Markov's Models of AI Systems Availability Considering Re-learning Processes[*]

Vyacheslav Kharchenko[1,*,†], Yuriy Ponochovnyi[2,†] and Heorhii Zemlianko[1,†]

[1] *National Aerospace University KhAI, Kharkiv, Ukraine*
[2] *Poltava State Agrarian University, Poltava, Ukraine*

Abstract

The article is devoted to the development of Markov models for assessing the readiness of artificial intelligence (AI) systems in critical areas, taking into account retraining procedures. A conceptual model of an information and control system with AI (AI-ICS) is proposed, which includes a state space, failures, and maintenance procedures, such as online verification and retraining. A feature of the developed single- and multi-fragment Markov models is that they allow for the assessment of AI-ICS readiness, taking into account various parameters, both traditional for software and hardware systems (failure and recovery rates), and parameters of planned and reactive retraining processes and the resulting change in the corresponding system indicators. It is shown that the multi-fragment model surpasses the single-fragment one in accuracy, demonstrating the ability to account for adaptation through retraining. Prospects for future research are discussed, including two-version structures that increase safety by reducing Common Cause Failures Risks, and the development of diversification technologies in the creation of AI.

Keywords

artificial intelligence, reliability, pre-learning, Markov models, two-version system, diversity, security, von Neumann paradigm

1. Introduction

1.1. Motivation and related works

Systems of artificial intelligence (AI) play a key role in critical areas, where they provide automation of complex processes, real-time analysis of large volumes of data, and decision-making under uncertainty. In medicine, AI is used for diagnosing diseases based on medical images and predicting pandemics; in transport, for controlling autonomous vehicles; in the energy sector, for optimizing resource distribution and conducting predictive maintenance; and in the defense and security sectors, for threat analysis, risk prediction, decision-making support, and humanitarian demining [1-3]. However, the dependability of these systems remains a serious challenge due to the certain imperfection of AI tools, given the vulnerability of their components and insufficient resilience to specific interferences, etc. Hardware may experience physical failures and degrade due to equipment failures as a result of aging processes or external influences. Software can lead to AI system failures due to design faults, and AI models can lose trustworthiness due to data drift, incorrect training, the limitations of the datasets on which they are trained, and insufficient resilience to cyberattacks, including so-called AI-powered attacks [4-6].

Errors in such systems can have catastrophic consequences, including loss of human lives, power grid blackouts with significant economic losses, or security breaches in defense systems due to erroneous decisions or cyberattacks [7, 8]. To address these problems, a modernization of the von Neumann paradigm (VNP) and its components [9] has been proposed by creating trustworthy AI systems from untrustworthy AI components through the use of the diversity principle (diverse

[*] *ProfIT AI'25: 5th International Workshop of IT-professionals on Artificial Intelligence, October 15–17, 2025, Liverpool, UK*
[1*] Corresponding author.
[†] These authors contributed equally.

v.kharchenko@csn.khai.edu (V. Kharchenko); yuriy.ponch@gmail.com (Y. Ponochovnyi); g.zemlynko@csn.khai.edu (H. Zemlianko)
0000-0001-5352-077X (V. Kharchenko); 0000-0002-6856-2013 (Y. Ponochovnyi); 0000-0003-4153-7608 (H. Zemlianko)

model and data architectures) and redundancy (component reservation) [10]. This approach, historically developed for software and hardware systems, particularly safety-critical instrumentation and control systems of NPPs (reactor trip systems), has evolved to modern AI systems, where diversity and redundancy ensure dependability and resilience in conditions of changing operating modes, and cyber and physical environment parameters.

Modern research is focused on increasing the protection of AI from attacks and model anomalies. In [11], taxonomy of AI resilience factors is proposed, and in [12], aspects of protection from cyberattacks in autonomous transport systems are investigated. The diversity of neural architectures, as in hybrid neural networks [13], and bio-inspired approaches [14] contribute to AI adaptability. Retraining and continuous learning, described in [15, 16], allow systems to adapt to new conditions. At the same time, the ethical and legal aspects reviewed in [17, 18] emphasize the need to provide a certain level of responsibility ensuring explainability and trustworthiness AI safety in autonomous vehicles [19] and critical infrastructure [20] requires new methods for penetration testing, demonstrative verification [21], and protection from attacks [22], which provide objective quantitative and qualitative assessment.

In general, a certain deficit of mathematical models for AI systems can be concluded, which allow the calculation of complex availability indicators that take into account various parameters of the AI itself and training processes, as well as software and hardware platforms, enabling the investigation of the dependencies of these indicators on parameter changes and the formation of recommendations for ensuring compliance with the requirements for such systems.

1.2. Objectives and approach

The goal of the study is the development of Markov models for assessing the readiness of non-redundant AI systems, taking into account changes in parameters due to retraining and the formation of recommendations for increasing readiness and dependability.

Objectives are the following:

- the development of a conceptual model of an information and control system with AI (AI-ICS) (section 2), which takes into account the features of the failure and recovery processes of its components and the system as a whole;
- the development and study of Markov models of the AI-ICS (section 3);
- the analysis of the modeling results of the respective advantages and limitations in the use of models, as well as the formulation of recommendations for the selection of system parameters to increase readiness (section 4);
- the determination of the main contribution, results, and directions for further research (section 5).

The research methodology is based on:

- the consideration of the AI-ICS as a set of the model part of artificial intelligence, its software and hardware implementation, and the environment with cyber-physical effects on different system components;
- the detailing of the AI-ICS failure and recovery model, taking into account the main factors and types of faults;
- the use of single- and multi-fragment Markov models that allow for the consideration of parameter changes during training and possible improvement of the model part of the AI system due to online verification processes.

2. AI-ICS conceptual model

A conceptual model of an information and control system with AI (AI-ICS) is based on a software and hardware platform and an AI model that performs control functions (Figure 1). It describes the system's state space, which includes working states S_w, failure states S_f, and maintenance states S_m:

$$S = S_w \cup S_f \cup S_m, \tag{1}$$

where $S_w = \{S_q, S_1^{'}\}, S_f = \{S_{HW}, S_{SW}, S_{AI}\}, S_m = \{S_{OV}, S_{RL}\}$.

Let's note that:

- S_{HW}, S_{SW}, S_{AI} – are the failure states of the hardware, software, and AI model;
- S_{OV} – are the online verification states;
- S_{RL} – are the retraining states.

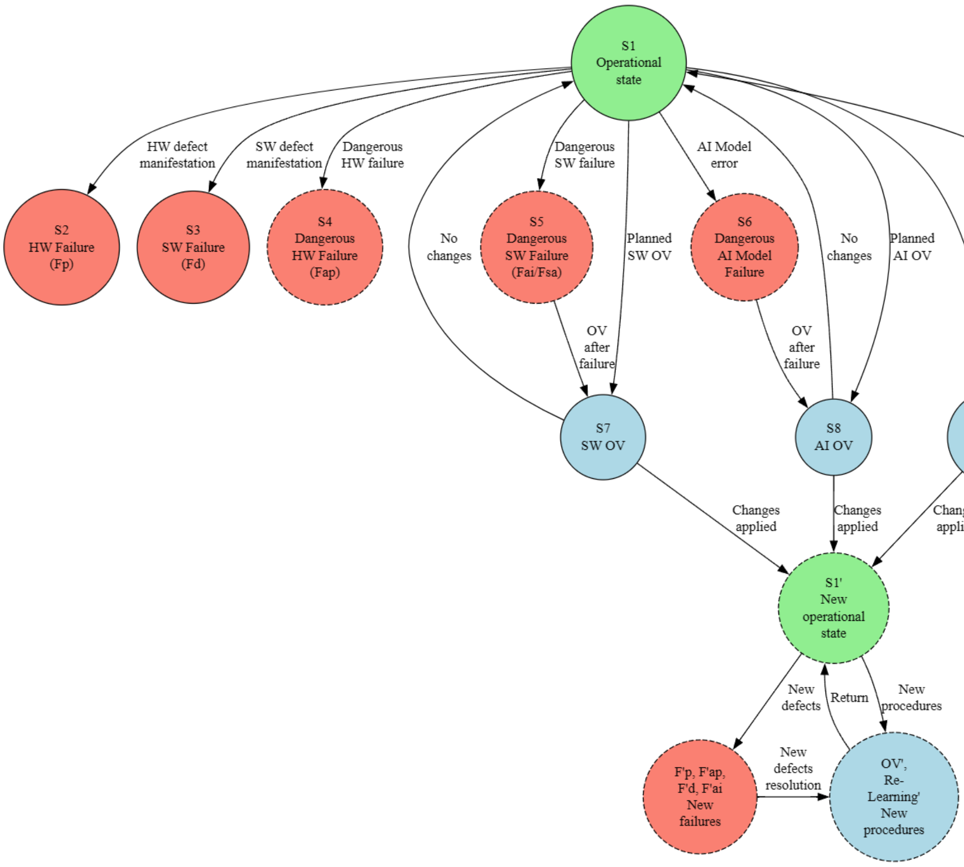


Figure 1: AI-ICS state graph describing transitions between operational states (S_1), failure states (S_2-S_6), and maintenance (S_7-S_9)

Online verification (OV) checks the system's compliance with requirements in real-world conditions, while Retraining (Re-Learning) adapts the AI model to new conditions or eliminates errors. Transitions between states are modeled as a mapping T: S⟶S, where the failure rates (λ_{HW}, λ_{SW}, λ_{AI}) and recovery rates (μ_{HW}, μ_{SW}, μ_{AI}) determine the system's reliability dynamics.

3. Markov models AI-ICS

For the study of AI systems, the functioning of the so-called single-version architectures was considered, which allows to simplify the models at the initial stage of the study and obtain adequate values of the input parameters of the model from today's available sources. Markov models have been developed to analyze the readiness of a single-version AI system, which take into account planned and reactive additional training.

3.1. One-fragment model

The state space of a single-fragment model (Figure 2) includes the following states:

- S_0: full operability;
- S_1: partial performance (eg reduced accuracy);
- S_2: incapacity (erroneous decisions);
- S_3: pre-training.

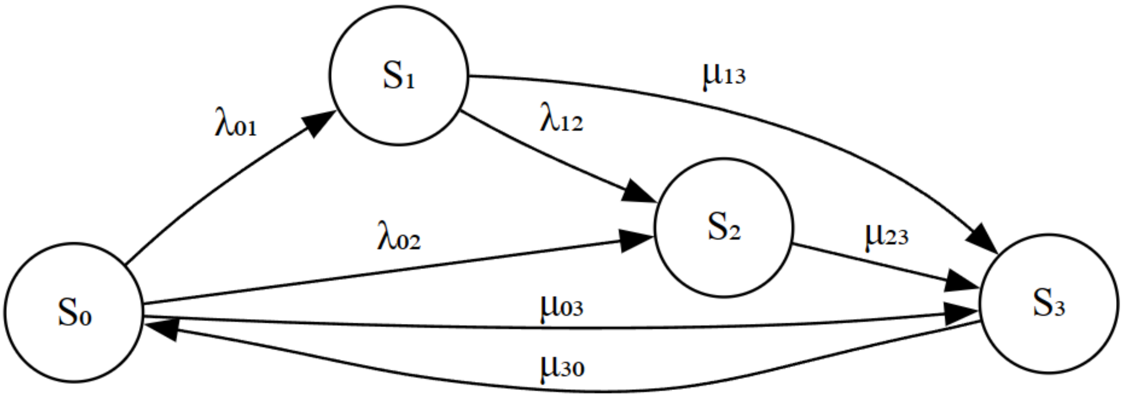


Figure 2: State graph of a single-fragment model

Transitions between states are described by intensities:

- λ_{01}: (degradation);
- λ_{02}: (critical rejection);
- λ_{12}: (deterioration);
- μ_{03}: (planned additional training);
- μ_{13}, μ_{23}: (reactive retraining);
- μ_{30}: (restoration).

The corresponding system of Kolmogorov-Chapman differential equations has the form:

$$\begin{gathered}
\frac{dP_0(t)}{dt} = -(\lambda_{01}+\lambda_{02}+\mu_{03})P_0(t)+\mu_{30}P_3(t),\\
\frac{dP_1(t)}{dt} = \lambda_{01}P_0(t)-(\lambda_{12}+\mu_{13})P_1(t),\\
\frac{dP_2(t)}{dt} = \lambda_{02}P_0(t)+\lambda_{12}P_1(t)-\mu_{23}P_2(t),\\
\frac{dP_3(t)}{dt} = \mu_{03}P_0(t)+\mu_{13}P_1(t)+\mu_{23}P_2(t)-\mu_{30}P_3(t),
\end{gathered} \tag{2}$$

with the conditions that $P_0(t = 0) = 1$ and for any moment t:

$$P_0(t)+P_1(t)+P_2(t)+P_3(t)=1, \tag{3}$$

Availability function can be presented by following formula:

$$A(t)=P_0(t)+P_1(t). \tag{4}$$

The parameters for conducting the simulation are provided by Table 1.

Table 1
Parameters of a single-fragment model

Parameter	Value (1/y)	Description
λ_{01}	0.01	Data drift
λ_{02}	0.005	Critical failures
λ_{12}	0.02	Deterioration of the condition
μ_{03}	0.033	Planned additional training (1 time/month)
μ_{13}	0.1	Reactive pre-training (10 hours)
μ_{23}	0.2	Reactive pre-training (5 hours)
μ_{30}	0.5	Recovery (2 hours)

3.2. Multi-fragment model

The multi-fragment model (Figure 3) expands the system's state space by adding a new post-retraining fragment (S_5-S_9) and the probability D_p=0.8 of successful scheduled retraining. Parameters λ_{56}', λ_{57}', λ_{67}' in the second fragment are reduced due to model updates (Table 2).

Figures should be centered, and their captions should be placed below them.

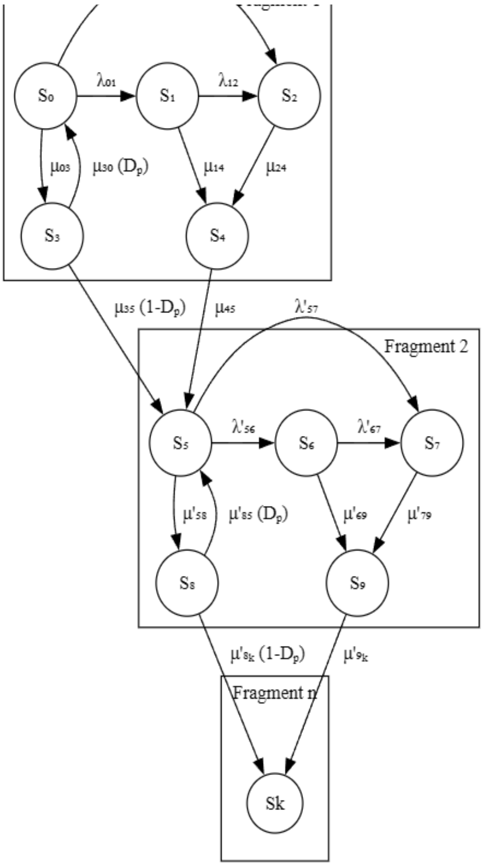


Figure 3: Graph of a two-fragment model

Table 2
Parameters of a multi-fragment model

Parameter	Value (1/y)	Description
λ_{01}	[0.01, 0.007, 0.006]	Data drift
λ_{02}	[0.005, 0.005, 0.005]	Critical failures
λ_{12}	[0.02, 0.017, 0.015]	Deterioration of the condition
μ_{03}	0.033	Planned additional training (1 time/month)
μ_{14}	0.1	Reactive pre-training (10 hours)
μ_{24}	0.2	Reactive pre-training (5 hours)
μ_{30}, μ_{35}, μ_{45}	0.5	Recovery (2 hours)
D_p	0.8	Success of planned pre-education

Thus, the multi-fragment model takes into account the separation of pre-learning states into for scheduled pre-learning, S_4 for reactive pre-learning in Fragment 1 (analogously, S_8 and S_9 in Fragment 2). After reactive retraining ($S_4 \longrightarrow S_5$), the system switches to Fragment 2, where the parameters λ_{56}', λ_{57}', λ_{67}' differ from λ_{01}, λ_{02}, λ_{12} due to model updates. After routine retraining With

S_3, the system returns to S_0 with probability D_p or moves to S_5 with probability $1-D_p$. Fragment 2 follows the structure of Fragment 1, but with new parameter values reflecting the effect of pre-learning.

4. Results of modeling and discussion

The single-fragment Markov model was implemented in MATLAB using the function fM1.m, which constructs vertex matrix V (defining states S_0 to S_3 with coordinates and colors for visualization) and edge matrix E (defining transitions with intensities λ_{01}, λ_{02}, λ_{12}, μ_{03}, μ_{13}, μ_{23}, μ_{30}). The script m_01.m sets global parameters (Table 1), builds the transition matrix A via matrixA.m, solves the Kolmogorov differential equations using ode15s with the stiffness handler stiff.m over a 100-hour interval, computes availability A(t) as (4), and plots A(t) along with individual state probabilities.

As a result of the investigating the first model by Matlab, it was found that (Figure 4):

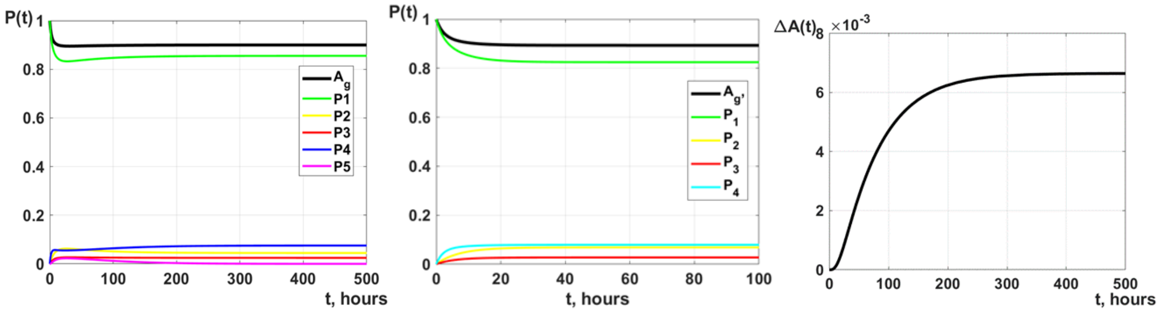
- the availability function A(t) drops rapidly from 1 to 0.919 in the first 5 hours due to high frequencies of transitions to the pre-learning state (S_3) and system degradation. In the future, A(t) stabilizes at the level of 0.893, with the system spending 82.5% of the time in the state of full working capacity (S_0), 6.9% – in partial working capacity (S_1), 2.7% – in incapacity (S_2) and 7.9% – in additional training (S_3);
- the relatively high probability of entering the S_3 state is due to planned additional training (μ_{03}=0.033), which significantly reduces availability. Reactive retraining (μ_{13}=0.1, μ_{23}=0.2) effectively reduces time in S_1 and S_2, but does not compensate for losses from S_3;
- to improve performance, it is recommended to reduce the frequency of scheduled retraining or speed up the recovery process.

Thus, the Markov model for a retraining AI system provides a methodological basis for analyzing its behavior, taking into account both planned and reactive adaptation strategies, but at the same time does not allow to assess the reliability of the AI system under the conditions of making changes to the system during the retraining process.

The multi-fragment model was implemented in MATLAB using the function fM2.m, which extends the single-fragment approach to n fragments (determined by array sizes of λ_{01}, λ_{02}, λ_{12}), constructing expanded V and E matrices with 5 states per intermediate fragment (S_0-S_4 or equivalents) and 4 for the last, incorporating probability D_p for branching in retraining transitions (μ_{30}, μ_{35}, μ_{45}). The script m_02.m sets array-based parameters for 3 fragments, builds A via matrixA.m, solves the ODE system using ode15s with stiff.m over a 500-hour interval, computes aggregated availability Ag(t) as the sum of probabilities for full/partial operability states, groups probabilities into P_1-P_5 for state categories, and plots Ag(t) with these groups. For comparison, m_02_1.m computes and plots the difference ΔA(t) between multi- and single-fragment availabilities. As a result of modeling the multi-fragment model (Figure 4), it was found that:

- the availability factor gradually decreases to 0.916 for the first 5.56 hours of operation. The rate of decrease of A(t) slows down with time, and then, at t = 500, it stabilizes at the level of 0.899, which indicates the achievement of a stationary state of the system;
- the sum of the probabilities of being in the states of full operability (S_1, S_6, S_{11}) decreases to 0.855 at t = 500, which is 85.5% of the initial value. This indicates that the system retains its initial operability for a significant part of the time, although it gradually loses it due to transitions to other states;
- the sum of the probabilities of being in the states of partial operability (S_2, S_7, S_{12}) increases to 0.045 (4.5% of the maximum value) at t = 500, indicating a limited time of the system in partially operable states. The system is completely inoperable for a small fraction of the time (0.025, or 2.5%);
- the sum of the probabilities of being in the states of planned retraining (S_4, S_9, S_{14}) demonstrates a significant fraction of the time spent on retraining or adaptation processes (0.075, or 7.5%). We note that the additional states that model the processes of reactive

retraining show an extremely low probability of the system being in these states (2.61×10^{-5}), associated with rare events.



a) b) c)

Figure 4: Modeling results: (a) change of A(t) for single-fragment model; (b) change of A(t) for multi-fragment model; (c) comparison of A(t) single- and multi-fragment models

In addition, we conclude that the high proportion of (S_4, S_9, S_{14}) (7.53%) can be due to intensive planned retraining processes with a frequency of μ_{03} = 0.033. This significantly affects the overall availability of the modeled system, reducing A(t). The states that reflect partial operability and inoperability remain at relatively low levels due to fast reactive processes (μ_{13} = 0.1 and μ_{23} = 0.2), which effectively reduce the time the system spends in these states.

Thus, retraining processes are the dominant factor limiting the system readiness. It is clear that the single-fragment model is simpler and predicts the system behavior faster, but underestimates the readiness in the long term due to the generalized approach.

The multi-fragment model, taking into account the dynamics of the parameters, provides higher accuracy and better adaptation to real conditions, which is confirmed by the higher value of A(t) in the steady state. For AI systems, where detail and long-term stability are important, the multi-fragment model is a better choice, although it requires more complex tuning.

5. Conclusions

The main contribution of the research is suggested Markov's models and results of their investigation. These models allow describing processes of re-learning and changing AI-ICS parameters that impact on system availability. Due to theses models can be improved accuracy of availability assessment.

The proposed Markov models allow assessing the readiness of AI-ICS, taking into account various parameters, both traditional for software-hardware systems, and parameters of re-learning processes and the resulting change in the corresponding indicators.

The multi-fragment model exceeds the single-fragment model in accuracy (A(t)=0.89999 versus 0.893348), demonstrating the possibility of taking into account adaptation through re-learning.

Future research can be aimed at:

- first, detailing Markov models by taking into account more complex failure scenarios due to cyberattacks or hardware degradation, which will increase the accuracy of predicting the behavior of AI systems;
- second, developing requirements and substantiating quantitative values for AI characteristics such as ethics and legality. These steps will contribute to the development of AI Safe and Secure Systems Engineering as a separate discipline that meets the current needs of critical industries;
- third, researching two-version structures that increase safety by reducing Common Cause Failures Risks, but their implementation is complicated by the need to develop diversification technologies when creating AI.

Acknowledgements

This research was partly supported by project «Dependability and resilience of intelligent UAV-USV complexes with combined application strategies», 0124U000945, MES of Ukraine.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] A. Fathollahi, "Machine learning and artificial intelligence techniques in smart grids stability analysis: A Review," Energies, vol. 18, no. 13, p. 3431, Jun. 2025. doi:10.3390/en18133431.

[2] D. Chumachenko et al., "Methodology for assessing the impact of emergencies on the spread of infectious diseases," Radioelectronic and Computer Systems, vol. 2024, no. 3, pp. 6–26, Aug. 2024. doi:10.32620/reks.2024.3.01.

[3] G. Fedorenko, H. Fesenko, V. Kharchenko, I. Kliushnikov, and I. Tolkunov, "Robotic-biological systems for detection and identification of explosive ordnance: Concept, general structure, and Models," Radioelectronic and Computer Systems, no. 2, pp. 143–159, May 2023. doi:10.32620/reks.2023.2.12.

[4] V. S. Nallapareddy and S. K. Katta, "AI-enhanced cyber security proactive threat detection and Response Systems," 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), pp. 1510–1514, Feb. 2025. doi:10.1109/icsadl65848.2025.10933436.

[5] E. N. Amora, J. C. Agoylo, J. A. Olaybar, J. C. Munasque, and P. D. Cerna, "Ai-driven real-time severity prediction for cyber attacks using machine learning," 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), pp. 1467–1472, Jun. 2025. doi:10.1109/icssas66150.2025.11081230.

[6] Y. Pushkarenko and V. Zaslavskyi, "Research on the state of areas in Ukraine affected by military actions based on remote sensing data and Deep Learning Architectures," Radioelectronic and Computer Systems, vol. 2024, no. 2, pp. 5–18, Apr. 2024. doi:10.32620/reks.2024.2.01

[7] A. M. Rahmani, B. Rezazadeh, M. Haghparast, W.-C. Chang, and S. G. Ting, "Applications of artificial intelligence in the economy, including applications in stock trading, market analysis, and Risk Management," IEEE Access, vol. 11, pp. 80769–80793, 2023. doi:10.1109/access.2023.3300036

[8] V. Moskalenko, V. Kharchenko, A. Moskalenko, and B. Kuzikov, "Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, models and methods," Algorithms, vol. 16, no. 3, p. 165, Mar. 2023. doi:10.3390/a16030165

[9] V. Kharchenko, O. Odarushchenko, Trustworthy AI Systems from Untrustworthy Components: Development von Neumann's Paradigm using Principle of Diversity. In Proceedings of the 4th International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2024), Cambridge, USA, Sept. 25-27, 2024, pp. 392-404.

[10] Diversity Strategies for Nuclear Power Plant Instrumentation and Control Systems, URL: https://www.nrc.gov/docs/ML1005/ML100541256.pdf

[11] O. Illiashenko, V. Kharchenko, I. Babeshko, H. Fesenko, and F. Di Giandomenico, "Security-informed safety analysis of autonomous transport systems considering AI-powered cyberattacks and protection," Entropy, vol. 25, no. 8, p. 1123, Jul. 2023. doi:10.3390/e25081123

[12] F. Mirzarazi, S. Danishvar, and A. Mousavi, "The safety risks of AI-driven solutions in Autonomous Road Vehicles," World Electric Vehicle Journal, vol. 15, no. 10, p. 438, Sep. 2024. doi:10.3390/wevj15100438

[13] A. Yanko, V. Krasnobayev, A. Hlushko, and S. Goncharenko, "Neurocomputer operating in the Residue Class System," Advanced Information Systems, vol. 9, no. 2, pp. 84–92, Apr. 2025. doi:10.20998/2522-9052.2025.2.11

[14] R. Naidu, "Bio-inspired AI Architectures: Lessons from nature for building next-Generation Learning Systems," IJARIIE, URL: https://surl.li/daolxs

[15] M. Joshi, "Adaptive Learning through Artificial Intelligence," SSRN Electronic Journal, 2023. doi:10.2139/ssrn.4514887

[16] J. Mendonça, F. Machida, and M. Völp, "Enhancing the reliability of perception systems using N-version programming and rejuvenation," 2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W), pp. 149–156, Jun. 2023. doi:10.1109/dsn-w58399.2023.00044

[17] B. K. Konidena, J. N. Malaiyappan, and A. Tadimarri, "Ethical considerations in the development and deployment of AI Systems," European Journal of Technology, vol. 8, no. 2, pp. 41–53, Mar. 2024. doi:10.47672/ejt.1890

[18] S. T. Mortaji and M. E. Sadeghi, "Assessing the reliability of Artificial Intelligence Systems: Challenges, metrics, and future directions," International Journal of Innovation in Management, Economics and Social Sciences, vol. 4, no. 2, pp. 1–13, Jun. 2024. doi:10.59615/ijimes.4.2.1

[19] J. Perez-Cerrolaza et al., "Artificial Intelligence for safety-critical systems in industrial and Transportation Domains: A survey," ACM Computing Surveys, vol. 56, no. 7, pp. 1–40, Apr. 2024. doi:10.1145/3626314

[20] M. Al-Hawawreh, Z. Baig, and S. Zeadally, "AI for Critical Infrastructure Security: Concepts, challenges, and future directions," IEEE Internet of Things Magazine, vol. 7, no. 4, pp. 136–142, Jul. 2024. doi:10.1109/iotm.001.2300181

[21] F. Wotawa, On the use of available testing methods for verification. In Proceedings of the Workshop on Artificial Intelligence Safety 2021 (SafeAI 2021). February 8, 2021. pp. 1-6. URL: https://ceur-ws.org/Vol-2808/Paper_29.pdf

[22] G. G. Shayea, M. H. Zabil, M. A. Habeeb, Y. L. Khaleel, and A. S. Albahri, "Strategies for protection against adversarial attacks in AI models: An in-depth review," Journal of Intelligent Systems, vol. 34, no. 1, Jan. 2025. doi:10.1515/jisys-2024-0277