Вісник ЛДУБЖД
Bulletin of Lviv State University of Life Safety
https://journal.ldubgd.edu.ua/index.php/Visnuk
ISSN 2078-4643 (print), ISSN 2708-1389 (online)

UDC 004.056:004.8
DOI: 10.32447/20784643.33.2026.16

А. І. Івануса, А. В. Ільків, Н. О. Маслова, В. С. Балацька, М. І. Николайчук
Львівський державний університет безпеки життєдіяльності, м. Львів, Україна
ORCID: https://orcid.org/0000-0001-9141-8039 – А. І. Івануса
https://orcid.org/0009-0009-6330-1664 – А. В. Ільків
https://orcid.org/0000-0002-9078-0973 – Н. О. Маслова
https://orcid.org/0000-0002-6262-6792 – В. С. Балацька
https://orcid.org/0009-0004-6313-5196 – М. І. Николайчук
ivaaanusa@gmail.com

АНАЛІЗ РОБОЧОГО ІНСТРУМЕНТАРІЮ УПРАВЛІННЯ РИЗИКАМИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ

Стрімкий розвиток і впровадження технологій штучного інтелекту зумовлює суттєві трансформації у сфері кібербезпеки, підвищуючи ефективність виявлення та протидії кіберзагрозам. Водночас використання інтелектуальних алгоритмів у процесах аналізу даних, автоматизованого реагування на кіберінциденти та прийняття управлінських рішень породжує нові види кіберризиків, пов'язаних із уразливістю моделей машинного навчання, непрозорістю алгоритмів, можливістю маніпулювання даними та порушенням інформаційної безпеки.

Додатково слід враховувати сучасні оцінки загроз у сфері штучного інтелекту, представлені у звіті ENISA Threat Landscape for Artificial Intelligence, де систематизовано ключові вектори атак на AI-системи, зокрема маніпуляції навчальними даними (data poisoning), використання adversarial-прикладів, компрометацію моделей та зловживання автоматизованими рішеннями. Це підтверджує необхідність адаптації традиційних підходів до управління ризиками з урахуванням специфіки інтелектуальних систем.

Сучасні дослідження також підкреслюють взаємозв'язок кібербезпеки з питаннями захисту інтелектуальної власності, прав людини та сталого розвитку цифрового середовища. Зокрема, у дослідженні Bani-Meqdad et al. доведено, що розвиток цифрового середовища супроводжується зростанням ризиків порушення прав інтелектуальної власності та несанкціонованого використання даних.

Крім того, розвиток технологій комп'ютерного зору створює нові можливості для підвищення ефективності систем безпеки. Наприклад, методи автоматичного виявлення полум'я на основі LBP-дескрипторів демонструють можливість високоточного виявлення небезпечних подій у реальному часі. Водночас застосування таких підходів потребує врахування ризиків маніпуляції відеоданими.

Додаткові виклики виникають у сфері біометричної ідентифікації. Використання методів Ateb-Gabor дозволяє підвищити точність оброблення біометричних зображень, проте створює ризики помилкової ідентифікації. Наявні наукові дослідження переважно зосереджуються на технічних перевагах застосування штучного інтелекту в кіберзахисті, тоді як питання комплексної оцінки ризиків, управління ними та нормативно-правового забезпечення кібербезпеки залишаються недостатньо систематизованими. Це зумовлює необхідність формування цілісного наукового підходу до аналізу ризиків використання штучного інтелекту в кіберпросторі з урахуванням технічних, правових та етичних аспектів.

У статті здійснено комплексний аналіз сучасних методик управління ризиками інформаційної безпеки з урахуванням специфіки використання технологій штучного інтелекту. Особливу увагу приділено інтеграції сучасних міжнародних стандартів управління ризиками AI, зокрема ISO/IEC 23894:2023, ISO/IEC 42001:2023 та NIST AI Risk Management Framework. Запропоновано класифікацію методик та адаптивну модель управління ризиками.

Метою дослідження є аналіз існуючого робочого інструментарію управління ризиками у сфері інформаційної безпеки з метою вивчення умов інтеграції систем штучного інтелекту в кібербезпекову інфраструктуру підприємств. Методологічну основу становлять методи системного та порівняльного аналізу, узагальнення, класифікації, а також ризик-орієнтований підхід до оцінювання кіберзагроз і вразливостей інтелектуальних систем.

У роботі використано системний, порівняльний та ризик-орієнтований підходи. Проаналізовано стандарти: ISO/IEC 27005:2022, ISO/IEC 23894:2023, ISO/IEC 42001:2023, NIST Risk Management Framework, NIST AI Risk Management Framework.

Наукова новизна одержаних результатів полягає у систематизації методик управління ризиками інформаційної безпеки та кіберзахисту за рівнем завдання шкоди, удосконаленні підходу до їх класифікації за характером оцінювання (якісні, кількісні, сценарні) та рівнем управлінської інтеграції. На відміну від існуючих досліджень, у роботі враховано новітні міжнародні стандарти управління ризиками штучного інтелекту (ISO/IEC 23894, NIST AI RMF), що забезпечує актуалізацію підходів до оцінювання AI-загроз. Запропонований системний підхід може бути використаний під час розроблення політик кібербезпеки, впровадження систем управління інформаційною безпекою та модернізації процедур оцінювання ризиків у державних і корпоративних інформаційних системах.

Ключові слова: кібербезпека; штучний інтелект; інформаційна безпека; машинне навчання; управління ризиками.

A. I. Ivanusa, A. V. Ilkiv, N. O. Maslova, V. S. Balatska, M. I. Nykolaichuk
Lviv State University of Life Safety, Lviv, Ukraine

ANALYSIS OF INFORMATION SECURITY RISK MANAGEMENT TOOLKIT USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES

The rapid development and widespread implementation of artificial intelligence technologies are driving significant transformations in the field of cybersecurity, enhancing the effectiveness of detecting and countering cyber threats. At the same time, the use of intelligent algorithms in data analysis, automated incident response, and managerial decision-making generates new types of cyber risks associated with the vulnerability of machine learning models, the opacity of algorithms, the possibility of data manipulation, and violations of information security.

Additionally, contemporary assessments of threats in the field of artificial intelligence, presented in the ENISA Threat Landscape for Artificial Intelligence report, identify key attack vectors targeting AI systems, including data poisoning, adversarial examples, model compromise, and the misuse of automated decision-making. This confirms the need to adapt traditional risk management approaches to the specific characteristics of intelligent systems.

Modern research also emphasizes the interconnection between cybersecurity and issues of intellectual property protection, human rights, and sustainable digital development. In particular, the study by Bani-Meqdad et al. demonstrates that the evolution of the digital environment is accompanied by increasing risks of intellectual property rights violations and unauthorized use of data.

Furthermore, the development of computer vision technologies creates new opportunities to enhance the effectiveness of security systems. For example, automatic flame detection methods based on Local Binary Patterns (LBP) descriptors demonstrate high-precision real-time detection of hazardous events. However, the application of such approaches requires consideration of risks related to video data manipulation.

Additional challenges arise in the field of biometric identification. The use of Ateb-Gabor methods improves the accuracy of biometric image processing but also introduces risks of false identification. Existing research primarily focuses on the technical advantages of artificial intelligence in cybersecurity, while issues of comprehensive risk assessment, risk management, and regulatory support remain insufficiently systematized. This necessitates the development of an integrated scientific approach to analyzing AI-related risks in cyberspace, taking into account technical, legal, and ethical aspects.

The article provides a comprehensive analysis of modern information security risk management methodologies, considering the specifics of artificial intelligence technologies. Particular attention is paid to the integration of international AI risk management standards, including ISO/IEC 23894:2023, ISO/IEC 42001:2023, and the NIST AI Risk Management Framework. A classification of methodologies and an adaptive risk management model are proposed.

The purpose of the study is to analyze the existing practical toolkit for risk management in the field of information security in order to examine the conditions for integrating artificial intelligence systems into enterprise cybersecurity infrastructures. The methodological basis includes system and comparative analysis, generalization, classification, and a risk-oriented approach to assessing cyber threats and vulnerabilities of intelligent systems.

The study employs systemic, comparative, and risk-oriented approaches. The following standards are analyzed: ISO/IEC 27005:2022, ISO/IEC 23894:2023, ISO/IEC 42001:2023, the NIST Risk Management Framework, and the NIST AI Risk Management Framework.

The scientific novelty of the obtained results lies in the systematization of information security and cybersecurity risk management methodologies based on the level of potential damage, as well as in improving their classification according to the nature of assessment (qualitative, quantitative, and scenario-based) and the level of managerial integration. Unlike existing studies, this work incorporates the latest international standards for artificial intelligence risk management (ISO/IEC 23894, NIST AI RMF), which ensures the actualization of approaches to assessing AI-related threats. The proposed systemic approach can be applied in the development of cybersecurity policies, implementation of information security management systems, and modernization of risk assessment procedures in governmental and corporate information systems.

Keywords: cybersecurity; artificial intelligence; information security; machine learning; risk management.

Вступ. Активна цифрова трансформація суспільства та впровадження технологій штучного інтелекту в державні, військові, фінансові та промислові системи зумовлюють появу нових класів кіберзагроз. На відміну від традиційних інформаційних систем, AI-системи характеризуються адаптивністю, здатністю до самонавчання та високим рівнем невизначеності поведінки, що ускладнює застосування класичних моделей оцінювання ризиків.

Аналіз сучасних досліджень свідчить про зростання кількості публікацій, присвячених безпеці машинного навчання, adversarial-атакам та управлінню ризиками.

Окремий напрям досліджень присвячений формалізації сценаріїв кібератак за допомогою моделей attack tree. У роботі Mauw та Oostdijk [21] запропоновано логічну модель дерев атак, що дозволяє формалізувати властивості безпеки. Подальший розвиток цього підходу представлено у дослідженні Kordy та ін. [20], де розроблено алгоритми кількісного аналізу дерев атак. Використання таких підходів є перспективним для аналізу загроз, спрямованих на системи штучного інтелекту.

Водночас відсутня узагальнена систематизація наявних методик управління ризиками з точки зору їх придатності до AI-середовища.

Метою дослідження є аналіз існуючого робочого інструментарію управління ризиками у сфері інформаційної безпеки з метою вивчення умов інтеграції систем штучного інтелекту в кібербезпекову інфраструктуру підприємств. Методологічну основу становлять методи системного та порівняльного аналізу, узагальнення, класифікації, а також ризик-орієнтований підхід до оцінювання кіберзагроз і вразливостей інтелектуальних систем.

Об'єктом дослідження є процеси управління ризиками у сфері інформаційної та кібербезпеки підприємств в умовах інтеграції систем штучного інтелекту. **Предметом дослідження** є методи, моделі та інструментарій управління ризиками інформаційної безпеки, а також умови й особливості інтеграції систем штучного інтелекту в кібербезпекову інфраструктуру підприємств.

Методи дослідження. У дослідженні виконано методи контент-аналізу, порівняльного аналізу та узагальнення. Методологічні підходи класифіковано за рівнями управління: технічний (модельний захист, безпека даних), організаційний (процеси управління ризиками), стратегічний (корпоративне та державне управління).

Наукова новизна дослідження полягає у систематизації методик управління ризиками інформаційної безпеки та кіберзахисту за рівнем завдання шкоди, удосконаленні підходу до їх класифікації за характером оцінювання (якісні, кількісні, сценарні) та рівнем управлінської інтеграції. На відміну від існуючих досліджень, у роботі враховано новітні міжнародні стандарти управління ризиками штучного інтелекту (ISO/IEC 23894, NIST AI RMF), що забезпечує актуалізацію підходів до оцінювання AI-загроз. Запропонований системний підхід може бути використаний під час розроблення політик кібербезпеки, впровадження систем управління інформаційною безпекою та модернізації процедур оцінювання ризиків у державних і корпоративних інформаційних системах.

Результати досліджень. Методика ISO/IEC 27005 є складовою сімейства стандартів ISO/IEC 27000 та підтримує впровадження системи управління інформаційною безпекою (ISMS).

Сильні сторони:

-сумісність з ISO/IEC 27001, що є критично важливим для організацій, які впроваджують сертифіковані системи управління;

-можливість інтеграції ризиків ШІ на рівні активів (моделі, дані, алгоритми);

-підтримка безперервного моніторингу ризиків.

Обмеження:

-відсутність конкретних інструментів кількісної оцінки ризиків ШІ;

-недостатня увага до сценаріїв атак на навчальні дані та моделі машинного навчання.

Методика ISO/IEC 27005 забезпечує системний і стандартизований підхід до управління ризиками інформаційної безпеки та є придатною для інтеграції ризиків використання штучного інтелекту в межах ISMS. Водночас вона потребує доповнення спеціалізованими інструментами для аналізу динамічних і модель-орієнтованих загроз ШІ [17].

Фреймворк NIST Risk Management Framework розроблений Національним інститутом стандартів і технологій США та інтегрує управління ризиками у життєвий цикл інформаційних систем [12; 26; 27].

Сильні сторони:

-орієнтація на безперервне управління ризиками;

-висока відповідність регуляторним вимогам;

-можливість оцінювання ефективності заходів безпеки ШІ-систем.

Обмеження:

-складність впровадження та значні ресурсні витрати;

-обмежена адаптація до динаміки самонавчальних моделей.

Фреймворк NIST RMF [10; 11] інтегрує управління ризиками у життєвий цикл

інформаційних систем, включно з системами на основі штучного інтелекту. Методика є ефективною для критичних і державних систем, однак характеризується високою складністю впровадження та значними ресурсними витратами.

Методика OCTAVE [1–4] зосереджується на організаційних і бізнес-ризиках та широко застосовується у державному секторі, забезпечує високий рівень відповідності регуляторним вимогам.2

Сильні сторони:

-урахування людського фактора;

-аналіз ризиків неправильного використання або надмірної довіри до ШІ;

-орієнтація на стратегічні наслідки для організації.

Обмеження:

-низька деталізація технічних ризиків ШІ;

-обмежена придатність для аналізу складних кіберзагроз.

Методика OCTAVE орієнтована на аналіз організаційних і бізнес-ризиків, що є важливим у контексті управлінського використання штучного інтелекту. Разом із тим вона має обмежені можливості для детального аналізу технічних уразливостей інтелектуальних систем.

FAIR (Factor Analysis of Information Risk) [15] є методикою кількісного аналізу ризиків, що дозволяє оцінювати кіберризики у фінансовому вимірі та підтримує обґрунтоване управлінське рішення.

Сильні сторони:

-підтримка управлінських рішень щодо інвестицій у безпеку ШІ;

-можливість оцінки економічних наслідків помилкових рішень ШІ;

-об’єктивізація ризиків.

Обмеження:

-потребує великого обсягу статистичних даних;

-складна для застосування в умовах нових і невивчених загроз ШІ.

Методика FAIR забезпечує кількісну оцінку кіберризиків у фінансовому вимірі та підтримує прийняття економічно обґрунтованих рішень щодо захисту систем штучного інтелекту. Її застосування ускладнюється потребою в достовірних статистичних даних та експертних оцінках.

Фреймворк COBIT [13] корпоративного управління IT, у межах якого управління ризиками розглядається як ключовий процес забезпечення відповідності та ефективності IT-діяльності.

Сильні сторони:

-чітка орієнтація на керівний рівень;

-інтеграція ризиків ШІ у систему корпоративного контролю;

-сумісність з іншими стандартами.

Обмеження:

-низька деталізація технічних загроз;

-непридатність для аналізу конкретних атак на ШІ.

COBIT є фреймворком корпоративного управління IT, який дозволяє інтегрувати ризики використання штучного інтелекту на стратегічному рівні управління. Водночас методика не призначена для аналізу конкретних технічних сценаріїв кібератак.

CRAMM – це класична методика аналізу та управління ризиками, що використовує формалізовані шкали оцінювання активів, загроз і вразливостей [6].

Сильні сторони:

-структурований підхід;

-детальний аналіз активів і загроз;

-придатність для складних систем.

Обмеження:

-трудомісткість і застарілі підходи;

-слабка адаптація до динамічних моделей ШІ.

CRAMM забезпечує формалізований і детальний аналіз ризиків на основі оцінювання активів, загроз і вразливостей. Методика є трудомісткою та менш адаптованою до швидкозмінних загроз, характерних для систем штучного інтелекту.

EBIOS Risk Manager [14] – це європейська методика управління ризиками, орієнтована на сценарний аналіз атак з урахуванням мотивації та можливостей зловмисників.

Сильні сторони:

-урахування мотивації та можливостей атакувальника;

-ефективність для аналізу атак на ШІ-моделі;

-стратегічна орієнтація.

Обмеження:

-складність методики;

-потреба у високій кваліфікації експертів.

EBIOS Risk Manager базується на сценарному аналізі ризиків з урахуванням мотивації та можливостей атакувальника, що робить її придатною для оцінювання складних атак на системи штучного інтелекту. Разом із тим методика потребує високого рівня експертної підготовки та значних аналітичних ресурсів.

Методика моделювання загроз Microsoft Threat Modeling (STRIDE) [28] широко застосовується під час проєктування програмних та інформаційних систем з метою реалізації принципу Secure by Design.

Категорії загроз STRIDE:

-Spoofing (підміна);

-Tampering (модифікація);

-Repudiation (відмова від дій);

-Information Disclosure (розголошення інформації);

-Denial of Service (відмова в обслуговуванні);

-Elevation of Privilege (підвищення привілеїв).

Сильні сторони:

-підтримка принципу Secure by Design;

-простота застосування;

-ефективність на ранніх етапах розробки ШІ-систем.

Обмеження:

-не враховує бізнес-контекст;

-обмежена після впровадження системи.

Методика STRIDE є ефективним інструментом ідентифікації загроз на етапі проєктування систем і підтримує реалізацію принципу Secure by Design для ШІ-рішень. Вона не охоплює повний цикл управління ризиками після впровадження системи.

Attack Tree Analysis [20–22; 30] це методика аналізу сценаріїв атак ризиків шляхом побудови дерева можливих сценаріїв реалізації атаки, що дозволяє виявити критичні вектори загроз та їх комбінації.

Сильні сторони:

-ефективне виявлення критичних векторів атак;

-придатність для аналізу складних багатокрокових атак на ШІ.

Обмеження:

-складність масштабування;

-потреба в експертних знаннях.

Attack Tree Analysis дозволяє моделювати сценарії реалізації кібератак і виявляти критичні вектори загроз, зокрема щодо компонентів штучного інтелекту. Обмеженням методики є її залежність від експертних знань і складність масштабування. Сучасні підходи до моделювання атак доповнюються фреймворком MITRE ATLAS, який орієнтований на аналіз атак саме на AI-системи.

Risk-Based Vulnerability Management [29] це підхід до управління вразливостями, який базується на оцінюванні ризиків з урахуванням реального бізнес-контексту та актуальних загроз.

Сильні сторони:

-урахування реального контексту експлуатації;

-ефективна пріоритезація вразливостей ШІ;

-підтримка оперативного управління ризиками.

Обмеження:

-залежність від актуальності threat intelligence;

-складність автоматизації для ШІ-моделей.

Risk-Based Vulnerability Management забезпечує пріоритезацію вразливостей з урахуванням реального бізнес-контексту та актуальних загроз, що є особливо важливим для систем із використанням штучного інтелекту. Ефективність підходу залежить від якості аналітики загроз і постійного оновлення даних.

Таблиця 1

Характеристики методик управління ризиками у сфері кібербезпеки

№	*Методика*	*Тип оцінювання*	*Рівень інтеграції*	*Перевага*	*Адаптивність до AI*
1	ISO/IEC 27005	Якісна	Організаційний	Сумісність з ISMS	Середня
2	NIST RMF	Мішана	Стратегічний	Регуляторна відповідність	Середня
3	OCTAVE	Якісна	Організаційний	Фокус на бізнесі	Низька
4	FAIR	Кількісна	Аналітичний	Фінансова оцінка ризику	Висока
5	COBIT	Якісна	Стратегічний	Корпоративне управління	Середня
6	CRAMM	Якісна	Технічний	Деталізований аналіз	Низька
7	EBIOS RM	Сценарна	Стратегічний	Аналіз цільових атак	Висока
8	STRIDE	Сценарна	Технічний	Secure by Design	Середня
9	Attack Tree	Сценарна	Аналітичний	Моделювання атак	Висока
10	RBVM	Мішана	Операційний	Динамічна пріоритезація	Висока

Систематизація методологій управління кіберризиками, пов'язаними зі ШІ. У результаті проведеного порівняльного аналізу встановлено, що сучасні методики управління ризиками кібербезпеки можуть бути систематизовані за трьома ключовими критеріями:

Тип оцінювання ризиків:

-якісні (ISO/IEC 27005, OCTAVE);

-кількісні (FAIR);

-сценарні (EBIOS, Attack Tree);

-мішані (NIST RMF, RBVM).

Рівень управлінської інтеграції

-стратегічний (COBIT, NIST RMF);

-організаційний (ISO/IEC 27005, OCTAVE);

-технічний/операційний (STRIDE [28], Risk-Based Vulnerability Management]);

-аналітичний (FAIR, Attack Tree Analysis).

Рівень адаптивності до ризиків штучного інтелекту

-низький (CRAMM, OCTAVE);

-середній (ISO/IEC 27005, COBIT, STRIDE);

-високий (EBIOS Risk Manager, FAIR, Risk-Based Vulnerability Management).

Уперше в межах дослідження запропоновано класифікацію методик за рівнем їх адаптивності до специфічних ризиків штучного інтелекту, що враховує такі фактори:

-можливість аналізу атак на навчальні дані (data poisoning);

-врахування вразливостей моделей машинного навчання;

-підтримку сценаріїв adversarial attacks;

-оцінювання автономності прийняття рішень;

-інтеграцію threat intelligence у реальному часі.

Встановлено, що традиційні стандартизовані підходи (ISO/IEC 27005 [6; 7], NIST RMF [10; 11]) забезпечують високий рівень управлінської системності, проте потребують доповнення спеціалізованими сценарними та кількісними інструментами для аналізу AI-специфічних загроз.

Порівняльна матриця методологій. На основі проведеного аналізу сформовано узагальнену порівняльну матрицю, яка дозволяє оцінити методики за такими параметрами:

-глибина аналізу технічних уразливостей;

-орієнтація на бізнес-контекст;

-можливість кількісної оцінки;

-придатність для критичної інфраструктури;

-рівень ресурсомісткості впровадження.

Отримані результати свідчать, що:

-FAIR є найбільш придатною для економічного обґрунтування інвестицій у захист AI-систем;

-EBIOS Risk Manager демонструє найвищу ефективність для моделювання цілеспрямованих атак на моделі ШІ;

-Risk-Based Vulnerability Management забезпечує оперативну адаптацію до динамічних загроз;

-STRIDE є ефективною на етапі проєктування AI-систем;

-COBIT та NIST RMF доцільні для стратегічного впровадження комплексної політики управління ризиками.

Таблиця 2

Характеристика наявності використання штучного інтелекту в методиках управління ризиками кібербезпеки

Методика	***Використання ШІ***	***Характер застосування***	***Рівень інтеграції***
ISO/IEC 27005	Опосередковане	Моніторинг ризиків	Низький-середній
NIST RMF	Опосередковане	Оцінка ефективності контролів	Середній
OCTAVE	Обмежене	Підтримка рішень	Низький
FAIR	Часткове	Прогнозування втрат	Середній
COBIT	Обмежене	Стратегічна аналітика	Низький
CRAMM	Майже відсутнє	Експертна оцінка	Дуже низький
EBIOS RM	Часткове	Сценарне моделювання	Середній-високий
STRIDE	Часткове	Threat modeling	Середній
Attack Tree	Часткове	Аналіз сценаріїв	Середній
RBVM	Активне	Threat intelligence	Високий

Запропонована адаптивна модель для управління кіберризиками, пов'язаними з ШІ. На підставі узагальнення результатів запропоновано адаптивну модель управління ризиками використання штучного інтелекту в кібербезпеці, що передбачає поетапну інтеграцію різних типів методик:

Етап 1 – Стратегічна рамка

Використання NIST RMF або ISO/IEC 27005 для формування політики та визначення контексту ризиків.

Етап 2 – Сценарний аналіз AI-загроз

Застосування EBIOS Risk Manager та Attack Tree Analysis для моделювання атак на моделі машинного навчання.

Етап 3 – Технічне моделювання загроз

Використання STRIDE для виявлення вразливостей на етапі проєктування.

Етап 4 – Кількісна оцінка впливу

Застосування FAIR для оцінювання економічних наслідків реалізації AI-ризиків.

Етап 5 – Операційна пріоритезація

Імплементація Risk-Based Vulnerability Management для безперервного моніторингу та оновлення ризикового профілю.

Запропонована модель забезпечує поєднання стратегічного, аналітичного та технічного рівнів управління ризиками, що підвищує адаптивність систем кібербезпеки до швидкоплинних AI-загроз.

Обговорення результатів проведеного аналізу. Проведений аналіз показав, що класичні фреймворки управління ризиками орієнтовані переважно на статичні активи та передбачувані сценарії загроз [8–12]. Сучасні AI-системи породжують нові ризики: data poisoning,

adversarial attacks, model drift. Дослідження у сфері adversarial machine learning підтверджують наявність нових класів загроз [25; 26]. Регуляторний підхід представлений AI Act ЄС [16], а також MITRE ATLAS [27]. Зокрема, сучасні дослідження у сфері аналізу зображень, відеоспостереження та біометричних технологій демонструють значний потенціал використання алгоритмів машинного навчання для підвищення ефективності систем безпеки. Наприклад, застосування відеоаналітики на основі LBP-дескрипторів дозволяє автоматично виявляти небезпечні події у відеопотоці з високою точністю [23; 24]. Аналогічно, використання математичних моделей Ateb-Gabor для фільтрації біометричних зображень підвищує ефективність розпізнавання та аналізу біометричних даних [25]. Проте інтеграція таких алгоритмів у системи безпеки водночас створює нові ризики, пов'язані з можливістю підміни біометричних даних, маніпуляцією відеопотоком та використанням adversarial-прикладів для обману систем машинного навчання.

Встановлено, що:

ISO та NIST забезпечують структуровану основу управління ризиками інформаційної безпеки та формують стандартизований підхід до побудови систем управління кібербезпекою;

FAIR дозволяє кількісно оцінювати ризики, проте складно адаптується до поведінкових моделей ШІ;

STRIDE та подібні методи ефективні для моделювання загроз, але потребують розширення для AI-архітектур.

Поглиблений аналіз продемонстрував, що трансформація ризик-менеджменту в умовах використання штучного інтелекту має відбуватися не лише на рівні технічних засобів захисту, а й на рівні управлінських та нормативних механізмів [19–21; 26]. На відміну від класичних інформаційних систем, AI-рішення функціонують у середовищі постійної зміни даних, що зумовлює емерджентні властивості та нелінійні ефекти. Це означає, що навіть за відсутності явної вразливості модель може демонструвати небажану або непередбачувану поведінку в нових контекстах застосування.

Порівняльний аналіз показав, що більшість традиційних методик базуються на припущенні стабільності активів і відносної передбачуваності сценаріїв атак. У випадку AI-систем ці припущення не завжди виконуються. Наприклад, ризики data poisoning або model drift можуть реалізовуватися поступово та залишатися непоміченими в межах стандартних циклів аудиту безпеки. Таким чином, виникає потреба у впровадженні безперервного моніторингу моделей, регулярної валідації навчальних вибірок і механізмів контролю якості вихідних результатів.

Окрему увагу слід приділити проблемі алгоритмічної непрозорості (black-box effect) [14; 15]. Багато сучасних моделей глибокого навчання характеризуються складною внутрішньою структурою, що ускладнює інтерпретацію прийнятих рішень. Це створює додаткові ризики для сфер із підвищеними вимогами до відповідальності та аудиту, зокрема у фінансовому секторі, охороні здоров'я та державному управлінні. Інтеграція принципів explainable AI у процеси ризик-менеджменту може суттєво підвищити прозорість та керованість AI-рішень.

Крім технічних аспектів, значну роль відіграють організаційні чинники: розподіл відповідальності за життєвий цикл моделі, визначення політик доступу до навчальних даних, регламентація тестування перед впровадженням у продуктивне середовище. Відсутність чітких процедур управління може нівелювати ефективність навіть найкращих технічних інструментів захисту.

Стратегічний рівень управління передбачає врахування регуляторних вимог, міжнародних стандартів та принципів цифрової етики [25; 28]. З огляду на глобальний характер AI-технологій, доцільним є гармонізоване поєднання стандартів інформаційної безпеки з галузевими регуляторними актами щодо штучного інтелекту. Це дозволить забезпечити баланс між інноваційністю та безпекою.

Окрему роль відіграє сучасне нормативне регулювання, зокрема AI Act Європейського Союзу (2024), який встановлює ризик-орієнтований підхід до класифікації AI-систем та визначає вимоги до їх безпеки.

Важливим доповненням до існуючих підходів є рекомендації OECD щодо управління ризиками штучного інтелекту [18], які передбачають впровадження принципів відповідального AI, зокрема прозорості, пояснюваності, надійності та підзвітності систем. Це дозволяє розширити класичні підходи до кібербезпеки, інтегруючи етичні та соціальні аспекти у процес управління ризиками [3–5].

Таким чином, результати дослідження підтверджують, що ефективне управління кіберризиками в AI-середовищі можливе лише за умови інтеграції технічних, організаційних та стратегічних інструментів, а також переходу від статичних моделей оцінювання до динамічних адаптивних підходів.

Висновки. У результаті проведеного аналізу здійснено систематизацію сучасних методик

управління ризиками у сфері кібербезпеки з урахуванням специфіки використання технологій штучного інтелекту.

Основні результати:

-Запропоновано класифікацію методик за типом оцінювання, рівнем управлінської інтеграції та адаптивністю до AI-ризиків;

-визначено сильні сторони та обмеження стандартизованих, сценарних і кількісних підходів;

-обґрунтовано доцільність комбінованого застосування різних типів методик для забезпечення комплексності оцінювання;

-розроблено адаптивну інтегровану модель управління ризиками використання штучного інтелекту в кібербезпеці.

Наукова новизна полягає у вдосконаленні методичного підходу до класифікації та інтеграції ризик-орієнтованих інструментів з урахуванням специфічних загроз штучного інтелекту, що дозволяє підвищити рівень обґрунтованості управлінських рішень у сфері кібербезпеки.

Практичне значення одержаних результатів полягає у можливості застосування запропонованої моделі під час розроблення політик інформаційної безпеки, модернізації систем управління інформаційною безпекою та впровадження процедур оцінювання AI-ризиків у державному та корпоративному секторах.

Інтеграція класичних та AI-орієнтованих методик дозволяє підвищити ефективність управління ризиками. Запропонована модель забезпечує адаптивність до сучасних кіберзагроз.

Подальші дослідження доцільно спрямувати на емпіричну апробацію запропонованої моделі та розроблення формалізованих метрик оцінювання ризиків інтелектуальних систем.

Список літератури:

1. Потій О. В., Горбенко Ю. І., Замула О. А., Ісірова К. В. Аналіз методів оцінки і управління ризиками кібер- і інформаційної безпеки. *Моделі, методи та засоби захисту інформації в інформаційно-комунікаційних системах : зб. наук. праць. Київ: ДУІКТ,* 2021. С. 5-24. https://doi.org/10.30837/rt.2021.3.206.01. URL: https://duikt.edu.ua/uploads/l_1066_72351971.pdf
2. Єфіменко І. В. Інформаційна безпека держави в умовах воєнного стану: особливості забезпечення. *Успіхи і досягнення у науці*. 2025. № 10(20). С. 96-105. https://doi.org/10.52058/3041-1254-2025-10(20)-96-105
3. Права людини в епоху штучного інтелекту : посібник. *Офіс Омбудсмана України*. Київ, 2023. 41 с. URL: https://ombudsman.gov.ua
4. Про схвалення Концепції розвитку штучного інтелекту в Україні : розпорядження Кабінету Міністрів України від 02.12.2020 № 1556-р. Верховна Рада України. URL: https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text
5. Штучний інтелект у системі кібербезпеки. *Cybersecurity: Education, Science, Technique*. 2023. URL: https://knu.ua
6. Information security, cybersecurity and privacy protection – Guidance on managing information security risks : ISO/IEC 27005:2022. Geneva : International Organization for Standardization, 2022. URL: https://www.iso.org/standard/80585.html
7. Information technology – Security techniques – Information security management systems – Requirements : ISO/IEC 27001:2015. Geneva : International Organization for Standardization, 2015. URL: https://www.iso.org
8. Information technology – Artificial intelligence – Guidance on risk management : ISO/IEC 23894:2023. Geneva : International Organization for Standardization, 2023. URL: https://www.iso.org/standard/77304.html
9. Information technology – Artificial intelligence – Management system : ISO/IEC 42001:2023. Geneva : International Organization for Standardization, 2023. URL: https://www.iso.org
10. Risk Management Framework for Information Systems and Organizations : NIST Special Publication 800-37 Rev. 2. Gaithersburg : National Institute of Standards and Technology, 2018. https://doi.org/10.6028/NIST.SP.800-37r2. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf
11. Guide for Conducting Risk Assessments : NIST Special Publication 800-30 Rev. 1. Gaithersburg : National Institute of Standards and Technology, 2012. https://doi.org/10.6028/NIST.SP.800-30r1. URL: https://csrc.nist.gov/pubs/sp/800/30/r1/final
12. Artificial Intelligence Risk Management Framework (AI RMF 1.0) : NIST AI 100-1. Gaithersburg : National Institute of Standards and Technology, 2023. https://doi.org/10.6028/NIST.AI.100-1. URL: https://www.nist.gov/itl/ai-risk-management-framework
13. COBIT 2019 Framework: Introduction and Methodology. Schaumburg : ISACA, 2020. URL: https://www.isaca.org
14. EBIOS Risk Manager: The Method. Paris : ANSSI, 2018. URL: https://cyber.gouv.fr/securisation/analyse-des-risques/methode-ebios-rm/
15. FAIR: A Framework for Revolutionizing Your Risk Analysis. Center for Internet Security, 2025. URL: https://www.cisecurity.org/insights/blog/fair-a-framework-for-revolutionizing-your-risk-analysis

16. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act). Official Journal of the European Union. 2024. URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

17. ENISA Threat Landscape for Artificial Intelligence. European Union Agency for Cybersecurity, 2023. URL: https://www.enisa.europa.eu

18. Artificial Intelligence Risk Management Framework (AI RMF 1.0). OECD.AI Policy Observatory, 2023. URL: https://oecd.ai/en/catalogue/tools/artificial-intelligence-risk-management-framework-ai-rmf-10

19. Bani-Meqdad M. A. M., Senyk P., Udod M., Pylypenko T., Sylkin O. Cyber-environment in the human rights system: modern challenges to protect intellectual property law and ensure sustainable development of the region. International Journal of Sustainable Development and Planning. 2024. Vol. 19, No. 4. P. 1389–1396. DOI: https://doi.org/10.18280/ijsdp.190416

20. Kordy B. et al. Efficient Algorithms for Attack Trees. // Lecture Notes in Computer Science. 2006. URL: https://doi.org

21. Mauw S., Oostdijk M. Foundations of Attack Trees. Information Security and Cryptology – ICISC 2005. Berlin ; Heidelberg : Springer, 2006. P. 186-198. https://doi.org/10.1007/11734727_17

22. Schneier B. Attack Trees. // Dr. Dobb's Journal. 1999. URL: https://www.schneier.com/academic/archives/1999/12/attack_trees.html

23. Nazarkevych M., Nazarkevych H., Karovic V. Ateb-Gabor Filtering Method in Fingerprint Recognition. // Procedia Computer Science. 2019. Vol. 160. P. 30-37. https://doi.org/10.1016/j.procs.2019.09.440

24. Maksymiv O., Rak T., Peleshko D. Video-based flame detection using LBP-based descriptor: influences of classifiers variety on detection efficiency. // International Journal of Intelligent Systems and Applications. 2017. Vol. 9, No. 2. P. 42-48. https://doi.org/10.5815/ijisa.2017.02.06

25. Shokri R. et al. Adversarial Machine Learning: Security Risks and Methods. 2023. URL: https://arxiv.org

26. Biggio B., Roli F. Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning. Pattern Recognition. 2018. Vol. 84. P. 317-331. https://doi.org/10.1016/j.patcog.2018.07.023

27. MITRE ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems. MITRE, 2023. URL: https://atlas.mitre.org

28. STRIDE Threat Modeling. Microsoft, 2020. URL: https://www.microsoft.com

29. Risk-Based Vulnerability Management. Gartner, 2022. URL: https://www.gartner.com

30. Hubbard D. W., Seiersen R. // How to Measure Anything in Cybersecurity Risk. Hoboken: Wiley, 2016. DOI: https://doi.org/10.1002/9781119162315. URL: https://www.wiley.com

References:

1. Potii O. V., Horbenko Yu. I., Zamula O. A., Isirova K. V. Analiz metodiv otsinky i upravlinnia ryzykamy kiber- i informatsiinoi bezpeky [Analysis of methods for assessing and managing cyber and information security risks]. *Modeli, metody ta zasoby zakhystu informatsii v informatsiino-komunikatsiinykh systemakh : zbirnyk naukovykh prats*. Kyiv: DUIKT, 2021. S. 5–24. DOI:https://doi.org/10.30837/rt.2021.3.206.01URL: https://duikt.edu.ua/uploads/l_1066_72351971.pdf [in Ukrainian].

2. Yefimenko I. V. Informatsiina bezpeka derzhavy v umovakh voiennoho stanu: osoblyvosti zabezpechennia. [Information security of the state under martial law: Features of ensuring]. *Uspikhy i dosiahnennia u nautsi.* 2025. No. 10(20). S. 96–105. DOI: https://doi.org/10.52058/3041-1254-2025-10(20)-96-105 [in Ukrainian].

3. Prava liudyny v epokhu shtuchnoho intelektu : posibnyk. *Ofis Ombudsmana Ukrainy*. Kyiv, 2023. URL: https://ombudsman.gov.ua [in Ukrainian].

4. Pro skhvalennia Kontseptsii rozvytku shtuchnoho intelektu v Ukraini : rozporiadzhennia Kabinetu Ministriv Ukrainy vid 02.12.2020 No. 1556-r. Verkhovna Rada Ukrainy. URL: https://zakon.rada.gov.ua/laws/show/1556-2020-%D1%80#Text [in Ukrainian].

5. Shtuchnyi intelekt u systemi kiberbezpeky. Cybersecurity: Education, Science, Technique. 2023. URL: https://knu.uaInternational Organization for Standardization. (2022). *Information security, cybersecurity and privacy protection – Guidance on managing information security risks* (ISO/IEC 27005:2022). https://www.iso.org/standard/80585.html [in Ukrainian].

6. International Organization for Standardization. (2022). *Information security, cybersecurity and privacy protection – Guidance on managing information security risks* **(**ISO/IEC 27005:2022). https://www.iso.org/standard/80585.html

7. International Organization for Standardization. (2015). *Information technology – Security techniques – Information security*

management systems – Requirements (ISO/IEC 27001:2015). https://www.iso.org

8. International Organization for Standardization. (2023). *Information technology – Artificial intelligence – Guidance on risk management* (ISO/IEC 23894:2023). https://www.iso.org/standard/77304.html

9. International Organization for Standardization. (2023). *Information technology – Artificial intelligence – Management system* (ISO/IEC 42001:2023). https://www.iso.org

10.National Institute of Standards and Technology. (2018). *Risk Management Framework for Information Systems and Organizations* (NIST Special Publication 800-37 Rev. 2). https://doi.org/10.6028/NIST.SP.800-37r2

11. National Institute of Standards and Technology. (2012). *Guide for conducting risk assessments* (NIST Special Publication 800-30 Rev. 1). https://doi.org/10.6028/NIST.SP.800-30r1

12. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

13. ISACA. (2020). *COBIT 2019 framework: Introduction and methodology*. https://www.isaca.org

14. ANSSI. (2018). *EBIOS Risk Manager: The method*. https://cyber.gouv.fr/securisation/analyse-des-risques/methode-ebios-rm/

15. Center for Internet Security. (2025). *FAIR: A framework for revolutionizing your risk analysis*. https://www.cisecurity.org/insights/blog/fair-a-framework-for-revolutionizing-your-risk-analysis

16. European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act). Official Journal of the European Union. *https://eur-lex.europa.eu/eli/reg/2024/1689/oj*

17. European Union Agency for Cybersecurity. (2023). *ENISA threat landscape for artificial intelligence*. https://www.enisa.europa.eu

18. Organisation for Economic Co-operation and Development. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. OECD.AI Policy Observatory. https://oecd.ai/en/catalogue/tools/artificial-intelligence-risk-management-framework-ai-rmf-10

19. Bani-Meqdad, M. A. M., Senyk, P., Udod, M., Pylypenko, T., & Sylkin, O. (2024). Cyber-environment in the human rights system: Modern challenges to protect intellectual property law and ensure sustainable development of the region. *International Journal of Sustainable Development and Planning, 19*(4), 1389–1396. https://doi.org/10.18280/ijsdp.190416

20. Kordy, B., et al. (2006). *Efficient algorithms for attack trees*. https://doi.org

21. Mauw, S., & Oostdijk, M. (2006). Foundations of attack trees. In D. H. Won & S. Kim (Eds.), *Information Security and Cryptology – ICISC 2005* (pp. 186-198). Springer. https://doi.org/10.1007/11734727_17

22. Schneier, B. (1999). Attack trees. *Dr. Dobb's Journal*. https://www.schneier.com/academic/archives/1999/12/attack_trees.html

23. Nazarkevych, M., Nazarkevych, H., & Karovic, V. (2019). Ateb-Gabor filtering method in fingerprint recognition. *Procedia Computer Science, 160*, 30-37. https://doi.org/10.1016/j.procs.2019.09.440

24. Maksymiv, O., Rak, T., & Peleshko, D. (2017). Video-based flame detection using LBP-based descriptor: Influences of classifiers variety on detection efficiency. *International Journal of Intelligent Systems and Applications, 9*(2), 42-48. https://doi.org/10.5815/ijisa.2017.02.06

25. Shokri, R., et al. (2023). *Adversarial machine learning: Security risks and methods*. arXiv. https://arxiv.org

26. Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition, 84*, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

27. MITRE. (2023). *MITRE ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems*. https://atlas.mitre.org

28. Microsoft. (2020). *STRIDE threat modeling*. https://www.microsoft.com

29. Gartner. (2022). *Risk-based vulnerability management*. https://www.gartner.com

30. Hubbard, D. W., & Seiersen, R. (2016). *How to measure anything in cybersecurity risk*. Wiley. https://doi.org/10.1002/9781119162315

Оглядова стаття.
Надійшла до редакції 26.03.2026.
Прийнята до друку 29.04.2026.
Опублікована 25.05.2026.