 transactions totaling £20 million [38]. Operations similar to this one in the defense sector can be carried out to extract information or coerce decisions that benefit the enemy.

4. *Optimization of logistics and resource utilization by the enemy*. AI can significantly improve the efficiency of logistics processes and troop management. The enemy can use AI to optimize supply routes, allocate resources, forecast needs, and minimize losses, making their logistics more resilient and efficient. AI can also assist in planning operations, assigning tasks, and optimizing troop movements on the battlefield, providing a tactical advantage [28, p. 195].

5. *Deceptive actions and countermeasures*. AI provides new opportunities to mislead the enemy by creating convincing virtual copies of military equipment, fake radio signals, or simulating troop activity to confuse the enemy and force them to waste resources on false targets. At the same time, the enemy can develop its own AI to detect and neutralize enemy AI systems, for example, by jamming communications with drones or detecting vulnerabilities in algorithms [32].

6. *Robotics and swarms.* The use of large numbers of small, AI-controlled drones that interact with each other allows the enemy to overcome air defenses, conduct reconnaissance, or strike targets, which makes countermeasures much more difficult [32].

7. *Problems of verification and data "poisoning*" AI complicates the process of human verification of information, especially that generated by AI systems themselves ("deepfakes," "hallucinations"). AI "hallucinations" are situations where the model cannot distinguish real content (photos, videos, texts) from generated content, which creates the risk of making wrong decisions based on false data [31].

The enemy can also deliberately generate large amounts of falsified ("corrupted") data to "poison" the adversary's AI training datasets. This will cause the AI to make incorrect decisions or carry out incorrect or unnecessary military actions, which will negatively affect operations and results on the battlefield [39].

In addition, there is a risk that the enemy will obtain protected or classified information by analyzing a large array of other, seemingly unclassified data from open sources (for example, determining the age of Instagram users based on greetings from friends or determining location based on photos) [40, p. 206].

The use of AI by the enemy requires constant analysis and adaptation of one's own military strategies, significant investments in countermeasures, and the development of one's own AI technologies to maintain parity and ensure national security.

Discussions

Theoretical understanding of AI in general, as well as specific areas and spheres of its implementation (including military), is developing very intensively. Reports and analytical studies are devoted to this topic, which is constantly present, in particular, in such Ukrainian scientific publications as "Cybersecurity: Education, Science, Technology," "Modern Information Technologies in the Field of Security and Defense," "Modern Information Protection," "Artificial Intelligence," and others. It is one of the most frequently researched topics in both Ukrainian and international journals, with the number of scientific articles growing rapidly.

At the same time, Ukraine's AI implementation strategy (in the military affairs in particular) has not yet been the subject of research, as it (like the algorithms, models, standards, and protocols based on it) has not yet been adopted and made public. Thus, the article contributes not only to research but also to the development of relevant policy, emphasizing the need for a risk-based approach as its (AI

implementation strategy) foundation.

The preliminary (working) list of risks proposed in the article can serve as a basis for further work on the AI risk register and map specifically in the military affairs. In the future, the compilation, maintenance, and continuous refinement of AI risk registers, maps, and matrices in the military area should become an indispensable part of AI implementation in the defense sector. These issues will require additional research and expert discussions among policy developers and implementers, as well as scientists and all interested parties.

Conclusion

Awareness of the challenges posed by AI implementation, including cybersecurity threats, potential algorithmic biases, socioeconomic implications, and ethical issues, is the basis for a responsible approach to AI implementation. Careful analysis, the development of comprehensive risk mitigation strategies, continuous monitoring, and maintaining human control are key factors for the successful and safe use of AI technologies in Ukraine's defense sector.

A risk-based approach is a means of managing risks and minimizing potential dangers and threats with potentially serious consequences, including human casualties. According to this approach, the implementation of AI at any level-from strategies, doctrines, and standards to implementation in specific decisions and applications-should be based on preliminary risk management work that is as clear and detailed as resources and circumstances allow. At the same time, the forms (and results) of such work can be diverse and vary depending on the field and situation: a map, a register, a risk matrix, etc.

Recommendations

AI risk management should be carried out in two main areas. The first involves developing and implementing defenses and countermeasures for each identified risk, especially in the context of enemy use of AI. This includes improving cyber defenses, developing methods for detecting anomalies, and protecting models from targeted attacks. The second area focuses on creating and adhering to clear protocols, regulations, and security measures when AI is used by defense forces at all levels, from strategic to operational-tactical. This will limit and control the areas and conditions of AI use, minimizing potential errors and human influence. Future research should focus on developing explainable AI systems, dynamic risk assessment models, and countermeasures against hostile AI use.

To minimize the potential risks associated with the use of AI in the military affairs, it is advisable to take the following measures:

1. Develop a clear strategy for AI implementation based on a risk-oriented approach. The strategy should include a detailed analysis of potential internal and external risks, the definition of ethical norms and safety standards, as well as mechanisms for responding to incidents. It is important to clearly identify potential threats, especially those related to the specifics of military AI applications. A necessary prerequisite for such a strategy is a clearly defined legal framework for the use of AI, preferably in the form of a specific law.

2. Invest in cybersecurity and data protection. Reliable systems for protecting data and infrastructure from cyberattacks must be developed and implemented. This includes using verified data sources that have been pre-checked for compliance with security standards, which will reduce the risk of data poisoning attacks. It is also important to implement security monitoring tools for large language models (LLMs) to detect prompt manipulation and other attempts to circumvent restrictions. Protecting the model from attacks may include: a) *Adversarial Training* – an approach that involves training the model on specially modified data that can simulate potential attacks; b) *Differential Privacy* – a method that ensures the confidentiality of training data by masking the contribution of individual records to the sample, making it impossible to identify or use specific personal data in the

training process [40].

3. Ensure transparency and accountability of AI systems. Where possible, algorithms whose decisions can be explained should be used. There should be clear mechanisms for control and accountability for the development, implementation, and operation of AI systems.

4. Focus on training and professional development of personnel. State Special Transport Service employees should have the necessary knowledge and skills to work with AI systems, understand their potential risks, and be aware of methods for detecting AI-generated content. The risk of declining cognitive skills due to overreliance on AI should also be taken into account.

5. Conduct regular monitoring and auditing of AI systems. It is necessary to continuously monitor the operation of AI systems, identify and eliminate vulnerabilities, and ensure that their operation complies with established standards and ethical norms. Particular attention should be paid to identifying anomalies that may indicate potential risks or deviations in operation.

6. Establish cooperation with experts in the fields of AI, cybersecurity, and cognitive sciences. Involving qualified specialists will help in the development and implementation of safe and effective AI-based solutions, as well as in understanding the impact of cognitive biases.

7. Maintain human control over critical decisions. As noted above, "AI will never grow up; decision-making must always remain in human hands," and AI cannot be held responsible for its mistakes, which, in particular, have led to human casualties and other irreparable consequences or damage. It is important to use AI technologies as a tool to assist, rather than replace, human intelligence, especially where decisions can have serious consequences. Implementing these measures requires a comprehensive approach, continuous monitoring, and interagency cooperation to ensure the safe and effective use of AI in the interests of national security.

Insufficient attention to risks, disregard for a risk-oriented approach, and the creation of strategic and doctrinal instruments on other grounds (such as the reckless introduction of AI, as in the People's Republic of China) in the context of ongoing enemy aggression, combat operations, the widespread use of weapons by the Armed Forces of Ukraine, and the need to counter such use by the enemy, may have significant negative consequences, namely: significant personnel losses, civilian casualties, and, in combination with the targeted use of AI technologies by the enemy , reputational risks for the Defense Forces, undermining trust in the command on the part of personnel and the civilian population, as well as other serious consequences that cannot be clearly identified in advance at the current stage of development and implementation of this "pioneering" technology.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to express their gratitude to Major Igor Svetlichny, Head of the Department for Research on Reform and Development of the State Special Transport Service at the Research Center, for his comments on the final draft of the article, and to Lieutenant Svetlana Emelyanova, an officer in the Department for Research on Reform and Development of the State Special Transport Service at the Research Center, for her assistance in compiling the list of references and bibliography.

References

1. Graham, E. (2025, April 25). "*All hands on deck" moment in the US-China AI race: Pentagon's former digital chief*. Defense One. https://www.defenseone.com/technology/2025/04/us-needs-fast-track-ai-counter-china-dods-former-ai-lead-says/404367
2. Sklyarenko, V. (2023, May 23). *Different rules for different risk levels within the first law on artificial intelligence*. Liga:Zakon. https://biz.ligazakon.net/analitycs/228664_rzn-pravila-

dlya-rznikh-rvnv-riziku-v-ramkakh-pershogo-zakonu-pro-shtuchniy-ntelekt
3. Stokes, J. (2024). *The PLA's AI journey: A long and winding road*. Center for a New American Security. https://www.cnas.org/publications/reports/the-plas-ai-journey-a-long-and-winding-road
4. State Council of the People's Republic of China. (2017). *A new generation artificial intelligence development plan*. http://www.gov.cn/zhengce/content/2017-07/20/content_5211990.htm
5. Shumliansky, S., & Svetlichnyi, I. (2025). *Intellectualization of military art: The use of artificial intelligence by China in the military sphere and its geopolitical consequences*. Sinological Studies, 3. (forthcoming)
6. Fontes, R., & Kamminga, J. (2023, March 24). *Ukraine a living lab for AI warfare*. National Defense Magazine. https://www.nationaldefensemagazine.org/articles/2023/3/24/ukraine-a-living-lab-for-ai-warfare
7. Cabinet of Ministers of Ukraine. (2020, December 2). *On the approval of the Concept for the Development of Artificial Intelligence in Ukraine* (Decree No. 1556-p). https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text
8. Cabinet of Ministers of Ukraine. (2024, April 13). *On the approval of the Concept of the State Target Scientific and Technical Program for the Use of Artificial Intelligence Technologies in Priority Sectors of the Economy for the Period up to 2026* (Decree No. 320-p). https://www.kmu.gov.ua/npas/pro-skhvalennia-kontseptsii-derzhavnoi-tsilovoi-naukovo-tekhnichnoi-prohramy-z-vykorystannia-s320130424
9. Romanchyk, T. (2020). Danger, threat, risk: Analysis of the terminological apparatus of the theory of economic security. *Economic Bulletin of NTUU KPI, 24,* 262–267. https://ev.fmm.kpi.ua/article/download/192866/216457/489099
10. Kalashnikova, L. (2017). The correlation of the concepts of “uncertainty," "risk“, “challenge“, “threat“, “danger“ in the context of sociology of life safety. *Scientific and Theoretical Almanac “Grani",* *20*(5), 16–23. https://www.researchgate.net/publication/319201965_Spivvidnosenna_ponat_neviznacenist_rizik_viklik_zagroza_nebezpeka_u_konteksti_sociologii_bezpeki_zittedialnosti
11. Ministry of Defense of Ukraine. (2021). *Risk management in the Ministry of Defense of Ukraine and the Armed Forces of Ukraine: Methodological recommendations VST 01.054.001-2021(01)*. https://www.mil.gov.ua/content/pdf/vnytr_control/VST_01054001_-2021(01).pdf
12. DSTU ISO 31000:2018. (n.d.). *Risk management. Guidelines*. https://zakon.isu.net.ua/sites/default/files/normdocs/dstu_iso_31000_2018.pdf
13. Ministry of Defense of Ukraine. (2021). *NATO Risk Management Guide for Acquisition Programmes*. https://mod.gov.ua/diyalnist/normativno-pravova-baza/aramp-1-edition-1-nato-risk-management-guide-for-acquisition-programmes-idt
14. Ministry of Defense of Ukraine. (2019, April 2). *On the approval of the procedure for the organization of internal control and risk management in the system of the Ministry of Defense of Ukraine* (Order No. 145). https://mod.gov.ua/diyalnist/normativno-pravova-baza/nakaz-145-pro-zatverdzhennya-poryadku-organizacziyi-v-sistemi-mou-vnutrishnogo-kontrolyu-ta-upravlinnya-rizikami
15. National Institute for Strategic Studies. (2020). *National systems of risk and threat assessment: Best world practices, new opportunities for Ukraine*. https://niss.gov.ua/sites/default/files/2020-07/dopovid.pdf
16. Bondarenko, P. V. (2005). Risk map – an effective management tool. *Bulletin of Socio-Economic Research, 20,* 45–54. http://dspace.oneu.edu.ua/jspui/handle/123456789/2083
17. Artyshchuk, I. V. (2011). Approaches to building a risk map based on taking into account the impact of basic factors on the activities of a trading enterprise. *Entrepreneurship and Trade, 13,* 101–106. http://journals-lute.lviv.ua/journal/13_2011/26.pdf
18. Mohapatra, D. (2025). *How to build an AI risk register that actually works*. CognitiveView. https://blog.cognitiveview.com/how-to-build-an-ai-risk-register-that-actually-works/

19. Australian Government. (n.d.). *Risk assessment for use of AI*. https://www.digital.gov.au/policy/ai/risk-assessment
20. Ministry of Defense of Ukraine. (2019, April 2). *Methodological recommendations on the organization of internal control* (Order No. 145). https://www.mil.gov.ua/content/pdf/vnytr_control/n145_020419.pdf
21. Glushkova, O. (2023). Strategy for the development of artificial intelligence in Ukraine. *Journal of Artificial Intelligence, 2*(1). https://jai.in.ua/archive/2023/ai_mono.pdf
22. G.M. Dobrov Institute for Scientific and Technological Potential and History of Science, NAS of Ukraine. (2024). *Ukrainian scientific and technical foresight: Strategic priority directions and prospects for the development of science and technologies*. https://foresight.in.ua/UkrSciTechForesight2024.pdf
23. Zhmai, O. V. (2025). The impact of artificial intelligence on children and adolescents: Learning, development, well-being. *Journal of Artificial Intelligence, 1*(102), 10–26. https://doi.org/10.15407/jai2025.01.010
24. Horvat, T., & Dub, M. (2024). Cognitive biases: Essence and influence on financial decisions. *Entrepreneurship and Trade, 43,* 29–35. http://journals-lute.lviv.ua/index.php/pidpr-torgi/article/view/1722/1621
25. Yakymova, N. (2020). Cognitive biases and their impact on behavioral models of labor market participants. *Galician Economic Bulletin, 2*(63), 118–125. https://galicianvisnyk.tntu.edu.ua/pdf/63/814.pdf
26. Kyulyan, A. (2024, September 11). Why is artificial intelligence Anglocentric and why is it a big problem? *Teksty.org.ua*. https://texty.org.ua/articles/113389/chomu-shtuchnyj-intelekt-anhlocentrychnyj-ta-chomu-ce-duzhe-velyka-problema/?src=read_next&from=114815
27. Matvienko, L., & Khomenko, L. (2025). Risks of information leakage when using online machine translation services. *Cybersecurity: Education, Science, Technology, 3*(27), 284–293. https://csecurity.kubg.edu.ua/index.php/journal/article/view/730/616
28. Goncharuk, D. (2024). The use of artificial intelligence for military purposes. *Pravo.ua, 1,* 192–197. https://www.researchgate.net/publication/387066794_Zastosuvanna_stucnogo_intelektu_u_vijskovih_cilah
29. Harari, Y. N. (2024, October 28). Our AI future is WAY WORSE than you think. *The Rich Roll Podcast*. https://www.youtube.com/watch?v=_jl64f-821o&t=173s
30. Pylipiv, I. (2025, May 18). In June, the Ministry of Digital Transformation will launch an AI agent on the "Diia" portal. *Ekonomichna Pravda*. https://epravda.com.ua/tehnologiji/u-diji-z-yavitsya-shi-pomichnik-806820/
31. Shumei, L. (2025, April 12). China's AI-powered judges spark debate. *Global Times*. https://www.globaltimes.cn/page/202504/1332412.shtml
32. Horwitz, L. A. (2025, May 15). AI is the future of war. *Asia Times*. https://asiatimes.com/2025/05/ai-is-the-future-of-war/
33. Darienko, D., & Shyshkin, Y. (2024). The use of artificial intelligence for unauthorized access to containerized systems. *Modern Information Protection, 1,* 189–196. https://journals.dut.edu.ua/index.php/dataprotect/article/view/3180/3062
34. Wilbur, D. (2025, January 22). The challenge of AI-enhanced cognitive warfare: A call to arms for a cognitive defense. *Small Wars Journal*. https://smallwarsjournal.com/2025/01/22/the-challenge-of-ai-enhanced-cognitive-warfare-a-call-to-arms-for-a-cognitive-defense/
35. Tkachenko, O. et al. (2025). Modern studies of information influences in social networks. *Cybersecurity: Education, Science, Technology, 3*(27), 120–124. https://csecurity.kubg.edu.ua/index.php/journal/article/view/716/592
36. Yang, K., & Menczer, F. (2024). Anatomy of an AI-powered malicious social botnet. *Journal of Quantitative Description: Digital Media, 4,* 1–36. https://journalqd.org/article/view/5848/4519

37. Salvi, F. et al. (2025, May 19). On the conversational persuasiveness of GPT-4. *Nature Human Behaviour.* https://www.nature.com/articles/s41562-025-02194-6
38. Rogers, D. (2024, May 17). Arup named as victim of £20m deepfake scam. *Building*. https://www.building.co.uk/news/arup-named-as-victim-of-20m-deepfake-scam/5129479.article
39. Fox, A. (2024, April 22). The principles for the future of warfare and stand-off warfare. *Landpower*. AUSA. https://www.ausa.org/publications/principles-future-warfare-and-stand-warfare
40. Martseniuk, Y., Partyka, A., & Kret, T. (2025). Research of artificial intelligence vulnerabilities and creation of a comprehensive security model for an organization. *Modern Information Protection, 1*(61), 206–218. https://doi.org/10.31673/2409-7292.2025.018929.

Ризикоорієнтований підхід до впровадження штучного інтелекту в сфері оборони України (концептуальні засади)

Станіслав Шумлянський
кандидат політичних наук
старший офіцер
відділу досліджень проблем реформування та розвитку Держспецтрансслужби
Центр досліджень Державної спеціальної служби транспорту
02094, вулиця Якова Гніздовського, 5, Київ, Україна
e-mail: stanislavshum@icloud.com
ORCID: 0009-0008-2739-7614

Євген Хоменко
аспірант
начальник
Центр досліджень Державної спеціальної служби транспорту
02094, вулиця Якова Гніздовського, 5, Київ, Україна
e-mail: centrd@dsst.gov.ua
ORCID: 0009-0006-7006-3439

Сергій Попрієнко
Офіцер
відділу досліджень проблем ОВТ, застосування безпілотних комплексів та систем РЕБ
Центр досліджень Державної спеціальної служби транспорту
02094, вулиця Якова Гніздовського, 5, Київ, Україна
e-mail: centrd@dsst.gov.ua
ORCID: 0009-0006-8517-1045

Анотація. У статті розглядаються концептуальні засади ризикоорієнтованого підходу до впровадження штучного інтелекту у сфері оборони України. Актуальність дослідження обумовлена швидким розвитком технологій штучного інтелекту у військовій справі та потребою протистояти ворогу, що переважає. У зв'язку з цим, дана стаття спрямована на виявлення та обґрунтування основних засад ризикоорієнтованого підходу до впровадження штучного інтелекту у військовій сфері України. Провідним методом для дослідження є аналіз, що дозволяє розглянути внутрішні нетехнічні ризики, оскільки вони є менш очевидними та потенційно більш небезпечними, ніж технічні та зовнішні ризики. У статті визначено та

класифіковано ключові нетехнічні ризики, що виникають при впровадженні штучного інтелекту у військовій сфері. Зокрема, ідентифіковано ризики, пов'язані з недооцінкою загроз, невизначеністю розвитку штучного інтелекту, негативним впливом на людину (зниженням когнітивних навичок, надмірною довірою), а також проблемами, пов'язаними з автоматичним перекладом та ухваленням рішень. Також розглянуто зовнішні ризики, що виникають при використанні штучного інтелекту противником, такі як посилення ефективності розвідки, кібервійни та пропаганди, а також ризики, пов'язані з “отруєнням” даних. Обґрунтовано доцільність створення реєстру та карти ризиків для подальшого управління ними. Матеріали статті становлять практичну цінність для розробки політики впровадження штучного інтелекту у військовій сфері, створення реєстру та мапи ризиків, а також для подальшої роботи з ними.

Ключові слова: штучний інтелект, ризики штучного інтелекту, ідентифікація ризиків, інформаційні технології, ризикоорієнтованість, інтелектуалізація війни, карта ризиків, реєстр ризиків.