DOI: https://doi.org/10.30525/2256-0742/2025-11-2-273-286

INTEGRATIVE APPROACH TO THE ANALYSIS, MODELING, AND ENSURING CYBER SECURITY OF CRITICAL INFORMATION INFRASTRUCTURE UNDER MODERN THREATS

Valentyna Yashchuk[1], Yulia Demyanchuk[2], Victoria Savitska[3]

Abstract. The present study explores contemporary approaches to the analysis, modelling and assurance of cyber security of critical information infrastructure in the face of modern threats. This text focuses on the utilisation of intelligent methods and advanced technologies for the protection of critical information infrastructure (CII). The objective of the present study is to employ an integrative approach to analysing, modelling, and ensuring the cyber security of CII in the face of contemporary threats. The methodological foundation of the study is a comprehensive analysis of scientific literature dedicated to the application of intelligent methods and technologies for the protection of CII. This includes both fundamental theoretical developments and practical aspects of implementing modern cyber security approaches, as well as experimental modelling of fraud detection processes in the CII of the financial sector. This modelling was conducted by the authors using artificial intelligence (AI) methods. A comparative analysis of concepts proposed by modern researchers is given particular emphasis, with the objective of identifying the main trends, development prospects, and potential areas for improving existing CII protection systems. In relation to the outcomes of the extension of scientific sources, it was determined that intelligent methodologies founded on machine learning and AI represent pivotal technologies for the effective safeguarding of critical infrastructure within the financial sector against contemporary cyberattacks and security threats. The results of the modelling of fraud detection processes in the CII of the financial sector of the economy allow an assessment of the effectiveness of the methods used, determination of their advantages and limitations, and formulation of recommendations for further improvement of cyber protection systems for the CII of the financial sector of the economy. The employment of an integrative approach, encompassing threat analysis, simulation modelling, artificial intelligence, and contemporary cyber defence technologies, facilitates the enhancement of the security of critical infrastructure and the effective response to emergent threats. The proposed solutions have the potential to facilitate the development of highly effective cyber defence systems in various critical infrastructure industries, including the financial, energy, and government sectors.

Keywords: critical information infrastructure, cyber security, information security, intelligent learning methods, artificial intelligence, machine learning, neural networks.

JEL Classification: O18, F52

1. Introduction

The Critical Information Infrastructure is the basis for the functioning of modern government, industrial and financial systems. The perpetual escalation in the intricacy and magnitude of cyberattacks necessitates the formulation and execution of efficacious protective measures. In this regard, intelligent methods, such as machine learning, AI, and computational intelligence, which allow for the automation of monitoring, detection, and response to threats, are of particular interest.

The CII of the financial sector plays a pivotal role in ensuring the stable functioning of economic systems and the financial stability of states. It is an irrefutable fact that the disruption of the operation of banking systems, payment services, stock exchanges and

[1] Lviv State University of Life Safety, Ukraine
ORCID: https://orcid.org/0000-0003-2651-2019
[2] Lviv State University of Life Safety, Ukraine *(corresponding author)*
E-mail: y.demianchuk@gmail.com
ORCID: https://orcid.org/0000-0001-8722-5568
[3] Taras Shevchenko National University of Kyiv, Ukraine
E-mail: v.v.savitska@gmail.com
ORCID: https://orcid.org/0000-0002-1403-125X

This is an Open Access article, distributed under the terms of the Creative Commons Attribution CC BY 4.0

other financial organisations can have catastrophic consequences for societal well-being. Concurrently, the accelerating digital transformation of the financial sector and the pervasive implementation of advanced technologies are generating novel vulnerabilities for cyber threats.

Modern cyberattacks are becoming increasingly sophisticated. They employ methods such as social engineering, the exploitation of unknown zero-day vulnerabilities, distributed botnets and automated network scanning tools. Traditional defence mechanisms based on signature analysis and rule-based systems are often ineffective against new, previously unknown threats. This creates an urgent need for more flexible and intelligent cyber security systems.

The advent of artificial intelligence (AI), machine learning, and data mining techniques has engendered a paradigm shift in the realm of safeguarding critical infrastructure within the financial sector. These technologies facilitate the automated detection of anomalies and indicators of cyberattacks within substantial, heterogeneous data sets. They enable the identification of intricate patterns, the construction of predictive models of cyberattacks, and the dynamic adaptation to threats that are in constant flux.

Intelligent methods have the capacity to process various types of cyber data, including network traffic, event logs, equipment telemetry, SIEM system data, information from open sources, and much more. In their research, scholars employ a range of advanced deep learning technologies, including convolutional and recurrent neural networks, transformers, generative adversarial networks, clustering methods, and visual analytics. This facilitates the development of next-generation comprehensive solutions for intrusion detection, intelligent monitoring, and active protection of critical systems.

A thorough analysis of contemporary scientific publications reveals several key directions in the integration of intelligent methods into CII protection systems. The following list enumerates the aforementioned topics: the utilisation of machine learning techniques for the analysis of cyber threats; intelligent systems for the analysis of logs and user behaviour; the integration of deep learning into cyber security systems; the automation of incident response; and the use of AI for threat analysis in the financial sector (Sontan, Segun, 2024; Fotiadou, Velivassaki, 2021; Gnatyuk, Sydorenko, 2024; Vegesna, 2024; Selim, 2021; Kerimkhulle, Dildebayeva, 2023; Pinto et al., 2023).

A review of contemporary research has revealed that the integration of intelligent methods into CII protection systems is an effective means of enhancing security levels. Nevertheless, there are several challenges that still need to be addressed, including the explainability of AI system decisions, the risk of false positives, and the necessity to adapt algorithms to new types of attacks. Concurrently, there is an ongoing need for further research on issues related to the improvement of self-learning algorithms, the development of hybrid approaches that combine traditional security methods with intelligent technologies, and the increasing transparency and trust in AI-driven solutions in the field of cyber security.

Thus, integrating intelligent methods into CII protection systems is a relevant area of research with significant potential for further development and practical implementation in various economic and administrative sectors. The problems listed above and their relevance informed the selection of the article's topic, as well as its goals and objectives.

The aim is to explore the application of intelligent methods and technologies for protecting CII, and the methodological approaches involved. This encompasses fundamental theoretical developments and the practical implementation of modern cyber security approaches. These approaches form the basis for optimising experimental modelling processes for detecting fraud within the CII of the financial sector using AI.

In order to achieve the stated goal, the study will identify and address the following objectives: 1) to investigate existing intelligent methods based on machine learning and AI as key technologies for effectively protecting critical information infrastructure against modern cyberattacks and security threats; 2) to conduct research using machine or deep learning methods to enable the detection of fraud in the financial sector, thereby improving the accuracy and speed with which financial abuses are identified and ensuring more reliable protection of financial institutions; 3) to provide recommendations for integrating intelligent methods into the CII protection system to ensure the security, resilience and efficiency of the financial sector in the face of contemporary cyber threats.

The objective of the research is to ascertain the means by which to ensure the cyber security of CII in the face of contemporary threats. The subject of the research is modern intellectual methods, tools and technologies for protecting CII using AII and machine learning technologies from cyberattacks.

The scientific novelty of the research lies in the further development and substantiation of theoretical principles, methodological approaches, and the provision of practical recommendations for the application of intelligent methods and technologies to protect the CII of the financial sector. In particular, the following contributions have been made: a methodology for detecting unusual transactions using machine learning methods has been proposed; models for predicting fraudulent operations have

been developed; and the processing of imbalanced data has been optimised.

2. Materials and Methods

The methodological basis of the study is a comprehensive analysis of scientific sources devoted to the use of intellectual methods and new technologies in the field of CII protection. A systematic review of the scientific literature will be conducted, encompassing both fundamental theoretical developments and practical aspects of implementing modern approaches to cyber security. A comparative analysis of concepts proposed by modern researchers will be conducted, with a view to identifying the main trends, development prospects, and potential areas for improving existing protection systems.

In their exploration of emerging cyber security trends for CII, the authors (Sontan, Segun, 2024) highlight the application of artificial intelligence and machine learning in the detection of threats, the utilisation of blockchain applications, and advancements in cloud computing security. The examination of challenges and threats, including advanced persistent threats and the impact of quantum computing, is undertaken to identify potential vulnerabilities.

As asserted by the authors (Fotiadou, Velivassaki, et al., 2021), network intrusion detection is a pivotal element in ensuring the resilience and effective operation of information systems within critical infrastructure facilities. The study posits that complex threat models have the potential to inflict substantial harm on cyber systems. New deep learning formulations have been proposed to detect threats and alerts in network logs obtained from pfSense, which is open-source software that functions as a firewall within the FreeBSD operating system. Several powerful security services are integrated into pfSense, including a firewall, URL filtering and a virtual private network. The primary objective of this study is to analyse logs obtained from a local installation of pfSense software, with a view to providing a robust and efficient solution that monitors traffic flow based on patterns automatically learned through advanced deep learning (DL) architectures. Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are utilised to construct robust multi-class classifiers capable of assigning each new instance of the network log to the appropriate category. The effectiveness of the proposed approach is evaluated by conducting several quantitative experiments and comparing it with existing ones.

The study of IT incident management methods for CII highlights that protecting CII is of paramount importance in today's digitalised world, where the growing number of cyber threats poses significant risks to national security, the economy, and public safety. Nevertheless, methodologies for the management of IT threats have yet to be extensively explored. The authors propose a novel approach to IT incident management that integrates the STRIDE model with the multi-criteria decision-making method TODIM. The methodology is designed to identify, assess, and prioritize threats based on the criticality of infrastructure objects at various levels. It is evident that the proposed methodology is capable of enhancing the security of CII by providing a systematic approach to the prioritisation and management of threats. The paper proffers a pragmatic resolution for enhancing CII protection against the encroachment of emerging cyber risks.

The issue of anomaly detection in cyber-physical systems, especially in the context of critical infrastructure protection, has been extensively researched by both domestic and foreign scholars. The author V. Vegesna posits that conventional anomaly detection methodologies founded upon rules and threshold values frequently prove ineffective for complex cyber-physical systems characterised by voluminous heterogeneous data. As an alternative, machine learning approaches are proposed. The paper undertakes an analysis of various methods, including deep neural networks for time series analysis, isolation forests for outlier detection, probabilistic models, and hybrid approaches. The following paper sets out to describe the working principles, advantages, and limitations of each method with regard to the task of anomaly detection in cyber-physical systems of critical infrastructure. To illustrate this point, the management system of power grids is examined. The present study involves the experimental comparison of various machine learning methodologies, employing authentic data sets. The findings indicate that the hybrid approach is the most efficacious (Vegesna, 2024). The author concludes that the application of machine learning methods for critical infrastructure protection and anomaly detection in cyber-physical systems shows great promise. It is evident that there are several areas in which further research is required. These include the development of specialised methods for specific systems, the improvement of data processing methods, and the assurance of the security of machine learning systems.

The article by Selim G. E. I. et al., presented in the journal Multimedia Tools and Applications (Selim et al., 2021), aims to investigate the application of machine learning methods to detect and classify anomalies in critical Industrial Internet of Things (IIoT) infrastructure. The authors highlight the increasing importance of securing IIoT systems that control critical processes. Traditional intrusion detection systems are often ineffective for the IIoT due to the large volume and heterogeneity of the data, the

complex network topology and the requirement for real-time operation. One proposed solution is the use of machine learning algorithms to analyse IIoT data, detect anomalous events and classify them by type. The study examines three main algorithms: random forest, support vector machine, and multilayer perceptron. The text provides a comprehensive overview of the mathematical underpinnings, advantages, and limitations inherent to each method. In order to evaluate the effectiveness of the proposed methodology, experiments were conducted on a dataset containing real performance indicators from an IIoT-based control system of a gas processing plant. The findings suggest that the random forest algorithm attained superior accuracy in the classification of anomalous events, attaining a maximum of 96% accuracy with a relatively brief training period. The authors of the study propose several potential avenues for further performance enhancement and underscore the significance of safeguarding critical IIoT infrastructure through the implementation of machine learning methodologies. The necessity for additional research utilising larger sets of real data and incorporating the particularities of diverse industrial sectors is underscored.

The issue of information security in the Industrial Internet of Things (IIoT) environment is also considered in the work of Kerimkhulle, Dildebayeva, et al. (2023). The authors posit that the evaluation of information security risks in IIoT is rendered complex by several factors, namely the intricacy and heterogeneity of the system, the dynamic nature of the system, the distributed network infrastructure, the absence of standards and recommendations, and the heightened consequences of security breaches. These factors mean that assessing information s ecurity risks in the IIoT requires a comprehensive approach that is adapted to the specifics and requirements of a given system and industry. It is therefore proposed that special risk assessment methods are used and that the context and specifics of the system are taken into account.

The present paper puts forward a novel method for the assessment of information security risks in IIoT systems. The approach is founded on the mathematical framework of fuzzy set theory. The article under scrutiny employs an analytical approach to evaluate the information security threats to IIoT systems, meticulously identifying the most salient criteria. The rules for decision-making are formulated as logical expressions containing input parameters. Three fuzzy inference systems are employed: one for assessing the probability of threat realization, another for estimating potential damage, and a final one for evaluating the overall information security risk for the IIoT system.

The proposed scientific approach has the potential to serve as a foundation for the development of expert decision support systems for the design of IIoT systems.

In order to evaluate the effectiveness of the proposed methodology, experiments were conducted on a dataset containing real performance indicators from an IIoT-based control system of a gas processing plant. The findings suggest that the random forest algorithm attained superior accuracy in the classification of anomalous events, attaining a maximum of 96% accuracy with a relatively brief training period. The authors of the study propose several potential avenues for further performance enhancement and underscore the significance of safeguarding critical IIoT infrastructure through the implementation of machine learning methodologies.

In their article (2023), Pinto et al. devote significant attention to reviewing the machine learning methods applied in modern intrusion detection systems (IDS), including supervised, unsupervised, hybrid and deep learning techniques. For each category, the authors provide detailed descriptions of the operating principles of the algorithms, highlighting their strengths and weaknesses in intrusion detection tasks. Furthermore, the article underlines the significance of data preprocessing and feature selection when dealing with network traffic data and security event logs. The article discusses the most promising methods and open challenges in intrusion detection systems (IDS), such as the handling of large data volumes, the addressing of concept drift, and the adaptation to the specifics of various critical infrastructures. The authors conclude that deep learning methods, in particular convolutional and recurrent neural networks, demonstrate the highest level of accuracy in detecting cyberattacks and anomalies. However, it is emphasised that the successful application of these models is contingent upon the availability of substantial sets of high-quality data.

In the study by M. Aragonés Lozano, M. Pérez Llopis, and M. Esteve Domingo (Aragonés Lozano, Pérez Llopis, Esteve Domingo, 2023), a system for detecting and analysing cyber threats targeting critical infrastructure is presented, utilising machine learning methods. This approach is rooted in the concept of "threat hunting", which involves the proactive search for indications of malicious activity. The authors emphasise that critical infrastructure is an attractive target for malicious actors, and traditional intrusion detection systems frequently prove ineffective against contemporary cyber threats. The system proposed in the article consists of modules for data collection, processing, machine learning and the visualization of results. In order to detect anomalies and classify threats, the authors employ a range of algorithms, including random forest, isolating forest and single-class support vector machine (SVM) algorithms. The system has been tested on real

data from the telecommunications infrastructure of a large operator.

The findings of the experiments demonstrate the high effectiveness of the machine learning ensemble in classifying various types of cyber threats, including the detection of new, previously unknown attacks. In the following discussion, the advantages of the proposed approach are considered by M. Aragonés Lozano et al., including its ability to generalise and adapt to evolving threats. However, the limitations of the approach are also discussed, including the need for continuous system updates as new types of attacks emerge. The study emphasises the promising potential of machine learning methods for effectively countering cyber threats to critical infrastructure.

The article by K. Raval et al. (2023) presents a comprehensive examination of cyber security issues concerning critical infrastructures in the context of modern threats, including those related to the use of artificial intelligence technologies. The authors analyse the most common cyberattacks that pose a threat to critical infrastructure systems, including denial-of-service attacks, malware, attacks on industrial control systems, and cyber espionage. A particular focus is allocated to the identification of emergent threats that are concomitant with the progression of AI. The article examines approaches to ensuring the cyber security of critical infrastructures using artificial intelligence (AI) and machine learning methods. The analysis encompasses a range of subjects, including intrusion and anomaly detection, vulnerability identification, incident response, protection of industrial control systems, and the safeguarding of AI itself against attacks. These subjects underscore the necessity of studying AI-related threats. A plethora of examples from the extant literature on the use of AI for security purposes across various sectors underscores the relevance of this research topic. The article presents unresolved problems and future research directions, including ensuring the interpretability and understandability of AI solutions, developing machine learning methods in adversarial learning environments, developing standards and risk assessment methodologies, and addressing privacy and ethics issues in the use of AI.

In their study, Alqudhaibi et al. (2023) propose a slightly different proactive approach to predicting cyber threats with a view to protecting critical infrastructure in the context of Industry 4.0. The fundamental objective is to analyse the motivations and behavioural patterns of potential attackers by employing machine learning methodologies. The authors of the research confirm that the ability to predict future cyberattacks can be enhanced by a comprehensive understanding of the motivations and characteristics of the individuals responsible for perpetrating them. The developed system consists of modules for the collection and enrichment of data on cyber incidents, the analysis of attacker motivations, the prediction of threats, and the assessment of risks.

Machine learning algorithms, including random forest and neural networks, are employed for prediction, with training being conducted on historical data concerning attacks and motivations. The experimental evaluation yielded results that demonstrated the proposed approach to be more accurate in predicting threats than traditional methods, with an accuracy of approximately 85–90% for various types of attacks. In conclusion, the authors discuss the study's limitations, such as the need for continuous data updates and potential inaccuracies in determining the actual motivations of attackers. The importance of adhering to ethical standards in data collection and usage is also emphasised. The present study (Balatska, Opirskyy, 2024) raises the issue of protecting government centralized systems used for data storage and processing. It is particularly important to note the vulnerability to cyberattacks, the insufficient protection of personal data, the lack of user control over their information, and the non-compliance with international standards such as the General Data Protection Regulation (GDPR). This creates the need for the adoption of new technological solutions capable of addressing these shortcomings and ensuring increased reliability of state registries. In the article (Balatska, Opirskyy, 2024), the implementation of blockchain technologies in government registration systems is proposed as a solution to enhance data security, transparency, and reliability. The authors further posit that this implementation would also give users greater control over their personal information. Blockchain technology enables the creation of decentralised systems in which data cannot be altered without proper authentication and all transactions are recorded, thereby reducing the risk of unauthorised access and fraud. A key element of the study is developing a mathematical model to quantitatively assess the effectiveness of implementing blockchain in government registration systems. The model is capable of evaluating indicators such as security, transparency, reliability, and data processing speed, thus facilitating comparisons with traditional centralized systems. The analysis demonstrates that blockchain has the potential to significantly reduce corruption risks, ensure full transparency of governmental processes, and increase public trust in state institutions. Furthermore, the implementation of blockchain technology aligns with international data security and protection standards, such as the GDPR, establishing it as a promising tool for public administration. The study also identifies key areas for blockchain development, such as enhancing system scalability, optimising data processing costs and integrating with other technologies for public governance. The data obtained demonstrate the significant potential of blockchain solutions for

transforming public registries, and the developed model is proposed for use as a tool for further research and implementation of the technology in public institutions.

The present study (Korniyenko, Galata, Ladieva, 2019) investigates the main issues of information security of critical information resources and the reasons for their emergence. The process of constructing a mathematical model to counter threats in the protection system of critical information resources is carried out using a Markov chain. The research calculates the probability of the information system's mathematical model being in one of four states: no threat has occurred; a threat has occurred but was not realised; the threat was realised; the threat occurred but was countered by the protection system. The utilisation of numerical analysis in the examination of the proposed methods has been demonstrated to facilitate the identification of threats pertinent to the system under scrutiny, thus substantiating their practical application. A notable disadvantage of the proposed method is the necessity to consider the behaviour of the system with regard to each type of threat in isolation, and the inability to ascertain the behaviour of the simultaneous action of multiple threats. The authors also note that studying the impact of each threat separately enables more detailed analysis of each type and determination of its probability of occurrence. Given the operation of systems for protecting critical information resources and the significant changes in the nature of modern threats, it is necessary to design and implement information security systems that take into account trends in cyber threat development.

The monograph by Yevseiev, Hryshchuk, Molodetska and Nazarkevych (2022) explores the methodology of co-operative modelling of conflict interactions among security system agents. It demonstrates the concept of modelling the structure and functioning of security systems for critical infrastructure facilities. The following methodology is proposed for the assessment of forecasts of social impact in regional communities. A strategy for counteracting strategic manipulation of public opinion during decision-making by social network actors is developed, based on a conceptual model of guided self-organisation of social networks. Algorithms for sparsifying the critical infrastructure identification system, along with their software implementation, have been developed.

In their scientific work (Shevchenko, 2023), the authors argue that individual decision-making is an element of managing any societal process; therefore, theories of cognitive science are applied across various fields, including information and cyber security systems. The study proposes the development of a cognitive "danger-risk" model for managing information risks in information and cyber security systems. The present study examines the main stages of developing the "danger-risk" cognitive model in the field of information and cyber security. These stages comprise the identification of complex situations and problems, the construction of a cognitive map, the modelling and verification of the model, and the analysis of adequacy and dynamic situations. A "danger-risk" model, which is not based on any theory, has been developed and its elements have been defined.

The article by Zhang et al. (2024) is dedicated to the problem of link prediction in graphs that combine information security requirements with data on cyber security threats. The authors propose a novel link prediction method based on edge propagation. The key components of the proposed approach are identified, including the encoding of graph nodes and edges into vector representations, iterative information propagation between node vector representations through intermediate edges, the formation of vector representations for node pairs that account for their shared connections at various distances, and the training of a classifier (logistic regression) based on these vector representations to predict the presence of a link.

Experiments on both real and synthetic datasets have demonstrated that the proposed method outperforms alternative approaches to link prediction in graphs. The utilisation of vector representations for the encoding of semantics, the consideration of link structures at varying levels, the capacity to operate with incomplete data and heterogeneous graphs, and the interpretability through the employment of explicit node pairs proposed by the researchers substantiate the primary advantages of the method. In conclusion, potential future research directions are discussed, including the extension of the method to account for dynamics and its integration with decision-making systems.

The utilisation of graph theory for the visual analysis and modelling of cyberattacks is also presented in the article by M. Rabzelj, C. Bohak, L. Južnič, A. Kos, and U. Sedlar (2023). The authors propose an approach based on the creation of a graph model in which nodes represent entities (vulnerabilities, attack vectors, attacker actions), and edges reflect possible transitions between them during the course of an attack. The article delineates the pivotal components of the approach, which include a cyberattack graph model comprising diverse node types and weighted edges; algorithms for automatic graph construction based on input data; graph analysis methods such as shortest path search, identification of critical nodes, and clustering; visualisation of the cyberattack graph using various representations; and an interface for cyber security analysts.

Experiments with real-world data have demonstrated the application of the approach for risk assessment, planning of preventive measures, and analysis of the consequences of cyberattacks. The authors of the study identified several key advantages of the graph-based model, including its ability to provide clear visualisation of complex processes and interconnections, facilitate the identification of critical components and vulnerabilities, support "what-if" analysis and scenario modelling, and integrate diverse threat data from various sources. These advantages suggest that further research in this area is promising, including integration with intrusion detection systems and the application of machine learning methods.

The article by Balatska, Poberezhnyk, Petriv and Opirskyy (2024) explores the potential of using blockchain technology in the context of single sign-on (SSO) systems. SSO systems enable users to authenticate once and access various digital resources. The problem statement addresses contemporary trends in security and access management, including the risk of data compromise and the inefficiencies inherent in information exchange between SSO systems. The primary objective of the present study is to develop and implement technological solutions with the aim of enhancing the security, resilience, and efficiency of SSO systems in the digital environment. Moreover, seminal studies are examined, emphasising the significance of blockchain adoption and innovation in user identification and authentication.

In their seminal work, Berardi et al. (2023) set out a novel approach to the critical issue of network security in the context of Time Sensitive Networking (TSN), with a particular focus on the security implications of the Precision Time Protocol (PTP). The researchers analyse the main security threats to PTP, including leader spoofing, man-in-the-middle attacks, distributed attacks, and denial-of-service attacks. A taxonomy of these threats across different architectural levels is proposed. The development of a secure PTP implementation prototype involves the utilisation of cryptographic protocols and certificate-based authentication mechanisms. The experimental evaluation demonstrated that the level of security overhead was acceptable. D. Berardi et al. place significant emphasis on the critical importance of securing PTP and other TSN components in order to protect industrial control systems from malicious interference. Future research directions include the integration of the proposed security mechanisms into existing systems and the development of methods to counter distributed time synchronization attacks.

In their 2023 study, authors T. Kim and W. Pak explored the application of transformer-based deep learning methods for Network Intrusion Detection (NID). The article proposes a novel approach in which incoming network data is transformed into images and processed by a set of parallel transformer networks. Each transformer is specialised in detecting a specific type of attack or anomaly in the network traffic. The outputs of the parallel transformers are then combined using fully connected layers prior to the final classification being obtained. Experiments conducted on the NSL-KDD and CSE-CIC-IDS2018 datasets demonstrate that the proposed method outperforms several existing deep learning-based approaches to NID. The advantages of this approach include high detection performance for various types of attacks, scalability, and interpretability of model decisions. The authors also note limitations of the approach, such as high computational requirements and the need for prior data filtering.

The conducted analysis confirms the necessity of addressing challenges related to the development of modern intelligent approaches for the protection of CII. A comprehensive review of the scientific literature (Sontan, Segun, 2024; Fotiadou, Velivassaki, 2021; Gnatyuk, Sydorenko, 2024; Vegesna, 2024; Selim, 2021; Kerimkhulle, Dildebayeva, 2023; Pinto et al., 2023) has been conducted, and the following conclusions can be drawn:

The utilisation of machine learning algorithms and AI methods represents a pivotal approach in the identification of anomalous activities, the detection of cyberattacks, and the assurance of the protection of CII.

A number of methodologies have been proposed for the transformation of heterogeneous data (network traffic, event logs, device information) into formats that are suitable for analysis using computer vision approaches and image processing techniques.

Deep learning methods have been extensively applied, encompassing convolutional neural networks, recurrent networks, autoencoders, and transformers. The efficacy of these technologies in the domains of anomaly detection and attack classification has been demonstrated. In order to enhance the performance and accuracy of intrusion detection systems, ensemble models are employed, which combine several machine learning algorithms targeted at different threat types and security aspects.

Architectural solutions and mechanisms are being developed to provide asymptotic security management of critical infrastructure, offering adaptability and self-configuration capabilities in protection systems under dynamic threat conditions.

Specialised security solutions are being designed for online financial services. These solutions include fraud prevention mechanisms based on scoring models, anomaly detection and biometric technologies.

The security aspects of the Industrial Internet of Things and critical infrastructure components are analysed within the Industry 4.0 framework.

Comprehensive security risk assessments of critical facilities are being carried out using approaches based

on fuzzy logic, cognitive modelling and hierarchical hypernetworks. These approaches allow for the uncertainty and interdependence of various risk factors to be considered.

Methods of protecting network time synchronisation protocols and other critical infrastructure components requiring high-precision timing are being explored using cryptographic mechanisms.

Significant attention is paid to interpreting and visualising machine learning model outputs for critical infrastructure protection. This includes visual analysis of cyberattack graphs.

Proactive cyber security strategies involving the analysis of adversary motivations and the prediction of potential threats using AI methods are being investigated.

It is evident from the analysis of the extant literature that intelligent methods based on machine learning and AI are pivotal technologies for ensuring effective protection of CII against modern cyberattacks and security threats. However, despite the significant achievements in applying intelligent machine learning and AI methods to protect CII, there remain a number of issues and unresolved challenges that require further research and development. The following main ones can be highlighted:

1. Ensuring the interpretability and transparency of AI models used to make critical cyber security decisions. Methods for visualising, analysing neural network activations and extracting interpretable rules from deep learning "black boxes" must be advanced.

2. Development of methods for assessing the confidence and uncertainty of intelligent threat detection systems to improve reliability and enable informed response decisions.

3. Development of active and online learning methods to ensure timely adaptation of security systems to changing conditions and new types of threats without the need for complete retraining.

4. Improvement of methods for automatic feature and representation extraction based on unsupervised learning in cyber security tasks.

5. Investigation of the resistance of machine learning algorithms to attacks aimed at introducing errors into the decision-making process (adversarial attacks). Methods to ensure resistance to such attacks are needed.

6. Improvement of approaches to sensitive-private training to combine heterogeneous confidential data from different subjects without disclosing private information.

7. Investigation of methods for integrating expert knowledge into machine learning models by hybridising with logical and productive rules to increase interpretability and efficiency.

8. Development of scalable platforms for testing, verification and comparison of different methods of intelligent protection using realistic data sets and attack scenarios.

9. Development of methods for planning coordinated actions of systems for active protection of critical infrastructure facilities based on situational analysis and predictive models of attackers.

10. Research on the integration of specialised secure and reliable AI methods with traditional information security technologies to build integrated cyber security systems.

In view of the aforementioned points, it is imperative to emphasise the practical implementation of these methodologies in real-world settings. A pivotal implementation domain pertains to the identification of fraudulent activities within the financial sector, which poses a substantial threat to critical information systems. The authors conducted a study that utilised machine learning or deep learning methods to detect financial fraud. It is noteworthy that the Prosecutor General's Office of Ukraine (PGO) has reported that more than 38,000 cases were initiated under Article 190 "Fraud" during the initial five months of 2024. This figure is 1.6 times higher than the total number of cases that were recorded during the entire year of 2021, when 23,847 cases were reported. It is noteworthy that nearly one in five fraud-related criminal proceedings were referred to the courts. Fraud-related criminal cases increased by 10% compared to the same period last year. Of all registered cases in 2024, more than 7,000 (18%) proceeded to court. Additionally, over 4,500 cases were closed (The number of fraud cases this year is already 1.6 times higher than for the entire year of 2021, 2024).

According to the 2023 Cost of Data Breaches report by IBM Security (Storchak, 2024), phishing and compromised credentials represent the two most prevalent initial attack vectors in data breaches. As illustrated in Figure 1, the ratio of open fraud cases filed in court is shown as of June 1, 2024.

In the process of practical application of the aforementioned methods in real-world scenarios, the dataset used was the Credit Card Fraud Detection dataset (https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud) (Credit Card Fraud Detection). The project, titled "Credit Card Fraud Detection", aims to identify instances of credit card fraud. The dataset is available on the Kaggle platform for use in machine learning tasks and competitions. It was created to analyse and build models for detecting credit card fraud. The dataset contains information on both legitimate and fraudulent transactions. Consisting of 284,807 credit card transactions, it includes just 0.172% (492) fraudulent transactions. The data includes numerical features obtained through principal component analysis (PCA).

The data was taken from systems that process credit card transactions. The dataset includes 28 anonymised numerical features, which were generated using the principal component analysis (PCA) algorithm. There

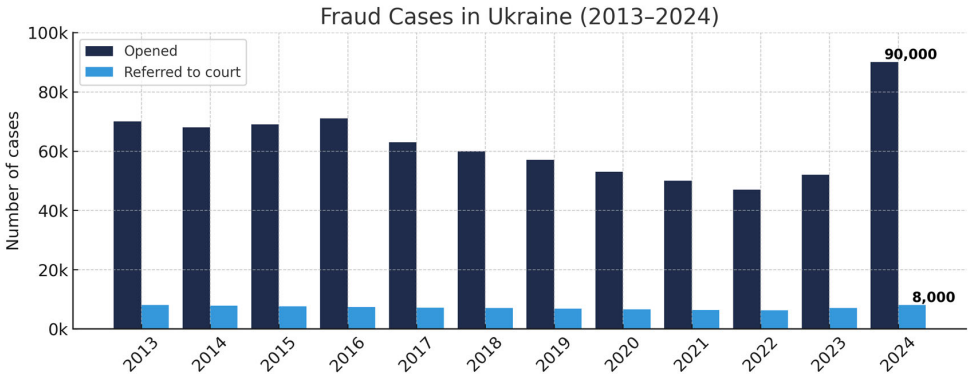

Figure 1. Number of fraud cases, June 1, 2024

are also two additional columns. "Time" indicates the time of the transaction in relative units, while "Amount" indicates the transaction amount. The "Class" column indicates the type of transaction: 0 – legitimate; 1 – fraudulent. It is noteworthy that merely 0.17% of transactions are determined to be fraudulent, thereby illustrating an archetypal imbalanced classification problem. The objective of the analysis on the Kaggle platform is threefold: firstly, to detect anomalous transactions using machine learning methods; secondly, to build models for predicting fraudulent activities (classifiers); and thirdly, to optimise performance on imbalanced data. The dataset is well-suited for the development of fraud detection algorithms, including, but not limited to, logistic regression, gradient boosting (e.g., XGBoost, LightGBM), neural networks, and ensemble models. Furthermore, it offers a valuable opportunity to explore techniques for handling class imbalance, including oversampling (SMOTE), undersampling, and class weighting. Additionally, it facilitates the evaluation of precision, recall, and other metrics that are critical to the modelling of such systems.

Conducting such research is of critical importance for the financial sector, as it helps to reduce losses from fraud, increase customer trust in banks and payment systems, and improve the speed and accuracy of transaction processing. In the event of this binary classification task being addressed directly by means of the application of any machine learning algorithm without the necessity of prior data analysis and with Accuracy serving as the evaluation metric, calculated by formula (1), a result of approximately 99% may be obtained – even in the absence of correct predictions by the algorithm for any instances of class "1".

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where TP is a true positive result, the number of correctly predicted instances of the target class;

TN is a true negative value, the number of correctly predicted instances of a non-target (zero) class;

FP is false positives, the number of falsely detected instances belonging to the target class;

FN is a false negative result, the number of falsely predicted instances belonging to a non-target class.

Data analysis includes the selection of class balancing methods (or justification for not using them), a reasonable choice of evaluation metrics, and analysis of machine and deep learning methods. The study tested more than 20 machine and deep learning models in six different scenarios:

1. Without emission pre-processing and without class balancing.
2. With emission pre-processing, but without class balancing.
3. Without emission preprocessing and with class balancing using class weights.
4. With emission preprocessing and class balancing using class weights.
5. Without emission preprocessing and with class balancing using SMOTE.
6. With emissions pre-processing and class balancing using SMOTE.

The primary evaluation metric for the models is ROC-AUC, as it most accurately reflects information about the true positive rate (TPR) and the false positive rate (FPR), depending on the decision threshold. When converting continuous output to a binary indicator, a threshold must be set at which the value 0 switches to 1; although 0.5 is commonly used, it may not be optimal in cases of class imbalance. To provide a general assessment of models independent of a specific threshold, AUC-ROC is applied, which represents the area under the curve reflecting the relationship between TPR (2) and FPR (3).

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{FP + TN} \quad (3)$$

TPR reflects recall, while FPR indicates the proportion of negative class instances incorrectly classified by the algorithm. In the case of

a perfect classifier, the AUC-ROC equals 1 (FPR = 0, TPR = 1). If the classifier produces random predictions, the AUC-ROC approaches 0.5, when TP and FP are equal. Each point on the ROC curve corresponds to a specific threshold, and the area under the curve serves as an indicator of the algorithm's quality: the larger the area, the higher the quality. The gradient of the curve is also a salient factor, as the objective is to optimise TPR and minimise FPR, with the curve approaching the point (0,1). As demonstrated in Figure 2, the current model exhibits deficiencies in comparison to the ideal, which is represented by an isosceles triangle oriented along the diagonal. Furthermore, it has been observed that errors in class 0 can contribute to the enhancement of the detection of class 1. The optimal balance between TPR and FPR is to be determined on an individual basis for each specific case.

The act of balancing between TPR and FPR is a complex task. It is hypothesised that lowering the threshold for identifying the target class 1 may result in the blocking of a significant number of borderline transactions, which in turn may lead to a reduction in customer loyalty. Conversely, minimising blocks may increase the risk of fraud and result in customer dissatisfaction. Consequently, the determination of the decision threshold should be informed by the unique requirements of the banking institution in question. All models will be trained according to a unified principle: transformation methods will be adapted on the training set and only applied to the validation set. Each model will be constructed as a pipeline of transformations, with a balanced dataset split. To illustrate this point, if the original dataset contains 0.17% of the target class, it is essential that the same ratio is maintained for both the training and validation sets. Initial transformations include scaling using StandardScaler and RobustScaler.

For each model, a learning curve must be constructed in order to demonstrate how performance varies with the size of the training dataset. This should include a confidence interval for the ROC-AUC metric. The learning curve shows how training set size affects model performance and cross-validation. This allows the following two key questions to be answered: 1. How does the model's performance change as the amount of training data increases? 2. How sensitive is the model to errors due to variance compared to bias? Furthermore, for each model, a grid search will be applied to optimise hyperparameters, and training will only be conducted after selecting the best hyperparameter values. The findings of the examined models are outlined in Table 1.

The KNeighbors Classifier model does not support class weighting; however, the most optimal results were demonstrated using the SMOTE method, which proved to be almost unaffected by the presence of outliers, although the optimal result was obtained with outliers included. In a similar manner, the BaggingClassifier, AdaboostClassifier, and GradientBoosting models are unable to utilise class-weighted datasets. The ROC-AUC metric for BaggingClassifier demonstrated superior outcomes on the dataset without undergoing any transformations. Conversely, the ROC-AUC metrics for AdaboostClassifier and GradientBoosting exhibited elevated values when SMOTE was implemented. The selection of the most appropriate model was based on the ROC-AUC metric. Initially, the optimal model was selected from a range of

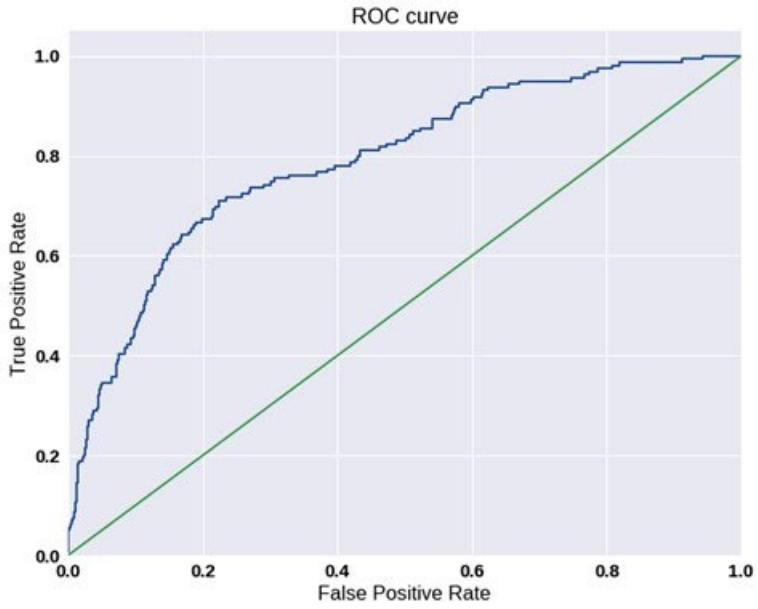

Figure 2. ROC-AUC

Table 1

ROC-AUC metric for models with different datasets

Model	Outlier condition	No transformations	With SMOTE	With class weighting
LogisticRegression	with outliers	0.972423	0.972235	0.973915
LogisticRegression	without outliers	0.976971	0.975332	0.973418
SVC	with outliers	0.964920	0.979727	0.977112
SVC	without outliers	0.962544	0.976860	0.969799
KNeighborsClassifier	with outliers	0.953505	0.957686	-
KNeighborsClassifier	without outliers	0.951753	0.957471	-
DecisionTreeClassifier	with outliers	0.918165	0.907020	0.887544
DecisionTreeClassifier	without outliers	0.925287	0.924379	0.919717
StackingClassifier	with outliers	0.943349	0.952068	0.972654
StackingClassifier	without outliers	0.953258	0.949473	0.977153
BaggingClassifier	with outliers	0.972654	0.972139	-
BaggingClassifier	without outliers	0.977153	0.975305	-
AdaboostClassifier	with outliers	0.918156	0.952031	-
AdaboostClassifier	without outliers	0.925261	0.954966	-
GradientBoosting	with outliers	0.927943	0.972651	-
GradientBoosting	without outliers	0.914239	0.981783	-
NeuralNetwork	with outliers	0.977698	0.972294	0.972206
NeuralNetwork	without outliers	0.980729	0.974456	0.976503

machine learning methods, followed by the selection of the best-performing neural network approach. As a result, five machine learning models with the highest performance were selected:

– (37) GradientBoostingClassifier trained on a SMOTE dataset cleaned of outliers;

– (8) SVC trained on a SMOTE dataset without outlier removal;

– (27) BaggingClassifier trained on the original dataset with outliers removed;

– (10) SVC trained on the original class-weighted dataset without outlier removal;

– (1) LogisticRegression trained on the original dataset with outliers removed. Figure 3 presents the performance chart for the selected models.

Further analysis reveals two key questions, depending on task priorities: how to ensure the detection of all fraudulent operations, and how to maximise the detection of fraudulent transactions while minimising the identification of legitimate operations as fraudulent. Based on this, two top-performing models can be identified:

– In the first case, the leader is GradientBoosting Classifier 37;

– In the second case, the leaders are BaggingClassifier 27 or SVC 8.

A significant number of false positives – transactions erroneously identified as fraudulent – have the potential to cause more harm than good. Consequently, BaggingClassifier 27 is identified as the optimal

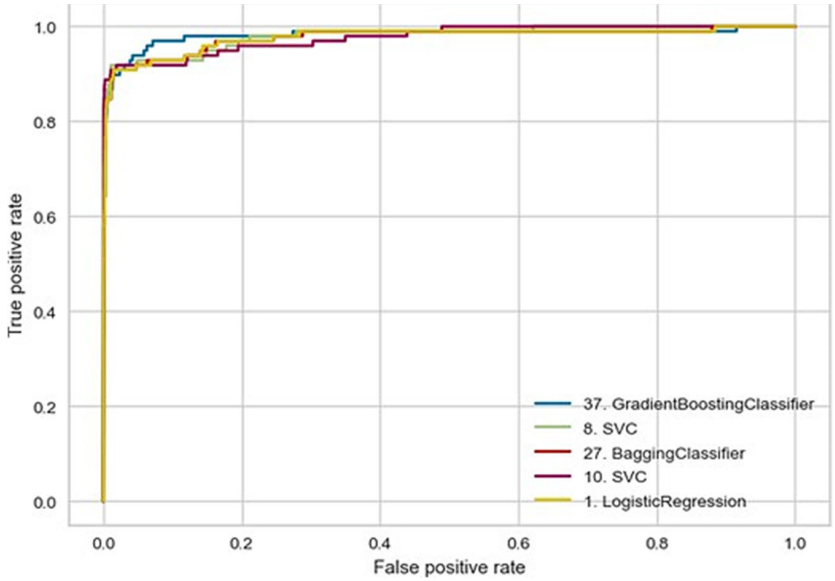

Figure 3. ROC curve

model. The optimal data processing options for the neural network were as follows: the original dataset with outliers removed; the original dataset with outliers retained; and class weighting applied to the dataset with outliers removed. The ROC curve for the neural network is presented in Figure 4. The Model 39 demonstrated superior performance in comparison to the other models.

Figure 5. ROC curve for the neural network and BaggingClassifier

It is evident that the neural network model demonstrates the optimal performance. Consequently, the primary model selected is Neural Network 39, which was trained on the original dataset with outliers removed.

3. Discussion

It is evident that the experiment conducted corroborates the issues highlighted by the authors, as previously delineated. In order to facilitate the decision-making process in the domain of cyber security, it is imperative to ensure the transparency and interpretability of AI models. The evaluation of the trustworthiness and reliability of intelligent system outputs is a pivotal aspect in the development of protection algorithms for CII. It is imperative to investigate methodologies for integrating expert knowledge into machine learning models through hybridisation with logical and production rules to enhance model interpretability and efficiency.

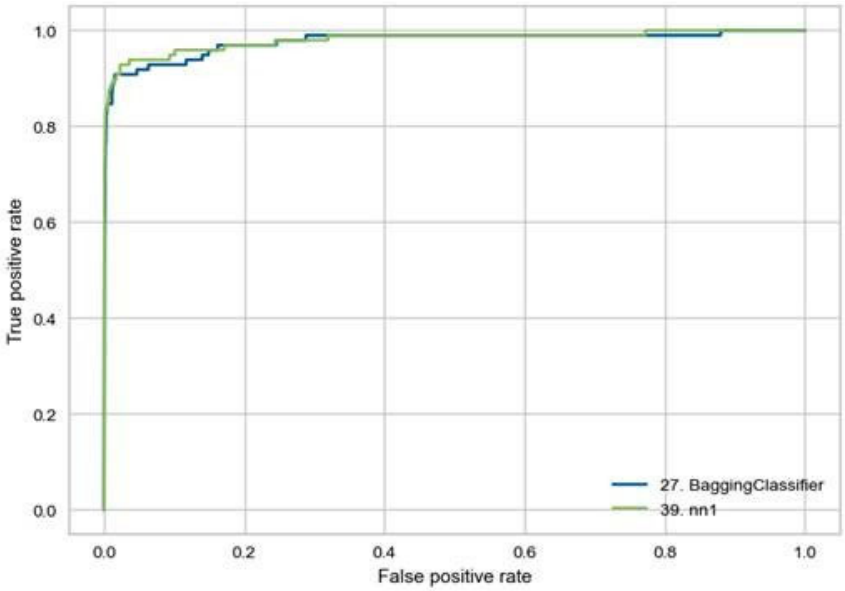

Figure 4. ROC curve for nn

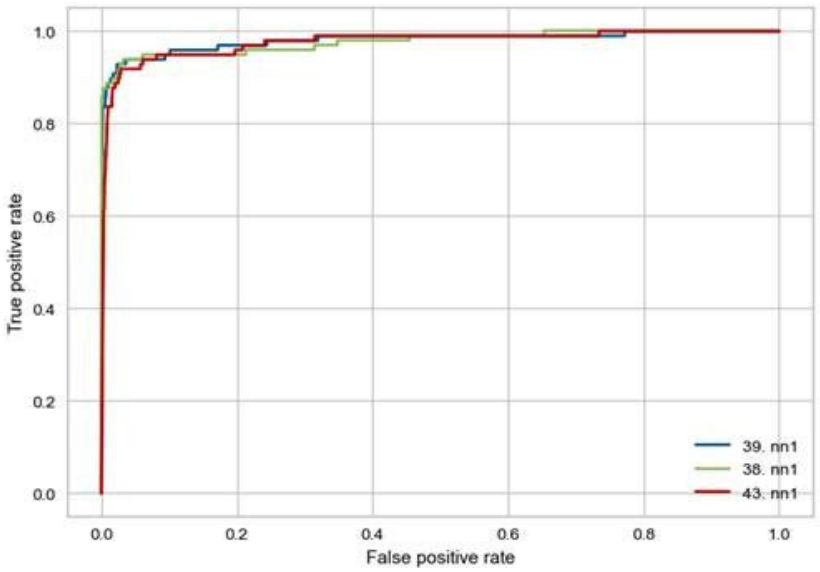

Figure 5. presents the final comparison of the best machine learning and deep learning models

The creation of scalable platforms for the testing, verification and comparison of various intelligent security methods using realistic datasets and attack scenarios remains a highly relevant task. The integration of specialised secure and trustworthy AI methods with traditional information security technologies is of great importance for the construction of comprehensive cyber security systems.

4. Conclusions

The conducted research confirms the need to apply intelligent methods to ensure the protection of CII in the financial sector. The main directions of solving this issue are as follows:

– Enhancement of the security of financial systems through the development of methods and technologies aimed at countering cyber threats and minimising risks. This involves the implementation of both technical solutions (such as intrusion detection and prevention systems, data encryption methods) and organisational measures (including information security policies and cyber incident response plans).

– Creation of reliable and threat-resistant systems capable of operating effectively under high load and in dynamic cyberspace conditions. This requires the development of intelligent security systems based on fault-tolerant architectures, self-healing algorithms and adaptive response mechanisms.

– Ensure the interpretability of AI systems used to protect financial systems. It is important to develop algorithms that are understandable to experts and users, including the use of models with transparent operating principles, as well as mechanisms for auditing and explaining decisions.

– Building a high level of trust in AI technologies among users and regulators. As the financial sector serves critical needs of citizens, businesses, and the state, the introduction of AI technologies should be accompanied by confidence-building measures such as quality control, certification, and minimising the risks of possible malfunctions or erroneous decisions that could lead to significant financial and reputational losses.

In view of this, the integration of intelligent methods into CII security systems is a key area of ensuring the security, stability and efficiency of the financial sector in the face of modern cyber threats. To increase confidence in artificial intelligence systems in financial sector CII, it is advisable to implement the following measures:

– Introduction of mechanisms for auditing and monitoring the operation of artificial intelligence systems by independent experts to verify the correctness of their work.

– Securing the transparency and comprehensibility of the AI system. This involves disclosing to users and regulators the decision-making logic of AI, the algorithms used, and the data sources.

– Development and implementation of ethical standards and principles for the use of AI in critical financial sector information infrastructure systems, which will help to strengthen confidence that these systems are used responsibly and in the best interests of users.

– Active involvement of users and regulators in the development and implementation of artificial intelligence technologies.

The following will be aimed at improving the efficiency of CII in the financial sector by increasing confidence in artificial intelligence systems:

– Extension of the scope of artificial intelligence technologies and their penetration into critical industries.

– Improvement of the quality and reliability of financial services while reducing the risks associated with them.

– Improving the competitiveness of financial institutions that implement artificial intelligence technologies.

– Stimulation of investments in the development of artificial intelligence technologies for the financial sector.

Among the drawbacks of using AI systems in the CII of the financial sector, the following should be noted:

– Additional costs for financial institutions to ensure transparency and oversight of AI systems.

– Potential problems in achieving full transparency and interpretability of complex AI systems.

– The risk of user distrust of innovations in the financial sector due to its conservatism.

Therefore, increasing confidence in the use of artificial intelligence systems in financial sector CII through the implementation of comprehensive measures is an important and promising area that can bring significant benefits, but requires careful consideration and implementation.

At the same time, the relevance of the topic is confirmed by the identification of future research areas, in particular:

– Conducting a comprehensive analysis of threats and risks specific to financial sector CII in order to identify priority areas for the development of protection methods.

– Exploring advanced artificial intelligence technologies (machine learning, deep learning, neural networks) and their application to protect financial systems.

– Development of methodologies for assessing the reliability and fault tolerance of intelligent CII protection systems.

– Researching approaches to ensuring transparency and interpretability of artificial intelligence systems in the field of cyber security of financial institutions.

Thus, despite significant progress, the protection of critical financial sector infrastructure using intelligent methods remains an active interdisciplinary research area. It requires solving fundamental problems related to the security, reliability, interpretability, and trust in artificial intelligence systems, which underscores the relevance of this fundamental research.

References:

1. Sontan, A., & Segun, S. (2024). Merging trends in cyber security for criticalinfrastructure protection: a comprehensivereview. *Computer Science & IT Research Journal*, Volume 5, Issue 3.

Fotiadou, K., Velivassaki, T., Voulkidis, A., Skias, D., Tsekeridou, S., & Zahariadis, T. (2021). Network Traffic Anomaly Detection via Deep Learning. *Information*, 12(5). 215 p.

Gnatyuk, S., Sydorenko, V., Polozhentsev, A., & Sokolov, V. (2024). Method for managing IT incidents in critical information infrastructure facilities. *CPITS-II 2024: Workshop on Cyber security Providing in Information and Telecommunication Systems II*. P. 323–336.

Vegesna, V. (2024). Machine Learning Approaches for Anomaly Detection in Cyber-Physical Systems: A Case Study in Critical Infrastructure Protection. *International Journal of Machine Learning and Artificial Intelligence*. Vol. 5. № 5. P. 1–13.

Selim, G., et al. (2021). Anomaly events classification and detection system in critical industrial internet of things infrastructure using machine learning algorithms. *Multimedia Tools and Applications*. Vol. 80. № 8. P. 12619–12640.

Kerimkhulle, S., Dildebayeva, Z., Tokhmetov, A., Amirova, A., Tussupov, J., Makhazhanova, U., & Salykbayeva, A. (2023). Fuzzy Logic and Its Application in the Assessment of Information Security Risk of Industrial Internet of Things. *Symmetry*, 15 (10). P. 1958.

Pinto, A. et al. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*. Vol. 23. № 5. P. 2415.

Aragonés Lozano, M., Pérez Llopis, I., & Esteve Domingo, M. (2023). Threat hunting system for protecting critical infrastructures using a machine learning approach. *Mathematics*. Vol. 11. № 16. P. 3448.

Raval, K. J., et al. (2023). A survey on safeguarding critical infrastructures: Attacks, AI security, and future directions. *International Journal of Critical Infrastructure Protection.* P. 147.

Alqudhaibi, A., et al. (2023). Predicting cyber security threats in critical infrastructure for industry 4.0: a proactive approach based on attacker motivations. *Sensors*. Vol. 23. № 9. P. 4539.

Balatska, V., & Opirskyy, I. (2024). Blockchain as a tool for transparency and protection of government registries. *Ukrainian Scientific Journal of Information Security*. Vol. 30, issue 2. P. 221–230.

Korniyenko, B. Y., Galata, L., & Ladieva, L. (2019). Mathematical Model of Threats Resistance in the Critical Information Resources Protection System. In: International Conference on Intelligent Tutoring Systems [online]. Kingston: ITS. Available at: https://ceur-ws.org/Vol-2577/paper23.pdf

Yevseiev, S., Hryshchuk, R., Molodetska, M., & Nazarkevych, M. (2022). Modeling of security systems for critical infrastructure facilities: monograph. Kharkiv: PC Technology Center. 196 p.

Shevchenko S., et al. (2023). Information Security Risk Management using Cognitive Modeling. *Cyber security Providing in Information and Telecommunication Systems*. Vol. 3550. P. 297–305.

Zhang, Y. et al. (2024). Edge propagation for link prediction in requirement-cyber threat intelligence knowledge graph. *Information Sciences*. Vol. 653. Pp. 119770.

Rabzelj, M., Bohak, C., Južnič, L., Kos, A., & Sedlar, U. (2023). Cyberattack Graph Modeling for Visual Analytics. *IEEE Access*. Vol. 11. Pp. 86910–86944.

Balatska, V., Poberezhnyk, V., Petriv P., & Opirskyy, I. (2024). Blockchain Application Concept in SSO Technology Context, CPITS-2024: Cyber security Providing in Information and Telecommunication Systems. P. 38–49.

Berardi, D., et al. (2023). Time sensitive networking security: issues of precision time protocol and its implementation. *Cyber security*. Vol. 6. № 1. P. 8.

Kim, T., & Pak, W. (2023). Deep Learning-Based Network Intrusion Detection Using Multiple Image Transformers. *Applied Sciences*. Vol. 13(5). P. 2754. DOI: https://doi.org/10.3390/app1305275

The number of fraud cases this year is already 1.6 times higher than for the entire year of 2021 (2024). Available at: https://opendatabot.ua/analytics/fraud-2024-5

Storchak, Y. (2024). Insider Threat Statistics for 2024: Reports, Facts, Actors, and Costs [online]. Available at: https://www.ekransystem.com/en/blog/insider-threat-statistics-facts-and-figures

Credit Card Fraud Detection. Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Received on: 05th of April, 2025
Accepted on: 18th of May, 2025
Published on: 10th of June, 2025