

Intelligence versus threats: The role of AI agents in ensuring cybersecurity of information and telecommunication systems^{*}

Nikolay Ievlev^{1,†}, Vitalii Klymenko^{1,†}, Anatolii Morozov^{1,†}, Vitalii Yashchenko^{1,†},
and Hennadii Hulak^{1,2,*,†}

¹ *Institute of Mathematical Machines and Systems Problems of the National Academy of Sciences of Ukraine, 42 Ac. Glushkov ave., 03187 Kyiv, Ukraine*

² *Borys Grinchenko Kyiv Metropolitan University, 18/2 Bulvarno-Kudriavska str, 04053 Kyiv, Ukraine*

Abstract

The article is devoted to the use of AI-based intelligent agents for cybersecurity of information and telecommunication systems (ITS). The prospects for their use in situational centers, SOC architectures, SOAR systems and crisis conditions are considered. The importance of regulatory framework, transparency of AI and human-machine interaction for the resilience of ITS to modern and future threats is noted. In conclusion, recommendations are given for the implementation and effective use of intelligent agents in cybersecurity, including the need for continuous personnel training, ensuring transparency and controllability of AI systems, as well as the development of multi-level protection mechanisms focused on the adaptability and resilience of ITS in the face of modern and future threats. Particular attention is paid to neural-like multidimensional multi-connected receptor-effector growing networks capable of self-learning, contextual response and adaptation to new threats.

Keywords

adaptive security, security profiles

1. Introduction

Information and telecommunication systems (ITS) have become an integral part of the functioning of modern society. Their development ensures the stable operation of the digital economy, e-commerce, government infrastructure, healthcare, transport, energy and defense systems. Against the backdrop of global digitalization, the role of ITS is constantly growing, and along with it, the requirements for ensuring their security are increasing.

With the constant expansion of the functionality of ITS and the introduction of new technological solutions, their vulnerability to cyber threats is also increasing. Modern cyberattacks are becoming increasingly sophisticated, targeted and large-scale. They are capable of disrupting the operation of key information systems, causing economic and reputational damage, and threatening national security. This problem is especially acute in wartime, when cyberspace becomes a battlefield in hybrid conflicts. Here, cyberattacks acquire strategic significance, aimed at destabilizing communications, supply chains, command centers and control infrastructures. Against this backdrop, artificial intelligence (AI) acts as a dual factor: on the one hand, it provides powerful tools for strengthening cybersecurity, but on the other, it itself becomes a source of new risks and attack vectors. Intelligent agents with self-learning, adaptation, and autonomous decision-making capabilities can revolutionize threat detection, attack prediction, and incident response mechanisms in real time. However, the same technologies can be used by attackers to automate attacks, bypass traditional security systems, and create difficult-to-detect malware.

^{*} *CPITS-II 2025: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, October 26, 2025, Kyiv, Ukraine*

^{*} Corresponding author.

[†] These authors contributed equally.

✉ ievlev@i.ua (N. Ievlev); klimentko@immsp.kiev.ua (V. Klymenko); amorozov@immsp.kiev.ua (A. Morozov); vitaliy.yashchenko@gmail.com (V. Yashchenko); h.hulak@ukr.net (H. Hulak)

ORCID 0000-0002-9364-9495 (N. Ievlev); 0000-0001-6951-4091 (V. Klymenko); 0000-0002-3923-9495 (A. Morozov); 0000-0002-4168-7190 (V. Yashchenko); 0000-0001-9131-9233 (H. Hulak)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Therefore, there is a need for a comprehensive understanding of the role of AI agents in ensuring the cybersecurity of ITS. Their use requires the development of new security architectures adapted to the rapidly changing threat landscape, as well as ethical, legal, and technical frameworks capable of controlling the behavior of autonomous systems. This becomes especially relevant in the context of emergency situations, when the normal rhythm of system operation is disrupted, and decision-making must be prompt, accurate, and minimize damage. The purpose of this article is to study the impact of AI agents on the cybersecurity architecture of information and telecommunication systems, identify key risks of their use and suggest directions for further research in the field of forming intelligent cyber defense. Particular attention is paid to the analysis of the specifics of using such technologies in wartime and emergency situations, where the effectiveness of ITS protection is directly related to the stability of the state and society as a whole.

2. Intelligent agents in its: concept and functions

In the context of information and telecommunication systems, intelligent agents (AI agents) are software and hardware modules capable of autonomous analysis, decision-making, and interaction with other components of the digital infrastructure. They use artificial intelligence methods, such as machine learning, natural language processing, pattern and anomaly recognition, to solve a wide range of problems related to ensuring the stability and security of ITS.

AI agents can exist both as separate software components (e.g., chatbots or voice assistants), and be part of complex incident management systems (e.g., SOAR platforms — Security Orchestration, Automation, and Response). Their functionality covers several critical areas.

Automation of traffic monitoring. AI agents are capable of analyzing network traffic in real time, identifying suspicious activities, intrusion attempts, data leaks, or malware infections. They reduce the workload of SOC (Security Operations Center) analysts by filtering routine events and forwarding only significant incidents for review.

Intelligent incident response. Modern agents can not only detect threats, but also automatically initiate response actions — blocking IP addresses, isolating infected nodes, changing security policies. In SOAR systems, such actions can be programmed as response scenarios based on risk assessment and incident context.

Consulting users and operators. Chatbots and voice AI assistants built into support services provide information and technical assistance on cybersecurity issues, train employees in the basics of digital hygiene, help identify phishing emails, incorrect settings or suspicious behavior in the workplace.

Detection of abnormal behavior and deviations from the norm. Using behavioral analysis methods, AI agents are able to detect actions that are not typical for the normal operation of users and systems, including slow internal attacks, movement of threats within the network, and insider actions.

Support for situational awareness and crisis management. In emergency situations (ES) or wartime, AI agents play a key role in information support for decision-making. They collect and analyze data from diverse sources, aggregate messages, classify and rank risks, and generate analytical reports in real time [1, 2].

The integration of intelligent agents is carried out at all levels of the ITS: from user devices to cloud infrastructures. In network components, they provide intelligent traffic control [3], in cloud platforms — scalability and resilience of protection, in SOC — support for cyber operators [4], and in situational centers — the formation of an operational situational picture.

Such agents are of particular importance in military conflicts or large-scale emergencies, when control and communication systems are subject to multiple targeted attacks. In these conditions, AI agents provide: accelerated filtering and categorization of information flows, including social networks, sensor networks and telecommunication channels; forecasting the development of cyber threats taking into account the dynamics of attack actions and previous scenarios; support for

coordination of actions between departments, allowing for the rapid transfer of relevant information to various services and decision-making based on a comprehensive analysis.

Intelligent agents thus become critical elements of the resilience of ITs, especially in scenarios where time delays and human error can lead to catastrophic consequences. Their implementation requires not only technological adaptation, but also a rethinking of the digital security architecture as a system with elements of cognitive autonomy.

3. Data protection in situation centers: Cyber and physical security

Situation centers (SCs) are key elements of operational management and coordination in times of crisis and war.

This paper pays special attention to the construction of situational centers (SC) as key nodes for ensuring digital sovereignty, coordinating responses to cyber threats, and managing ITCS in emergency situations and wartime. These centers perform critical functions, from monitoring and filtering information flows to predicting threats and supporting decision-making in real time. Architecturally and conceptually, situational centers are close to industry cybersecurity centers (ICCS), described in a number of scientific papers. In particular, the article by Morozov et al. (2021) proposes an applied approach to building an ICCS as a distributed system for protecting corporate information resources with the ability to integrate at the industry and government levels. This model emphasizes prompt incident handling, collective cyber defense, setting up threat detection tools, and interaction with international response centers (CERT/SOC). Such centers provide information exchange, centralization of analysis, and automation of response to attacks, which significantly increases the resilience of critical infrastructure to external influences [5].

At the same time, Skiter et al. (2021) develop this concept towards a systems approach to creating an ICCS focused on the integrated management of the cybersecurity of critical infrastructure facilities (CIF). The authors propose a hierarchical multi-level protection model covering the base, segment and boundary levels and describe the principles of the architectural construction of the ICCS — integration, modularity, scalability, centralized management and survivability. The proposed model takes into account industry specifics and allows adapting the ICCS to conditions of high uncertainty and technological failures. The role of the analytical stack (from transactional systems to OLAP analytics) as the basis for decision-making at the strategic level is also emphasized [6].

Thus, combining the operational efficiency described by Morozov et al. (2021) and the methodological integrity of the approach of Skiter et al. (2021), situational centers can be viewed as a hybrid model of a new generation of ICCS — intelligent, resilient, capable of autonomous operation and rapid adaptation to the conditions of cyber warfare and hybrid threats.

Given the critical importance of the SC, information protection becomes a task of strategic scale. The threats of cyberattacks, intrusions, sabotage and leaks require a multi-layered security architecture that includes both cyber defense and physical measures.

Cyber threats with examples. Attacks on Ukrainian government IT systems in 2022 included DDoS, data erasure and intrusions into ministry systems. WannaCry (2017) affected global infrastructures using an NSA exploit [7]. Stuxnet (2010) was the first to demonstrate an attack on physical systems via a digital medium [8]. The SolarWinds campaign (2020) was a supply-chain attack that impacted 18,000 organizations [9].

Cybersecurity. Firewalls, IDS/IPS, encryption (TLS/VPN), MFA, auditing and logging of actions.

Physical security. Biometric access control, video surveillance, redundant server rooms and shielding. Incidents in Mariupol and attempted intrusions into EU data centers highlight the importance of physical security of infrastructure [10].

Integrating AI into the protection of situational centers provides: automatic filtering and classification of threats; support for real-time decision-making; coordination of teams and response to complex incidents.

Thus, effective security of the situational center requires a combination of digital and physical measures, and AI is becoming a key component of operational resilience.

4. New threats and vulnerabilities associated with AI

Despite the high efficiency of intelligent agents in ensuring cybersecurity, their active implementation in information and telecommunication systems is associated with the emergence of new threats and vulnerabilities. These risks are often complex in nature, as they are associated not only with the technical aspects of the implementation of AI, but also with the peculiarities of its training, operation and interaction with users.

The following areas pose a particular danger.

4.1. Phishing and social engineering based on generative models

One of the fastest growing areas of AI abuse is the use of generative models (including LLM – large language models). Generative models are used for targeted phishing attacks and deepfake information. Prompt-injection vulnerabilities allow the introduction of hidden commands into AI agents. Adversarial examples and data poisoning attacks can distort results and undermine trust in AI [11, 12].

In military situations, attacks on software and hardware supply chains, including the introduction of backdoors and hidden algorithms, are especially dangerous [13].

4.2. Prompt injection

Language models, especially those embedded in user interfaces or automated support systems, are susceptible to attacks based on the injection of hidden commands into textual interactions. An attacker can surreptitiously modify input data in a way that causes unwanted model behavior, ranging from revealing sensitive information to performing malicious actions.

Since models process text as a single meaning space, they are vulnerable to “reconfiguration” through carefully chosen phrases. This poses a serious threat to trusted AI interfaces that work with sensitive information.

4.3. Uncontrolled data collection and leakage

AI agents, especially those that are trained or operate online, are often involved in collecting, filtering, and analyzing large amounts of data, including personal, corporate, and confidential information. Without proper control over data flows, information can be leaked, either through misconfiguration or deliberate tampering.

In addition, the lack of transparency in how models use post-processing data creates challenges in complying with regulatory requirements (e.g. GDPR, NIS2, national data protection laws).

4.4. Manipulation of AI learning and behavior

Data poisoning and adversarial attacks allow influencing the AI learning process or introducing specially prepared input data that causes misclassification or incorrect decisions. In critical ITS, this can lead to undetected attacks, incorrect threat identification, false positives, or paralysis of response systems [14, 15].

Examples of such attacks include: *substituting* or throwing in training data with deliberately distorted statistics; *using* adversarial examples – input data that is indistinguishable to humans but causes model failures; *injecting* malicious patterns through external APIs, open data feeds, or supply chains.

4.5. Wartime Risks: Supply Chain Vulnerabilities and Sabotage AI

In military conflicts or cyber sabotage, threats associated with the introduction of malicious components into software or hardware modules supplied by external sources become especially relevant. Since AI systems often depend on cloud platforms, external libraries, open source code, and foreign-made equipment, supply chains become a critical element of risk.

Possible scenarios include: *intentional* infection of software components of AI agents with malicious code; *introduction* of backdoor algorithms into neural network models; *changing* AI behavior through remote control or inserting malicious parameters during their update.

Control over such chains requires the implementation of source code verification mechanisms, model integrity monitoring, certification of external supplies, and the creation of fault-tolerant mechanisms in the event of anomalies.

Thus, intelligent agents, despite their significant potential, require a systematic and comprehensive approach to ensuring their security. It is necessary not only to protect traditional elements of the ITS, but also to develop special methods for testing, auditing and monitoring the behavior of AI in real operating conditions, especially in crisis and military scenarios. Only in this case will AI become a reliable partner in ensuring digital sovereignty and sustainability of critical infrastructure.

5. AI as a defense tool

Artificial intelligence (AI) is becoming a key component of modern cybersecurity architecture, enabling a shift from reactive defense models to proactive, adaptive, and autonomous threat response strategies. Integrating AI into ITS not only improves incident response efficiency, but also significantly reduces the workload of information security specialists, increasing the overall resilience of digital infrastructure.

Let's look at the main areas of AI application as a cyber defense tool.

5.1. Anomaly detection and behavioral analysis

Traditional threat detection systems (IDS/IPS) rely on signature-based methods that are effective only against known attacks. In contrast, AI systems use behavioral analysis methods — machine learning and statistical modeling — to build basic activity profiles of users, devices, and applications.

AI agents are capable of detecting anomalies that deviate from normal behavior, including: unauthorized access attempts; *network* movements typical of lateral movement attacks; *internal* threats from compromised accounts or insiders; *unusual* activity outside of business hours, surges in data volumes, and other intrusion signals.

Such systems “learn” from normal behavior and are able to detect zero-day attacks and stealth intrusions that elude signature analysis.

5.2. Predictive analytics and threat forecasting

AI makes it possible to process large amounts of heterogeneous data — event logs, network traffic, reports on malicious activity, threat intelligence — in order to identify patterns and build predictive models [1].

Based on accumulated experience and statistics, AI can: *assess* the likelihood of incidents; *identify* new threats at an early stage; *suggest* preventive measures; *prioritize* vulnerabilities and risks depending on the business context.

This is especially important in situations where security resources are limited and it is necessary to effectively distribute attention among many potential threats.

5.3. Response automation and SOAR platforms

Security Orchestration, Automation and Response (SOAR) platforms use AI to automate the entire incident response cycle: from detection to mitigation. AI agents analyze events in real time, classify them by severity and, if there are clearly defined policies, initiate automated response scenarios.

Examples of actions: *isolating* an infected machine from the network; *blocking* suspicious accounts; *recreating* virtual environments; *notifying* responsible persons and generating reports.

Intelligent bots can also interact with users and operators in a chat, clarify information, confirm incidents and offer solutions.

5.4. Personnel training and AI simulators

The human factor remains one of the main reasons for successful cyberattacks. To improve the cyber literacy of employees, AI solutions are increasingly being used, including: *virtual* simulators that simulate realistic scenarios of phishing attacks, intrusions and other incidents; *personalized* training platforms that adapt materials to the employee's level of knowledge and behavior; *simulating* the actions of attackers in a controlled environment to assess resilience to attacks.

AI allows you to track progress, identify weaknesses in knowledge and generate recommendations for improving security awareness.

5.5. Support of situational awareness centers and decision-making

In crisis management, especially in emergency situations and wartime, AI agents are becoming integral assistants in situational awareness centers. Their functions include: *assessing* the operational situation based on the collection and analysis of data from various sources (telecommunication channels, sensors, intelligence reports, social networks); *visualizing* threats and generating scenarios for the development of events, which allows you to quickly assess the risks and consequences; *support* for coordination between departments — AI helps synchronize team actions, filter information flow, avoid duplication and errors; *intelligent* decision-making prompts based on the analysis of previous situations, assessment of the probability of success of countermeasures and available resources.

This functionality is especially important in conditions of high uncertainty and limited time for decision-making.

Thus, AI not only enhances the capabilities of existing means of protection, but also transforms the very paradigm of cybersecurity, transferring it to the mode of active, intelligent, adaptive defense. The creation of multi-level systems in which AI operates at the level of monitoring, analysis, response and training can significantly increase the resilience of ITS to modern and future threats.

6. Limitations and Challenges

Despite the significant potential of artificial intelligence in ensuring cybersecurity, its implementation is accompanied by a number of technological, ethical, legal and operational challenges. These limitations limit the effectiveness of AI in critical areas and require a systematic approach to address them — both at the technical architecture level and at the regulatory level.

6.1. The problem of explainability and the “black box”

Most modern AI models, especially deep learning neural network architectures, function as “black boxes”: they are able to make highly accurate decisions, but their logic often remains opaque to the user or system administrator.

This leads to the following consequences: *the inability* to interpret why the system classified a certain event as a threat or, conversely, ignored it; *difficulties* in forensic, legal and audit expertise;

decreased trust in the system on the part of operators and decision makers, especially in situations where the justification of actions is critical.

The development of Explainable AI (XAI) methods still lags behind the needs of real-world applications, especially in emergency situations where speed and transparency of decisions are critical.

6.2. Ethical and legal risks of autonomous AI systems

As AI agents become more autonomous, the problem of allocating responsibility for the consequences of their actions arises [16]. In cybersecurity, this is especially acute in scenarios of: *automatic* blocking of users or systems due to errors (false positives), leading to interruptions in operation; *autonomous* decision-making without human participation in a crisis; *unlawful* collection, storage or use of personal data, especially in wartime.

The legislative framework in most countries has not yet been adapted to the specifics of AI functioning, especially in terms of: *determining* the subjects of responsibility (developer, operator, customer); *certification* of AI for critical areas; *ethical* acceptability of decisions made by algorithms without human participation.

6.3. Difficulties in standardization and certification of AI solutions

Unlike traditional software, AI systems are trained on data and can change their behavior depending on the context and input data. This complicates: *verification* of the correctness and security of the model; *repeatability* of results during testing; *compliance* with pre-approved requirements.

The lack of universal standards in the field of cybersecurity of AI systems makes their official certification for use in critical facilities (for example, military, energy or transport infrastructure facilities) almost impossible. The problem is exacerbated by the lack of formalized methods for assessing AI vulnerabilities and scenarios for their exploitation by attackers.

6.4. Vulnerability of situational awareness centers and digital dependence

Situational awareness centers, where AI plays a key role in analyzing the operational situation and coordinating actions, themselves become potential targets of cyberattacks. Their high digital integration makes them vulnerable to the following impacts: *destructive* attacks on communications infrastructure; *disinformation* attacks capable of distorting input data; *DoS attacks* on elements of AI systems and support services; *violations* of the integrity and trust in the sources of data used by AI to build analytical models.

Dependence on external suppliers of cloud services or foreign software components is especially dangerous in the context of geopolitical instability and sanctions pressure.

Thus, the widespread introduction of AI in ITS requires not only technical improvements, but also the development of new approaches to: *explainability* of models; *formalization* of legal liability; *regulatory* certification of AI systems; *increasing* fault tolerance and cyber resilience of control systems.

The solution to these challenges must occur in parallel with the development of the technologies themselves, so that AI becomes not only a tool for enhancing security, but also remains safe, ethical and controllable in itself.

7. Regulation and standardization

The development of artificial intelligence in information and telecommunication systems (ITS), especially in the context of cybersecurity and critical infrastructure, requires reliable legal and technical regulation. In the context of growing threats, the autonomy of AI agents' decisions and their involvement in strategically important processes, the formation of transparent,

internationally recognized standards regulating the life cycle of AI systems — from development and training to operation and certification — is of particular relevance.

7.1. International standards: ISO/IEC 27001 and add-ons

The international standard ISO/IEC 27001 serves as the basis for building an information security management system (ISMS) and is used in most developed countries. In the context of AI solutions, the following add-ons are important: ISO/IEC 27005 (risk management); ISO/IEC TR 24028 — recommendations for the reliability and resilience of AI systems; ISO/IEC 23894 — AI risk assessment.

These documents introduce procedures for vulnerability analysis, real-time monitoring of AI solutions, requirements for model explainability, and algorithm auditing. However, their adaptation to rapidly evolving machine learning methods requires regular updates.

7.2. European regulation: GDPR and NIS2

Within the European Union, two documents play a key role.

GDPR (General Data Protection Regulation): directly affects AI systems, since they process personal data. In particular, GDPR requires: *explainability* of decisions made by automated means; *the possibility* of appeal and human intervention; *restrictions* on the automated processing of sensitive data; *ensuring* the right to be forgotten.

NIS2 (Directive on Security of Network and Information Systems): the new version of the directive strengthens the requirements for critical infrastructure operators, including in terms of managing AI systems, ensuring fault tolerance and rapid response. Measures are introduced to control supply chains, which is important for third-party AI components.

7.3. Regulations for defense and crisis systems

Systems used in defense, intelligence and anti-crisis activities require a special level of trust and control. In this area, the following apply: *closed* standards for automated military systems; *departmental* regulations for testing AI tools for resistance to sabotage, hidden bookmarks, adversarial attacks and information leaks; *requirements* for “integrated security” (security by design), especially for situation centers, early warning systems, communications and coordination.

The use of AI in such systems should be accompanied by mandatory certification for compliance with the requirements of not only functional, but also cyber resilience — taking into account scenarios of communication failure, power outages, active attacks on algorithms and infrastructure.

7.4. Prospects for creating AI certification standards in ITS

Work is underway on the international and national arenas to form specialized certification standards for AI components, including: *creating* uniform criteria for assessing the risk and reliability of algorithms; *standardizing* model testing processes (including attack modeling); *forming* requirements for logging and traceability of AI solutions; *developing* procedures for analyzing bias and uncontrolled behavior of AI in real operating conditions.

Particular attention is paid to AI agents used in situation centers and civil defense and emergency systems, where even a short-term failure can have large-scale consequences. These agents must be: *verified* for the ability to function adequately in conditions of data deficiency; *resistant* to falsification of input information; *capable* of explaining their decisions in a format understandable to a person in a stressful environment.

Thus, the formation of a regulatory and standardized environment for the use of AI in ITS is an essential condition for its effective and safe use. Without clear regulations, transparent evaluation procedures and reliable certification mechanisms, AI systems can turn from a security tool into a

potential vulnerability. Therefore, regulation must develop on a proactive basis — in sync with technological progress and the challenges of the modern digital environment.

8. Development prospects

In the rapidly changing cyber threat landscape, artificial intelligence continues to evolve, forming new concepts and architectures that can provide more flexible, adaptive and sustainable protection of information and telecommunications systems. Among the promising areas, several key development vectors stand out that can radically change approaches to cybersecurity, especially in times of crises and military conflicts.

8.1. Implementation of Zero Trust models with AI-verified behavior

The Zero Trust paradigm is based on the principle that no user or device should have automatic access to resources without constant verification. Integrating AI into this model allows: *to constantly* analyze the behavior of users and devices in real time, identifying anomalies even with valid credentials; *to dynamically* adapt access policies based on context (geolocation, time, type of action, profile risk); *to predict* potential threats and block suspicious activities before incidents occur [17, 18].

This approach significantly reduces the risk of internal attacks and minimizes the possibility of compromising critical resources, which is critical in conditions of limited response time [19, 20].

8.2. Hybrid response teams: synergy of humans and AI

The future of cyber defense is seen in the synergistic interaction of humans and AI. Despite the increased capabilities of automation, the role of an expert remains indispensable—especially in complex situations that require intuition, ethical judgment and strategic thinking.

Within the framework of hybrid teams: *AI agents* take on routine data analysis, primary filtering and automatic execution of standard procedures; *specialists* focus on making decisions in controversial or critical cases, interpreting results and developing strategies; *decision* support systems are used that visualize data and offer options based on machine learning and modeling.

This balance allows for increased response efficiency, reduced likelihood of errors and faster attack countermeasure cycle.

8.3. Creating digital twins of SOC's and situational centers with autonomous decision-making

The concept of digital twins is to create virtual models of real objects, processes or systems capable of autonomous operation and self-learning. In cybersecurity, this means: *developing* automated digital copies of SOC's (Security Operations Centers) and situational centers that can independently analyze incoming data, model the development of events and make decisions without constant human intervention; *using* such twins to test new attack and defense scenarios in conditions as close as possible to real ones; *ensuring* continuous monitoring and adaptation to new threats through integration with external sources of intelligence information and internal management systems.

Digital twins will significantly increase the speed and quality of reactions, as well as minimize the human factor in critical conditions.

8.4. Using Explainable AI (XAI) to Increase Transparency and Trust

One of the key challenges in implementing artificial intelligence in critical areas is the problem of explainability and transparency of decisions made by AI systems. Many modern models, especially those based on deep learning, function as “black boxes” — they are capable of producing highly

accurate results, but the internal mechanisms of decision-making remain hidden and incomprehensible even to their developers and users.

The Explainable AI (XAI) direction is designed to solve this problem. XAI is a technology and method that allows you to create AI models that not only produce results, but are also able to explain the reasons and logic behind their decisions in a language that is understandable to humans. Such transparency is especially important in conditions of high uncertainty, for example, during military operations or crisis situations, when operators and managers must quickly and confidently make decisions based on AI results. The use of XAI helps to increase the level of trust in intelligent systems, facilitates the interpretation of their findings, and ensures more effective human-machine interaction, which is critical for situation centers and operational control systems [21].

Despite the progress in XAI, classical models of explainable AI are still limited in their ability to interpret complex, rapidly changing, and highly contextual threats, especially in conditions of information deficit and uncertainty. This opens up space for the search for new architectures that can not only ensure decision transparency, but also reproduce elements of cognitive flexibility inherent in biological systems. In this context, of particular interest are multidimensional multi-connected receptor-effector neural-like growing networks that are not only highly adaptive and robust, but also potentially capable of developing their own forms of explainability by forming connections and causal structures based on experience and sensory feedback. Such networks may become the next step in the development of XAI – from machine interpretation to ontogenetically conditioned cognitive transparency.

8.5. Application of Multidimensional Multiconnected Receptor-Effector Neural-Like Growing Networks in Cybersecurity Problems

Modern threats in cyberspace are characterized by high dynamism, multi-levelness and unpredictability. Their effective detection and neutralization require intelligent systems capable of adaptive self-learning, generalization of unstructured information and autonomous response in conditions of incomplete and contradictory initial information. One of the promising approaches in this area is the use of multidimensional multiconnected receptor-effector neural-like growing networks (mmrenGN) [22], combining the properties of ontogenetic modeling, cognitive plasticity and high resistance to noise.

9. Multidimensional, multi-connected receptor-effector neural-like growing networks

MmrenGN are neural-like structures capable of continuous growth, topology modification, and self-organization based on incoming input data. Their architecture is based on the separation of input channels (receptors) and output response mechanisms (effectors), which ensures context-dependent interpretation of threats and scalable system response [22].

The concept of mmrenGN is based on the following main provisions

1. The network is a set of a large number of simple elements called neural-like. These elements are functionally similar to neurons of biological nervous systems and have the following properties: they are able to receive information, compare it with previously remembered information, analyze it, modify the stored data if necessary, and remember new, previously unknown information. Each neural-like element functions autonomously, but interacts with other network elements, forming a self-organizing structure capable of learning and adaptation.
2. Neural-like elements are connected to each other by unidirectional links, forming complex multi-connected structures. These links ensure the transmission of signals from one

element to another, creating conditions for the formation of stable paths for processing and storing information. The direction of the links allows you to set the sequence of activation of elements, which is important for the implementation of dynamic processes within the network.

3. Each neural-like element has multiple inputs and one output, through which signals are received and transmitted. The inputs are functionally similar to the dendrites of biological neurons, providing the reception of impulses from other neuron-like elements, and the output corresponds to the axon, transmitting the signal further along the network. This architecture provides a unified model of information processing and allows the implementation of mechanisms for the accumulation, filtering and transformation of signals.
4. The network includes several types of homogeneous neural-like elements, each of which performs specialized functions: Receptor elements are responsible for the reception and primary transformation of input information coming from the external environment. Sensory elements analyze and pre-process incoming signals, identify significant characteristics and transmit information further along the network. Local elements regulate the interaction between sensory and effector elements, participate in the formation of connections, signal transmission routes and adaptation of the network to changes. Effector elements generate control signals aimed at performing actions or reactions of the system in the external or internal environment.
This functional separation facilitates orderly processing of information and allows the system to respond adaptively to various external and internal stimuli.
5. The connections between neural-like elements are characterized by weights that reflect the significance of certain features of perceived concepts, objects, situations, etc. The weight of the connection determines the strength of the transmitted signal, affecting the degree of activation of the receiving element. These weights are not static — they can change during the functioning of the network, which provides the opportunity for learning, accumulation of experience and adaptation to changing environmental conditions. The mechanism of changing the weights serves as the basis for the formation of new associations, generalizations and increased efficiency of information processing.
6. In mmrenGN, based on the interaction of innate (unconditioned) reflexes and incoming sensory information, conditioned reflexes are formed in accordance with the theory of I. P. Pavlov. These conditioned reflexes are based on the fundamental principles of conditioned reflex activity of the brain. Due to this, mmrenGN has the ability to self-organize and self-learn, similar to biological nervous systems. Network learning can occur both under the influence of external factors and internal mechanisms (self-learning), and using external reinforcement — within the framework of learning with a teacher (reinforcement learning). This approach ensures the adaptability and stability of the system's behavior in changing conditions.
7. The network structure changes dynamically during its operation. New neural-like elements and connections can be added or removed depending on the analysis of incoming information and the current conditions of the external and internal environment. Such plasticity of the architecture ensures high adaptability, flexibility and the system's ability to self-regulate, allowing it to effectively respond to new tasks and changing circumstances.
8. All information in the multidimensional structure of mmrenGN is processed in parallel, which ensures high system performance and allows it to effectively solve complex problems in real time. Parallelism of processing helps reduce delays and increases the system's resistance to partial failures.
9. Due to the multidimensional architecture and high degree of connectivity, mmrenGN demonstrates the properties of emergence — when the behavior of the entire system goes beyond the simple sum of the functions of individual elements. This allows us to model

complex, nonlinear phenomena and processes, as well as to identify hidden dependencies and patterns that are not obvious at the level of individual components.

The principles of operation of mmrenGN as part of an intelligent system involve the following main stages

1. Network initialization.

At this stage, the initial structure of the network is formed, with embedded genetic instincts and unconditional reflexes. As a result, the initial number of active neuron-like elements, their types, initial connections between them and their weights are formed.

2. Life cycle:

- Perception of input data and internal parameters. Receptor elements receive input signals from the external environment and from internal sources of the system.
- Propagation of signals within the network. Input signals are propagated along the connections between elements, undergoing certain transformations depending on the weights of the connections and the activation functions of the elements.
- Generation of output actions. Effector elements, receiving signals from other network elements, generate output actions that perform actions, influencing the external environment or other systems.
- Network training and adaptation. During the operation of the network, the weights of the connections are adjusted and the structure of the network is changed (elements and connections are added or removed) in accordance with the results of the information analysis.
- The network continues to function, receiving new input data, processing it and generating output actions, while constantly learning and adapting to changing conditions.

Key capabilities of LDC mmrenGN in cyber defense

1. Ontogenetic recognition of anomalies and new types of attacks. Unlike classic machine learning algorithms, mmrenGN are able to form new internal representations as unknown patterns arrive, including in the absence of labeled data. This is especially relevant for zero-day attacks, hidden internal threats, and attacks using legitimate traffic.
2. Dynamic adaptation to changing threat scenarios. Due to the property of growing networks, the architecture can flexibly scale — activating new receptor chains when atypical inputs appear (for example, non-standard sequences of API calls, network packets with excess entropy, or mutations in malicious code).
3. Context-dependent incident response. The effector component of the mmrenGN implements a contextual response selection system — blocking, isolation, launching SOAR chains — taking into account the current situation, resource priorities, and previous experience. This helps minimize false positives and optimize the use of security resources.
4. Integration with situational centers and decision-making systems. In times of crisis or war, mmrenGN can be integrated into the architecture of situational centers, acting as autonomous cognitive nodes that process information flows from telecommunications, sensory, and intelligence sources. Their ability to hierarchically self-organize allows aggregating different types of signals into high-level risk assessments and threat development scenarios.
5. Operation in conditions of incomplete or noisy data. The use of neural-like models with a weak dependence on the formal structure of input data allows mmrenGN -based systems to function even in conditions of loss of communication with external sources or data distortion due to destructive influences.

Prospects and limitations

The use of multidimensional receptor-effector neural-like growing networks does not require significant computing resources at the stages of training and adaptation. However, the main limitation remains the high time costs due to the sequential nature of data processing on modern computer architectures. In this regard, for the effective practical application of mmrenGN, their hardware implementation is necessary, ensuring absolute parallelism of operation and, accordingly, high speed.

Despite this requirement, the potential of mmrenGN in the field of cybersecurity is difficult to overestimate. These networks open up opportunities for creating cognitively flexible AI agents of a new generation, capable of not only detecting and classifying threats, but also forming internal representations of the digital environment, evolving along with the changing landscape of cyber threats. Thus, only a comprehensive and systematic approach to the development, implementation and regulation of artificial intelligence will allow us to fully realize the potential of AI agents as key elements of cyber defense of information and telecommunication systems in the context of modern challenges, including military conflicts, crisis situations and large-scale destructive impacts. A special role in this process can be played by multidimensional receptor-effector neural-like growing networks that provide adaptive learning, cognitive flexibility and autonomous evolution in the context of a constantly changing cyber landscape. Their hardware-implemented architecture with potentially absolute parallelism opens the way to the creation of highly effective, quickly responding and self-developing intelligent agents of a new generation — agents capable of acting in conditions of uncertainty, information deficit and extremely compressed time frames, maintaining the stability of critical infrastructure and the digital sovereignty of the state.

10. Recommendations for the further development and implementation of AI in the cybersecurity of its

1. Development and implementation of standards and regulations. It is necessary to actively participate in the formation of international and national standards that will regulate the development, testing and operation of AI systems in critical infrastructure, taking into account the specifics of war and crisis times.
2. Investing in Explainable AI (XAI) research. To increase trust in AI solutions, it is important to develop technologies that can ensure the transparency and interpretability of AI decisions, which is critical in situations where decisions affect the safety and lives of people.
3. Creation of hybrid human-machine systems. It is important to develop architectures where AI solutions are accompanied by human control and support, ensuring a balance between automation and human oversight.
4. Upskilling human resources. It is necessary to invest in training and developing human resources that can effectively interact with AI tools, understanding their capabilities and limitations.
5. Strengthening supply chain controls. Rigorous verification and certification measures should be implemented for AI software and hardware components to minimize the risk of introducing malicious elements into critical systems.
6. Evolving Zero Trust models with AI verification. It is recommended to integrate AI into zero trust security models for continuous monitoring and adaptive access control.
7. Hardware implementation of mmrenGN and their application in cybersecurity tasks.
8. Ensuring fault tolerance and cyber resilience of systems. ITS should be designed in such a way that AI systems can function effectively even in conditions of limited resources, communication interruptions, and under the influence of active attacks.

11. Conclusion

In the modern world, where digital infrastructure is becoming the foundation of the economy, public administration and social life, the role of artificial intelligence in ensuring cybersecurity of information and telecommunications systems is of exceptional importance. AI agents are already transforming approaches to data and infrastructure protection, providing higher threat detection speed, response efficiency and system adaptability in the face of ever more complex cyberattacks.

This need is especially acute in wartime and crisis situations, when information flows are critical and the speed and accuracy of decision-making become a matter of national security. In such circumstances, intelligent agents play a key role in situational centers, providing filtering, analysis and forecasting of threats in real time, as well as coordination of actions of various structures.

However, the development and application of AI in the field of cybersecurity is associated with a number of serious challenges. Technological limitations, such as the problem of explainability of decisions, questions of trust in “black boxes”, vulnerability of AI models to adversarial attacks, require active research and implementation of new methods, such as Explainable AI (XAI). In addition, ethical and legal aspects related to the autonomy of AI and the distribution of responsibility require the development of clear regulations and standards. An important area is the development of unified international and national standards to ensure the certification and control of AI systems, especially in critical areas, including defense, energy and civil defense. The prospects for the development of cybersecurity are closely related to the integration of AI and humans into hybrid response systems, where AI acts not as a replacement, but as an amplifier of human capabilities. The synergy of human experience, intuition and strategic thinking with the computing power and speed of AI allows us to create multi-level, adaptive and resilient defense systems that can effectively counter both known and new threats. In the future, special attention will be paid to Zero Trust models, expanding the use of digital twins of situation centers and SOCs, as well as the development of explainable AI methods that increase trust and transparency of decisions. This will not only increase the sustainability and fault tolerance of the ITS, but also ensure reliable coordination and management in conditions of high uncertainty typical of war and crisis times.

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication’s content.

References

- [1] G. Sharma, The Role of Artificial Intelligence Technologies in Crisis Response, *J. Safety Crisis Manag.* 5(2) (2019) 120–135.
- [2] G. Sidiropoulos, Ch. Kiourt, L. Moussiades, Metis: Multi Agent Based Crisis Simulation System, *ArXiv*, 2020. doi:10.48550/arXiv.2009.03934
- [3] D. Verma, S. Calo, Using AI/ML to Gain Situational Understanding from Passive Network Observations, *arXiv*, 2019. doi:10.48550/arXiv.1910.06266
- [4] Cyrebro, AI Based SOCs: Smarter, Faster, and More Secure (Part 1), 2025. <https://www.cyrebro.io/blog/ai-based-socs-smarter-faster-and-more-secure-part-1/>
- [5] A. Morozov, A. Hrebennyk, E. Trunova, I. Skiter, E. Hulak, Design of Industry Centers of Cyber Security of Facilities of Critical Infrastructure, in: *Cybersecurity Providing in Information and Telecommunication Systems (CPITS-II-2021)*, vol. 2923, 2021, 27–37.

- [6] I. Skiter, H. Hulak, V. Grechaninov, V. Klymenko, N. Ievlev, System Approach to the Creation of Cybersecurity Centers of Critical Infrastructure, in: Cybersecurity Providing in Information and Telecommunication Systems (CPITS-II-2021), vol. 2923, 2021, 244–250.
- [7] Putting the Eternal in EternalBlue: Mapping the use of the infamous exploit. Trend Micro Security Blog.
- [8] D. Kushner, The real story of Stuxnet, IEEE Spectrum, 2013. <https://spectrum.ieee.org/the-real-story-of-stuxnet>
- [9] CIS, The SolarWinds cyber-attack: What you need to know, Center for Internet Security, 2021.
- [10] UNIAN, Communication systems destroyed in Mariupol, 2022. <https://www.unian.net>
- [11] I. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, in: 3rd Int. Conf. on Learning Representations (ICLR), 2015.
- [12] N. Carlini, D. Wagner, Adversarial examples are not easily detected: Bypassing ten detection methods, in: 10th ACM Workshop on Artificial Intelligence and Security, 2017.
- [13] N. Weaver, Supply chain attacks in the modern era, Wired, 2021. <https://www.wired.com/story/hacker-lexicon-what-is-a-supply-chain-attack>
- [14] P. Skladannyi, et al., Improving the Security Policy of the Distance Learning System based on the Zero Trust Concept, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3421 (2023) 97–106.
- [15] R. Syrotynskyi, et al., Methodology of Network Infrastructure Analysis as Part of Migration to Zero-Trust Architecture, in: Cyber Security and Data Protection, vol. 3800 (2024) 97–105.
- [16] L. Floridi, et al., AI4People—An Ethical Framework for a Good AI Society, Minds and Machines, 28(4) (2018) 689–707.
- [17] I. Hanhalo, et al., Adaptive Approach to Ensuring the Functional Stability of Corporate Educational Platforms under Dynamic Cyber Threats, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, vol. 3991 (2025) 481–491
- [18] P. Skladannyi, et al., Model and Methodology for the Formation of Adaptive Security Profiles for the Protection of Wireless Networks in the Face of Dynamic Cyber Threats, in: Cyber Security and Data Protection, vol. 4042 (2025) 17–36.
- [19] S. Shevchenko, Y. Zhdanova, O. Kryvytska, H. Shevchenko, Fuzzy Cognitive Mapping as a Scenario Approach for Information Security Risk Analysis, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3826, 2024, 356–362."
- [20] P. Petriv, I. Opirskyy, N. Mazur, Modern Technologies of Decentralized Databases, Authentication, and Authorization Methods, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3826, 2024, 60–71.
- [21] B.S. Raji, et al., Explainable AI: A Review of Methods and Applications, Machine Learning Research, 156 (2022) 1–49.
- [22] V. Yashchenko, Neural-like Growing Networks the Artificial Intelligence Basic Structure, Intelligent Systems in Science and Information 2014, Studies in Computational Intelligence, vol. 591, 2015, 41–55. doi:10.1007/978-3-319-14654-6_3