Secure Systems Design and Technology Development. Veliko Tarnovo, Bulgaria
Proceedings of the International Scientific Conference, 2026. Volume I, 223-228

Risks and Mitigation Measures in the Integration of Artificial Intelligence into the Command Chains of the Armed Forces

Emil Ivanov
Ministry of Defence
Sofia, Bulgaria
emotsiv@gmail.com
ORCID: 0000-0003-1658-9588

***Abstract* - This paper examines the key risks associated with the integration of Artificial Intelligence (AI) and Machine Learning (ML) systems into military command chains. The analysis covers the main components of the decision-making process, including data, algorithms and models, software and system integration, communication infrastructure, and the human factor. The study emphasises that while AI offers significant advantages in accelerating analysis, enhancing situational awareness, and supporting operational planning, its deployment also introduces new vulnerabilities. These include risks related to low-quality or manipulated data, malicious model modifications, software dependency vulnerabilities, reliance on communication networks, and excessive trust in automated recommendations.**

Keywords* - *artificial intelligence, command and control, cybersecurity, risk management.

I. Introduction

The integration of AI and machine learning into modern military systems that support elements of the command chain represents one of the most significant technological transformations in defence. The growing use of AI in data analysis, intelligence, surveillance, and planning fundamentally changes command and control processes. In a fast-paced and high-risk environment, the ability to rapidly process large amounts of information and identify relevant dependencies becomes a key operational advantage for those who employ such technologies. AI systems support commanders through automated information extraction, anomaly detection, scenario forecasting, and recommendation generation. This shortens decision-making time and improves planning quality at every level of the hierarchical chain [1], [2].

Alongside these advantages, implementing AI introduces new challenges related to security, reliability, and control within command chains. The military environment is extremely sensitive to errors and manipulation, which means that AI integration must be viewed as a complex organisational and operational issue, as well as a cybersecurity problem.

Strategic documents and recent studies on AI governance, adversarial machine learning, and responsible AI emphasise the need for reliable, responsible, and controllable AI systems in the defence sector [3]–[8]. This means that their use must guarantee transparency, traceability, resilience against manipulation, and preservation of the leading role of humans in critical decisions [9], [10]. From this perspective, the question is not whether AI will be integrated into command chains, but how to do so in a way that utilises its advantages without jeopardising the security and effectiveness of the military system.

Despite the increasing number of studies on artificial intelligence in defence, there is a lack of a systematic analysis of risks within the hierarchical chains of command as a whole system, which motivates the present study. The main contribution of this paper is the proposal of a structured analytical framework for examining AI-related risks in military command chains. The framework links vulnerabilities across data, algorithms and models, software and system integration, communication infrastructure, and the human factor within the overall decision-making process. In this way, the paper provides a system-oriented view of AI integration risks rather than analysing them as isolated technical problems.

II. Materials and Methods

The methodology combines conceptual modelling, risk analysis, and a comparative analysis of traditional and AI-assisted processes in command chains. The main analytical tool is a model that represents the interaction among key components of the decision-making process. Figure 1 shows an abstraction that embeds these components into a single system. It does not show

Online ISSN 3044-7771
https://doi.org/10.68302/std2026.vol1.60

hierarchical positions but a system of functions and logical steps that every military structure follows - from intent to action.

Strategy (defines intent and objective) → Operations (translates intent into a plan) → Tactics (organises concrete actions) → Execution (real action on the ground) → Feedback

Fig. 1. Basic logic of the command chain

The analysis is based on further dividing this system into several main layers: data, models, Command and Control (C2) systems, human, and communication infrastructure. This strategy allows the recognition of potential vulnerability points and tracing how risk can propagate along the chain.

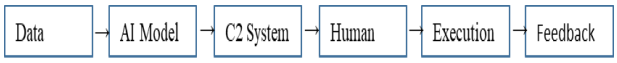

Fig. 2. Conceptual model of the command chain with AI

The analysis also includes an adaptation of the OODA loop, which assesses the influence of AI on the decision-making process. In this context, the Observe phase relates to data collection and processing, Orient to analysis and interpretation using AI models, Decide to decision-making involving both humans and AI, and Act to execution. This approach enables clearer tracking of AI's role in each phase and identification of specific risks.

III. Results and Discussion

AI integration into command chains shows that risks are not isolated but arise from interactions between different system components. A practical illustration of such interaction can be observed in Intelligence, Surveillance, and Reconnaissance (ISR) environments. In these settings, AI systems process data from multiple sources, such as Unmanned Aerial Vehicle (UAV) video feeds, satellite imagery, and ground-based sensors, in order to identify anomalies and detect potential threats. The processed information is integrated into a C2 system and presented to operators as part of the situational picture.

In practice, such systems do not operate under ideal conditions. Data may be incomplete, delayed, or affected by adversarial interference, which directly impacts the reliability of the AI-generated outputs. At the same time, disruptions in communication channels between ISR assets and the C2 system may lead to inconsistencies in the operational picture. This may result in delayed or incorrect assessments, especially in time-sensitive scenarios.

This example illustrates how risks propagate across the data, model, communication, and human layers of the system. It also highlights the importance of maintaining human oversight, as automated outputs must be interpreted within the broader operational context.

Based on that, the risks can be grouped into five main categories: data-related risks, model-related risks, software-related risks, communication infrastructure risks, and human-factor risks.

Table I Classification of Risks in AI-Enabled Command Chains

Technical	**Operational**	**Human**
Data	Communications	Training
Models	Delay	Trust
Software	Disruption	Interpretation
Cyberattacks	Information gaps	Responsibility

Risks can also be systematised through the OODA loop, where each phase introduces specific vulnerabilities.

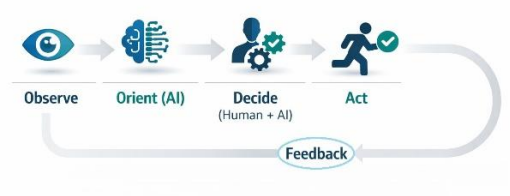

Fig. 3. OODA Loop with AI

A. Qualitative assessment of risk significance

In addition to the categorisation of risks, their significance can be assessed through a qualitative risk matrix combining estimated probability and potential operational impact. Although the assessment is conceptual and based on expert judgement supported by the analysed literature, it provides an indication of the relative criticality of different vulnerabilities in AI-enabled command chains. The qualitative levels are derived from the frequency and severity of risks discussed in the analysed literature and interpreted in the context of military command environments.

Table II. Qualitative Assessment of AI-Related Risks in Command Chains

Risk Category	**Estimated Probability**	**Operational Impact**	**Overall Risk Level**
Data poisoning	Medium-High	High	Critical
Model poisoning/ backdoor attacks	Medium	High	Significant
Communication disruption	High	Critical	Critical
Software vulnerabilities	Medium	Medium-High	Significant
Automation bias	Medium	High	Significant
Data obsolescence	High	Medium	Significant
Low model transparency	Medium	Medium-High	Moderate

As shown in Table II, communication disruptions and manipulated data represent among the most critical risks due to their combination of high probability and severe operational consequences. The assessment also indicates that software vulnerabilities, insufficient model transparency, and excessive reliance on automated outputs may significantly affect the reliability of military decision-making processes. Therefore, mitigation measures should prioritise vulnerabilities with both high likelihood and high operational impact, particularly in time-sensitive command environments.

B. Data-related risks

Data is the fundamental element of any AI system. In the context of command chains, its role is even more critical, as it forms the basis of situational awareness and supports decision-making.

One of the main risks concerns *data quality*. Incomplete, inaccurate, or biased data can lead to serious deviations in analysis and incorrect decisions [4]. Particularly dangerous are situations where models are trained on limited or inadequate datasets [2]. If the information base, on which a model is trained or fed, does not accurately reflect the operational environment, the system may generate misleading recommendations.

For example, a model trained on limited or one-sided data may fail to correctly recognise atypical but critical scenarios. This may lead to incorrect prioritisation, misjudgement of threats, or inaccurate identification of objects and actions relevant to the planning or execution of a military operation.

A major risk is *data poisoning*, where hackers introduce manipulated information into training or operational datasets to influence model behaviour [3]. Such manipulation can cause the system to underestimate certain signals, favour incorrect patterns, or generate recommendations that appear logical but are strategically inadequate.

This may cause systematic distortion of results and create a false operational picture. In a military environment, such manipulations may remain unnoticed until the system is used in a real operation, which is completely unacceptable.

Another essential aspect is classified information and sensitive *data protection*. Military systems process such information in general - intelligence data, operational plans, data on own and adversary forces and assets, communications, and disposition. All of this requires a higher level of security. The use of external platforms or shared resources may create a risk of information leakage or indirect data extraction through analysis of model behaviour, with strategic consequences [5]. Even when the raw data is not directly accessible, techniques such as model inversion and membership inference can be used to draw conclusions about the content of training data sets [4], thereby exposing information that must be protected.

The issue of operational data validity must also be considered. In a military operation, the rate of environmental change may be extremely high, meaning that data that was accurate hours or even minutes ago may already be outdated. If the AI system does not have reliable mechanisms for updating and verification, it may make a final decision based on an obsolete operational picture.

In this regard, data-related risks are not only technical. They can also directly affect the reliability of the command process and the integrity of the command chain. These risks require systematic measures for source verification, quality testing, anomaly detection, protection of sensitive and classified information, and limiting dependence on external or insufficiently transparent data streams.

C. Risks related to the algorithms and models

The models used in AI systems can be a source of significant vulnerabilities. One of the main problems is the transparency, especially the lack of it in complex models such as neural networks. The large algorithms and complex models are difficult to understand. They may provide an assessment or recommendation with high confidence, without it being clear which factors led to it. This makes it difficult to evaluate the reliability of the results and reduces operator trust. In the command structure, this is a serious limitation because the commander must not only receive a result but also have grounds to trust it or reject it. If the AI system cannot provide sufficiently clear logic behind its analysis, it risks becoming a “black box” that influences decisions without real possibilities for control [6], [8] and [10]. According to the U.S. Department of Defence, the lack of transparency in AI systems creates serious challenges for trust and control in command processes [8].

A serious threat is also posed by malicious model modifications, including model poisoning and backdoor attacks [3], [4]. In these scenarios, the model is compromised in such a way that under normal conditions, it functions as expected, but under certain input data or combinations of signals, it generates intentionally incorrect outputs. This is particularly dangerous because the compromise may remain hidden during standard testing and manifest itself at a critical moment.

D. Risks related to software and integration

AI solutions rarely exist as isolated modules. In practice, they belong to a broader software ecosystem that includes libraries, development frameworks, software dependencies, interfaces, databases, visualisation systems, and tools for communication with other platforms. It is precisely the complexity which increases the risk of vulnerabilities and compromise along the supply chain [3], [7]. If a software library contains a vulnerability, is maliciously modified, or becomes compromised during distribution, this may affect the final AI system as well. In environments where open-source tools are frequently used without sufficient verification, this risk is significant.

The integration of AI into C2 systems also represents a serious challenge. C2 systems are critical infrastructure for the armed forces, and the addition of AI components increases their complexity. Every new integration point (interface, API, data-exchange module, or external service) expands the attack surface. From here, an error or weakness

in a single component can lead to cascading failures across the entire C2 system.

The issue of compatibility and predictability must also be taken into account. Military systems are often built in different periods, by different manufacturers, and according to different standards. Adding AI functionalities to those environments may lead to unpredictable interactions, delays, data exchange errors, or disruption to existing control and protection mechanisms. Research by the NATO Cooperative Cyber Defence Centre of Excellence in the field of cybersecurity shows that AI components can become an independent vector for cyberattacks, especially when there is insufficient traceability of their origin and lifecycle [11]. In this sense, the security of AI cannot be separated from the security of the entire software architecture into which the AI is embedded.

E. Risks related to communication infrastructure

Communication infrastructure is also a critical element of command chains. Disruptions such as jamming, cyberattacks, and manipulation of the communication flow can lead to a distorted operational picture [5].

Many AI systems count on a continuous stream of input data, exchange between different levels of command, access to distributed databases, and synchronisation between multiple platforms and sensors, which makes them particularly vulnerable to such communication disturbances. There is also a risk of manipulation of the communication flow. If an adversary manages to inject false messages, alter information flow priorities, or introduce artificial noise into the communication link, this may affect both operators and AI systems. In the absence of adequate verification of the source and integrity of the data, this may lead to a distorted operational picture and to incorrect decision-making.

An additional problem is the dependence on external infrastructure - for example, cloud services, satellite channels, data centres, or inter-system connections controlled by third parties. In the military domain, such dependence may become a critical weakness if, during a crisis or conflict, access to these resources is restricted, compromised, or politically obstructed. Therefore, resilient communications are directly related to the reliability of AI.

F. Risks related to the human factor

Regardless of the degree of automation, the human factor remains a key element in command chains. The main risks include insufficient training, incorrect interpretation of results, and excessive trust in the systems.

The first issue concerns preparation and training. If operators and commanders do not sufficiently understand the logic, limitations, and conditions for using a given AI system, there is a risk of misinterpreting the results. They may accept a probabilistic assessment as a categorical fact, fail to recognise signs of anomalies, or use the system outside the scope for which it is intended.

An additional risk is excessive reliance on automated recommendations. The better an AI system performs in routine or well-structured situations, the more likely operators are to begin over-trusting it. This phenomenon, known as *automation bias*, can weaken critical thinking and lead to decisions being made by mechanically following recommendations without sufficient contextual assessment and verification, which contradicts the principles of responsible AI use [8], [9], [10].

The opposite risk is distrust of the system, especially if it has low transparency or has produced inconsistent results in the past. In such a case, AI may be formally integrated into the command process, but in practice, it is not used effectively.

There is also an issue with the distribution of responsibility, especially when decisions are partially generated by AI. When AI participates in the decision-making process, it must be clear who bears responsibility for the final decision. This requires a clear definition of roles and responsibilities. If roles are not clearly distinguished, there is a risk of organisational gaps in which no one assumes full responsibility for the consequences. According to analyses by RAND Corporation, the balance between automation and human involvement is critical for the reliability of military decisions [9]. This requires the creation of integrated measures that cover all levels of the system.

G. Comparison with existing studies and implications

Existing studies on artificial intelligence in defence and cybersecurity primarily focus on specific dimensions of AI adoption, including trustworthy AI, ethical principles, adversarial machine learning, software vulnerabilities, and autonomous decision-making [3], [6], [8], [10], [11], [14]–[17]. These works provide valuable insights into individual categories of risks but often address them separately or within broader AI governance frameworks.

In contrast, the present study proposes an integrated perspective by analysing risks across interconnected components of military command chains, including data, algorithms and models, software integration, communication infrastructure, and the human factor. This system-oriented approach allows identification of how vulnerabilities propagate through different layers of the decision-making process rather than emerging as isolated technical issues.

The qualitative risk assessment presented in Table 2 additionally supports prioritisation of vulnerabilities according to their estimated probability and operational impact. This contributes to a more structured understanding of AI-related risks in military environments and may support future development of quantitative risk evaluation methods.

Therefore, the findings suggest that resilience in AI-enabled command chains depends not only on technological robustness but also on maintaining transparency, human oversight, and adaptability across the entire operational system.

IV. Risk Mitigation Approaches

Effective and sufficient mitigation of the risks associated with integrating AI into command chains demands addressing vulnerabilities at both the system elements and across the decision-making cycle. The measures should cover all components and ensure technological resilience plus organisational reliability.

A. Data management and protection

Data quality is a fundamental requirement for the reliability of AI systems. It is necessary to establish strict data-management policies [3], [4], which include:

- verification of sources;
- validation of input data;
- detection of inconsistencies and anomalies;
- protection against manipulation (data poisoning).

Special attention should be paid to the processing of classified information. It is necessary to apply mechanisms for access control, cryptographic protection, and minimisation of exposed data.

B. Validation and transparency of models

AI models must undergo systematic validation, which includes:

- testing under different scenarios;
- checking resilience of the model against adversarial attacks;
- assessment of reliability and stability;
- traceability of the AI training process.

Regular testing and the use of explainable models increase trust and allow better control over decisions [6]. Where possible, methods for AI transparency (Explainable AI - XAI) should be used, as they allow better understanding of the logic behind the decisions being made. Even when complex models offer better performance, in situations involving critically important decisions, it is important for the commander to have at least some partially understandable logic explaining why a particular recommendation is given.

C. Software and integration security

The security of the software environment and the management of software risk require:

- control over the software dependencies, libraries, and external components used [7];
- regular vulnerability checks;
- the use of secure environments for development and deployment;
- clear separation between test and operational systems.

The integration of AI into existing military C2 systems must be gradual and carefully controlled, with minimisation of the attack surface. It is necessary to clearly separate tests and operations, limit the permissions of AI modules to the minimum required, and to continuously monitor their behaviour after deployment.

D. Communication resilience

The communication infrastructure should be designed with a high degree of resilience [5], including:

- redundant channels and cryptographic protection;
- protection against jamming and cyberattacks;
- mechanisms for verifying the authenticity of messages;
- the ability to operate in a degraded communication environment.

AI systems must be able to function under partial data loss without generating misleading results. AI in command chains should be designed in such a way that it can operate under degraded communications and partial loss of connectivity, so that it does not become a source of uncertainty.

E. The human in the decision-making process

In accordance with the principles of NATO and the U.S. Department of Defence, the application of human-in-the-loop is a key requirement for critical decisions [5], [8], [10], [11], [12] and [13]. This principle is also important for ensuring the controllability of the system, and this includes:

- maintaining human control in critical decisions;
- training personnel to work with AI;
- development of critical thinking;
- clear definition of responsibilities.

AI should be viewed as a tool for support, not as a replacement for human final judgment. The human factor should not be seen as a weakness that must be “replaced” by automation, but as a key element that must be supported through training, clear definitions and procedures, and developed critical thinking, alongside proper positioning of AI as an assisting instrument.

V. Conclusions

The integration of AI into the command chains of the armed forces provides significant opportunities for increasing efficiency, accelerating analysis, and improving situational awareness. These advantages, however, are accompanied by complex risks that affect all components of the system.

The analysis shows that the most significant vulnerabilities are related to data quality, model trustworthiness, software security, communication resilience, and the role of the human factor. The interaction between these elements determines the overall resilience of the command chain.

Successful integration of AI requires a balanced approach that combines technological capabilities with clear mechanisms for control, transparency, and

responsibility. It is particularly important to preserve the leading role of the human [10] in the decision-making process, using AI capabilities for support rather than replacement.

The present study contributes by proposing a structured framework for analysing AI-related risks in military command chains. This framework may serve as a basis for developing standardised risk-assessment approaches and future quantitative evaluation methods for AI-enabled defence systems.

REFERENCES

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson, 2021.

[2] I. Goodfellow et al., *Deep Learning*, MIT Press, 2016.

[3] Australian Cyber Security Centre, "Artificial Intelligence and Machine Learning Supply Chain Risks," 2025. [Online]. Available: https://www.cyber.gov.au/

[4] National Institute of Standards and Technology (NIST), "Adversarial Machine Learning: A Taxonomy and Terminology," 2025. [Online]. Available: https://www.nist.gov/

[5] NATO, "Summary of NATO's revised Artificial Intelligence (AI) strategy" 2024. [Online]. Available: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy

[6] European Commission, "Ethics Guidelines for Trustworthy Artificial Intelligence," 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/

[7] OWASP Foundation, "ML Security Top 10," 2024. [Online]. Available: https://owasp.org/www-project-machine-learning-security-top-10/

[8] U.S. Department of Defense, "Responsible Artificial Intelligence Strategy and Implementation Pathway," 2022. [Online]. Available: https://media.defense.gov/2024/Oct/26/2003571790/-1/-1/0/2024-06-RAI-STRATEGY-IMPLEMENTATION-PATHWAY.PDF

[9] RAND Corporation, "Artificial Intelligence and International Security," 2020. [Online]. Available: https://www.rand.org/

[10] U.S. Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence" 2020. [Online]. Available: https://www.war.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

[11] NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE), "Artificial Intelligence, Cybersecurity and Warfare," 2023. [Online]. Available: https://www.ccdcoe.org/uploads/doc/CyCon_2023_book_print.pdf

[12] NATO, "Summary of the NATO Artificial Intelligence Strategy", 2021, [Online]. Available: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy

[13] NATO Data and Artificial Intelligence Review Board (DARB), "Framework Documents," 2023. [Online]. Available: https://cdn.asp.events/CLIENT_Defence__8EE24275_D70D_4386_936BB8991B847FF8/sites/Navy-Leaders-2022/media/libraries/cne24-presentations/Day-3-UDSA-1115-OPTION-1-_PUBLIC_CNE_NATO_Data_and_AI_Policies.pdf

[14] G. Falco et al., "Cybersecurity Risks in Artificial Intelligence Systems," Computer, vol. 54, no. 6, pp. 84–88, 2021.

[15] A. Taddeo and L. Floridi, "How Artificial Intelligence can be a Force for Good in Defence," Science and Engineering Ethics, vol. 24, no. 6, pp. 1–15, 2020.

[16] M. Brundage et al., "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims," arXiv, 2020.

[17] M. Cummings, "Artificial Intelligence and the Future of Warfare," Chatham House Research Paper, 2021.