КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.29.872
УДК 004.056.55:004.8

Ободяк Віктор
к.т.н., доцент, доцент кафедри кібербезпеки
Сумський державний університет, Суми, Україна
ORCID ID: 0000-0002-8539-1252
v.obodyak@cs.sumdu.edu.ua

Отрощенко Михайло
аспірант кафедри комп'ютерних наук
Сумський державний університет, Суми, Україна
ORCID ID: 0000-0001-5064-6780
m.otroshenko@ias.sumdu.edu.ua

Любчак Володимир
к.ф.-м.н., доцент, завідувач кафедри кібербезпеки
Сумський державний університет, Суми, Україна
ORCID ID: 0000-0002-7335-6716
v.liubchak@dcs.sumdu.edu.ua

МОЖЛИВОСТІ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АУДИТУ ТА УПРАВЛІННЯ РИЗИКАМИ КІБЕРБЕЗПЕКИ

Анотація. У статті досліджуються можливості використання штучного інтелекту (ШІ) в аудиті та управлінні ризиками кібербезпеки в умовах цифрової трансформації. Традиційні підходи до аудиту інформаційної безпеки, засновані на ручному зборі даних і періодичних перевірках, показують не достатню ефективність в сучасних цифрових системах. Вони обмежені в масштабуванні, чутливі до людського фактору та не здатні забезпечити постійний моніторинг. Інтеграція технологій штучного інтелекту дозволяє здійснювати автоматичне виявлення аномалій, проактивну оцінку ризиків, генерацію рекомендацій, а також аналіз великих масивів як структурованих, так і неструктурованих даних (журнали подій, мережевий трафік, текстові звіти тощо). Розглянуто приклади застосування моделей машинного навчання та глибокого навчання в практиках аудиту, включно з використанням рекурентних і згорткових нейронних мереж, алгоритмів кластеризації, а також методів обробки природної мови для виявлення порушень політик безпеки. Важливу роль відіграє розвиток концепції ситуаційної обізнаності в мережі, що дозволяє прогнозувати поведінку системи та потенційні загрози на основі історичних і реальних даних. Окрім технічних досягнень, у дослідженні також розглянуто етичні виклики використання штучного інтелекту в аудиті, зокрема: непрозорість алгоритмів, ризик упередженості, загрози конфіденційності та труднощі з відповідальним делегуванням рішень. Підкреслено важливість розробки пояснюваних моделей штучного інтелекту та запровадження етичних принципів, які регулюють застосування автоматизованих систем прийняття рішень у галузі кібербезпеки. Досліджено що штучний інтелект виступає двосторонньою технологією: його можна застосовувати як для забезпечення захисту, так і для здійснення атак. У роботі наведено приклади реальних інцидентів, у яких зловмисники використовували генеративні моделі для реалізації шахрайських дій. Метою дослідження є виявлення як потенціалу, так і обмежень застосування ШІ в аудиті кібербезпеки, а також формування рекомендацій щодо впровадження технологій ШІ з урахуванням технічних, етичних і нормативних вимог. Автори зазначають що найбільш дієвим підходом визнано гібридну модель аудиту, яка поєднує штучний інтелект із людською експертизою. Це забезпечує вищу точність, оперативність і адаптивність у реагуванні на сучасні кіберзагрози, що є критично важливим для зміцнення кіберстійкості організацій в умовах сучасного цифрового середовища.

Ключові слова: штучний інтелект; аудит кібербезпеки; управління ризиками; виявлення аномалій; машинне навчання.

ВСТУП

У сучасних умовах цифрової трансформації питання кібербезпеки набуває стратегічного значення для державного сектору, приватного бізнесу та об'єктів критичної інфраструктури. Зі зростанням складності інформаційних систем та зловмисних атак класичні підходи до аудиту та управління ризиками дедалі частіше виявляються неефективними. Ці методи є трудомісткими, мають обмежене охоплення, працюють у режимі постфактум і не здатні забезпечити гнучке реагування на нові типи загроз.

На цьому тлі стрімкий розвиток технологій штучного інтелекту (ШІ) відкриває нові можливості для підвищення ефективності аудиту, забезпечення безперервного моніторингу та проактивного управління кіберризиками. Алгоритми машинного навчання, обробка природної мови, а також методи ситуаційної обізнаності дозволяють ШІ-системам автоматично ідентифікувати аномалії, оцінювати рівень ризику та генерувати рекомендації для подальших дій.

Постановка проблеми. Попри високий потенціал, застосування ШІ в сфері кібербезпеки супроводжується низкою викликів: від непрозорості прийняття рішень (ефект «чорної скриньки») до ризиків конфіденційності, упередженості моделей, залежності від якості даних та високих обчислювальних витрат. Також невирішеними залишаються питання інтеграції ШІ в існуючі системи аудиту, забезпечення нормативної відповідності та участі людини в контурі прийняття рішень. Тому постає необхідність комплексного дослідження як технологічних, так і етичних аспектів впровадження ШІ у практику аудиту.

Аналіз останніх досліджень і публікацій. Останні дослідження [**Помилка! Джерело посилання не знайдено.**]–[**Помилка! Джерело посилання не знайдено.**] вказують на успішне застосування методів supervised learning для виявлення відомих загроз, unsupervised learning — для детекції невідомих аномалій, a reinforcement learning — для динамічного управління політиками безпеки. Платформи на кшталт McAfee MVISION вже реалізують автоматизований аналіз мережевого трафіку та генерують звіти про вразливості у реальному часі. У вітчизняних публікаціях [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**] також наголошується на важливості впровадження ситуаційної обізнаності в мережі (network situation awareness), що дозволяє передбачати інциденти до їхнього прояву на основі потокового аналізу поведінки користувачів.

Разом із цим підкреслюються етичні та соціальні ризики, що виникають у процесі прийняття рішень на основі непрозорих моделей [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**], зокрема проблема explainability (пояснюваності) та fairness (відсутності упередженості), що наразі є ключовими у формуванні довіри до ШІ-рішень.

Мета статті. Метою даної роботи є дослідження можливостей, переваг та обмежень використання штучного інтелекту для підвищення ефективності аудиту та управління ризиками кібербезпеки. Зокрема, увага зосереджується на практичних аспектах реалізації інтелектуальних систем моніторингу, аналізі методів прийняття рішень у реальному часі, забезпеченні прозорості ШІ-рішень та дотриманні етичних стандартів.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Роль ШІ в управлінні ризиками кібербезпеки

Ефективне управління кіберризиками сьогодні вимагає від організацій швидкого виявлення, класифікації та нейтралізації загроз у постійно змінному цифровому середовищі. Штучний інтелект відкриває нові можливості для виконання цих завдань завдяки здатності аналізувати великі обсяги даних, розпізнавати приховані закономірності та розробляти адаптивні стратегії реагування.

У контексті оцінки та управління ризиками кібербезпеки штучний інтелект застосовується через низку навчальних підходів, які демонструють високу ефективність у виявленні як відомих, так і нових загроз. Одним з найбільш поширених методів є навчання з учителем (supervised learning), який використовує марковані набори даних для ідентифікації та класифікації типових векторів атак, таких як фішинг, сканування портів або використання експлойтів. Цей підхід довів свою результативність у задачах виявлення шкідливого програмного забезпечення, що підтверджується результатами досліджень, зокрема у роботах [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**].

На відміну від цього, навчання без учителя (unsupervised learning) не потребує попередньо маркованих даних і дозволяє системам виявляти аномальні шаблони поведінки, які можуть вказувати на нові або приховані загрози, включаючи lateral movement, приховані канали зв’язку та повільно діючі атаки типу APT (Advanced Persistent Threat). Цей підхід широко застосовується в системах моніторингу поведінки користувачів та аналізу мережевих аномалій, як зазначено в роботах [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**].

Третім підходом, що набирає актуальності, є навчання з підкріпленням (reinforcement learning), яке дозволяє системам адаптивно вдосконалювати свої стратегії захисту шляхом безперервної оптимізації політик реагування в умовах змінного кіберсередовища. Така методологія забезпечує динамічну адаптацію до нових векторів атак та є ефективною в задачах управління мережею, що підтверджено в роботі [**Помилка! Джерело посилання не знайдено.**].

Таким чином, різні методи машинного навчання, в залежності від їхньої структури та доступності даних, доповнюють одне одного в процесі створення адаптивної та надійної системи виявлення загроз і підтримки стратегічного управління кіберризиками. Систематизовану інформацію щодо перелічених методів представлено у табл. 1.

Таблиця 1

Методи для аналізу великих обсягів даних у кібербезпеці за допомогою ШІ

	Застосування	**Переваги**	**Недоліки / Виклики**
Навчання з учителем (Supervised learning)	Виявлення фішингу, експлойтів, шкідливого ПЗ	Висока точність для відомих шаблонів загроз	Потребує якісних маркованих даних, низька ефективність щодо невідомих атак

Навчання без учителя (Unsupervised learning)	Виявлення APT, lateral movement, прихованих каналів, поведінковий аналіз користувачів	Може виявити невідомі або нові типи загроз	Високий рівень хибних спрацьовувань, складність інтерпретації результатів
Навчання з підкріпленням (Reinforcement learning)	Адаптивне управління мережею, динамічна оптимізація політик реагування	Гнучка адаптація до нових умов та атак, ефективна в реальному часі	Складна реалізація, потрібні моделі середовища, можливі ризики під час фази навчання

Приклад платформи MVISION від компанії McAfee демонструє практичну результативність застосування штучного інтелекту в галузі аудиту та управління кіберризиками. Завдяки використанню алгоритмів машинного навчання система здійснює безперервний аналіз мережевого трафіку, ідентифікує аномальні дії, адаптується до нових типів загроз і автоматично генерує звіти про вразливості. Вбудована інтеграція з глобальними джерелами розвідки кіберзагроз дозволяє платформі швидко реагувати на нові атаки. Окрім технологічних рішень, компанія приділила увагу навчанню персоналу, сприяючи формуванню сталої культури кібербезпеки в межах організації [**Помилка! Джерело посилання не знайдено.**].

Експерти, опитані в рамках дослідження, підкреслили значні переваги ШІ для аудиту кібербезпеки: підвищення точності, швидкості, масштабованості та зменшення кількості хибнопозитивних спрацювань. Старший консультант однієї з фінансових установ зазначив, що автоматизація на основі ШІ «не тільки покращує глибину аудиту, але й підвищує його стратегічну цінність, дозволяючи виявляти системні проблеми до того, як вони стануть критичними».

Окрім технічних аспектів, застосування штучного інтелекту в управлінні ризиками порушує низку етичних питань. Одним із напрямків розвитку є впровадження методів пояснюваного штучного інтелекту (XAI), що спрямовані на покращення зрозумілості та обґрунтованості рішень, прийнятих системами машинного навчання. Не менш важливим є захист персональних даних, що використовуються для навчання систем і виявлення загроз. Щоб уникнути порушення норм конфіденційності, організації мають впроваджувати чітко регламентовані підходи до збору, обробки й зберігання даних відповідно до стандартів, таких як GDPR, ISO 27001 тощо [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**].

Ітеративна природа алгоритмів штучного інтелекту забезпечує можливість безперервного навчання систем на основі нових даних та результатів аудитів, що

дозволяє перейти до моделі постійного, проактивного аудиту, в якій не просто фіксуються відхилення, а здійснюється оцінка поведінкових змін і моделювання ситуацій. Зокрема, системи розширюють можливості Network Situation Awareness, що дозволяє прогнозувати ризики до їх реалізації, базуючись на інтелектуальному аналізі потоків даних у мережі та поведінку користувачів [**Помилка! Джерело посилання не знайдено.**]. Це створює концептуальні засади для формування адаптивної, аналітичної та проактивної стратегії управління кібербезпекою. Зокрема, відповідно до внутрішніх документів компанії Meta, з якими ознайомилося агентство NPR, планується впровадження системи штучного інтелекту, здатної автоматизовано оцінювати до 90% змін, що вносяться до таких додатків, як Instagram і WhatsApp, з погляду їх впливу на конфіденційність користувачів.

Такий ШІ-орієнтований підхід потенційно підвищує оперативність розгортання оновлень, однак, за оцінками окремих колишніх керівників компанії, він супроводжується зростанням рівня ризику, оскільки негативні побічні ефекти змін можуть залишитися непоміченими до моменту їх фактичного прояву [**Помилка! Джерело посилання не знайдено.**]. У цьому контексті особливої актуальності набуває забезпечення балансу між автоматизацією та участю людини в процесі прийняття рішень. Максимальна ефективність впровадження ШІ в управління кіберризиками досягається саме завдяки синергії між інтелектуальними системами та людською експертизою, що дозволяє здійснювати інтерпретацію складних ситуацій з урахуванням стратегічних, нормативних і етичних чинників.

ШІ в аудиті кібербезпеки

Традиційний аудит кібербезпеки, як правило, базуються на ручному зборі та аналізі даних, що обмежує їх ефективність у масштабованих та динамічних цифрових середовищах. Такі аудити є трудомісткими, залежать від суб'єктивних суджень аудитора, схильні до людських помилок і не здатні забезпечити безперервний моніторинг. Крім того, класичні перевірки часто мають обмежений обсяг охоплення: аналізується лише вибірка активів або подій, що створює ризик пропуску критичних вразливостей.

Інтеграція штучного інтелекту у процеси аудиту кібербезпеки поступово трансформує цю практику з переважно ручної, реактивної діяльності на проактивний, динамічний та високотехнологічний процес. На відміну від традиційного аудиту, який переважно базуються на статичних контрольних списках і періодичних перевірках, підходи на основі ШІ забезпечують постійне, безперервне спостереження за системами, виявлення аномалій у режимі реального часу та глибокий аналіз поведінкових даних. Ці можливості дозволяють не лише виявляти відомі вразливості, а й вчасно розпізнавати нові, раніше невідомі загрози в умовах сучасного кіберпростору [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**].

Завдяки здатності працювати з неструктурованими даними, зокрема текстами журналів подій чи мережевого трафіку, моделі на базі рекурентних і згорткових нейронних мереж дозволяють виявляти складні шаблони, які можуть вказувати на порушення або аномалії. Це значно перевершує можливості ручного аналізу або сигнатурних систем, які часто пропускають нетипові сценарії [**Помилка! Джерело посилання не знайдено.**]. На рис. 1 представлено функціональну схему такої системи, яка може бути адаптована для проведення аудиту кібербезпеки в реальному часі.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023

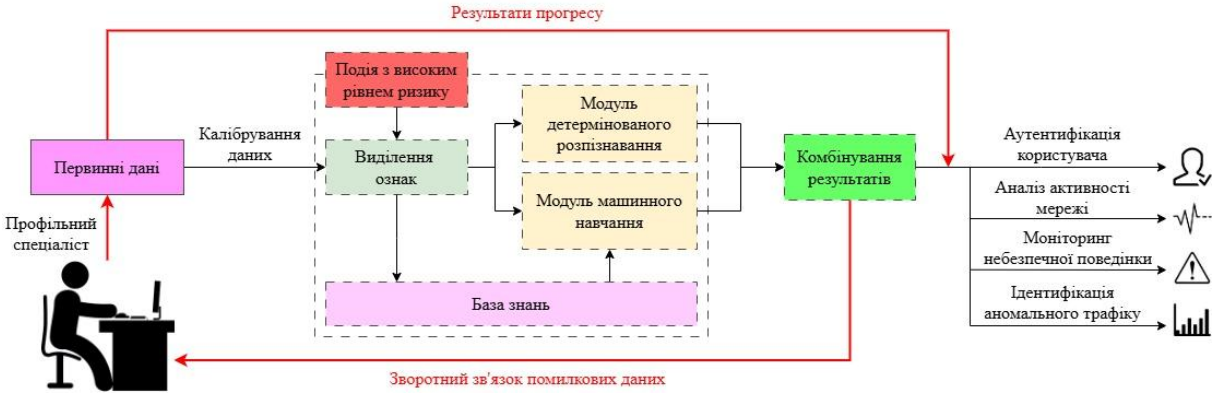

Рис. 1. Функціональна схема інтелектуальної аудиторської системи

Вхідні мережеві дані проходять попередню обробку, калібрування та виділення ключових ознак, після чого модулі розпізнавання, та машинного навчання аналізують ознаки та формують оцінку ризику. Узгодженість їх результатів формує рівень довіри до виявленої події. Події з високим ризиком заносяться до бази знань для подальшого навчання, а їх оцінка використовується для автентифікації, моніторингу поведінки та оцінки ситуаційної обізнаності в мережі. У разі невірного або недостатньо впевненого розпізнавання події система повертає зворотний зв’язок, а також надає можливість ручного втручання, що забезпечує гібридну адаптацію до аудиту мережі.

Особливо цінним є використання обробки природної мови (NLP) для аналізу аудиторських звітів, внутрішньої документації та журналів безпеки. NLP-моделі здатні автоматично витягувати ключові події, порушення політик безпеки або аномальні дії користувачів, що забезпечує аналітикам точніший і контекстуалізований огляд ситуації [12]. Це дозволяє оперативно реагувати на кіберінциденти ще до того, як вони набудуть широкого масштабу.

Разом з тим, використання ШІ в аудиті кібербезпеки породжує низку етичних і регуляторних викликів. Насамперед, це проблема прозорості алгоритмів прийняття рішень: багато моделей, особливо на базі глибокого навчання, є складними для інтерпретації, що ускладнює аудиторам перевірку обґрунтованості висновків. Відповідно, зростає інтерес до напрямку пояснюваного ШІ (XAI), який дозволяє краще зрозуміти, чому система вважає певну поведінку ризикованою або підозрілою [**Помилка! Джерело посилання не знайдено.**].

Додатково постають питання конфіденційності — використання великих масивів операційних даних для навчання або моніторингу вимагає суворого дотримання нормативних стандартів у сфері захисту персональної інформації. Розв’язання цих питань передбачає не лише технічну реалізацію (наприклад, через анонімізацію або федеративне навчання), а й запровадження політик етичного використання ШІ [**Помилка! Джерело посилання не знайдено.**].

Нарешті, хоча автоматизація здатна значно покращити точність і швидкість аудиту, людський фактор залишається критично важливим. Лише фахівець може інтерпретувати складні ризикові сценарії, узгодити технічні висновки із бізнес-контекстом і забезпечити прийняття зважених управлінських рішень. Таким чином, майбутнє кібераудиту, ймовірно, полягатиме у гібридній моделі співпраці між інтелектуальними системами та людьми, яка поєднує масштабованість ШІ з глибиною людського аналізу.

Етичні та технічні виклики

Попри безсумнівний потенціал штучного інтелекту у сфері кібербезпеки, його впровадження супроводжується низкою суттєвих етичних, соціальних та технічних викликів. Ці проблеми набувають підвищеної актуальності у сфері аудиту та управління ризиками, де штучний інтелект не тільки ідентифікує загрози, а й обробляє конфіденційні дані, впливає на прийняття рішень щодо безпеки та може проявляти упередженість у процесах навчання і функціонування моделей.

Одним із головних викликів є відсутність прозорості та пояснюваності моделей, особливо тих, що базуються на глибокому навчанні. Так званий ефект «чорної скриньки» ускладнює інтерпретацію рішень, які приймаються моделями ШІ, знижуючи рівень довіри з боку аудиторів, регуляторів і користувачів [**Помилка! Джерело посилання не знайдено.**]. Така непрозорість може унеможливити оцінку легітимності дій у випадках, коли система штучного інтелекту рекомендує блокування користувача чи ізоляцію системи. Одним із перспективних шляхів вирішення цієї проблеми є розвиток пояснюваного штучного інтелекту (Explainable AI, XAI), який дозволяє отримувати інтерпретовані висновки без суттєвої втрати точності.

Іншим важливим етичним аспектом є ризики конфіденційності, пов’язані з обробкою великих масивів персональних або корпоративних даних. Аудиторські системи на базі ШІ можуть неусвідомлено або свідомо накопичувати та аналізувати чутливу інформацію, створюючи нові вектори для витоків або несанкціонованого доступу. Це ставить організації перед викликом дотримання вимог щодо захисту даних, навіть у тих країнах, де відсутнє жорстке регулювання на кшталт GDPR. Серед технічних рішень, що знижують ризик порушення конфіденційності, виділяються федеративне навчання та диференційна конфіденційність, які дозволяють навчати моделі на розподілених даних без централізованого збору [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**].

Технічною загрозою є також можливість використання ШІ зловмисниками. Як зазначається в дослідженнях [**Помилка! Джерело посилання не знайдено.**], [**Помилка! Джерело посилання не знайдено.**], генеративні моделі можуть застосовуватися для створення шкідливого коду, автоматизованого фішингу, deepfake-атак і навіть для обходу систем захисту. Зокрема доступ до LLM-моделей дозволяє кіберзлочинцям масово генерувати персоналізовані фішингові повідомлення або підробляти біометричні дані. Один із задокументованих випадків — атака 2020 року, під час якої зловмисники згенерували голос керівника компанії для шахрайського переказу коштів (Symantec, 2020). Згідно зі статистикою IBM X-Force, близько 30% кіберінцидентів вже пов’язані з використанням інструментів ШІ [**Помилка! Джерело посилання не знайдено.**]. Це підкреслює необхідність розробки захисних рішень, які самі ґрунтуються на ШІ, а також важливість етичного контролю й регуляторного нагляду за його використанням.

Крім того постає проблема упередженості даних і моделей, яка може виникати як на етапі підготовки тренувального датасету, так і під час оновлення моделі у процесі експлуатації. У результаті моделі можуть демонструвати дискримінаційні або неадекватні рішення, особливо в ситуаціях, що відхиляються від «нормального» патерну даних. Як зазначалось в роботі [**Помилка! Джерело посилання не знайдено.**], для подолання цієї проблеми необхідно розробляти чіткі процедури перевірки якості даних, формувати мультиконтекстуальні тестові набори та здійснювати регулярну перевалідацію моделей.

Серед можливих організаційних рішень проблеми — впровадження етичних настанов, які б регламентували допустиме використання ШІ в аудиті та безпеці. На рівні міждержавної координації, такі ініціативи вже були запропоновані OECD (2021) та IEEE (2019), де

наголошується на принципах прозорості, нагляду людини, технічної надійності та недискримінаційності. Деякі компанії вже реалізують внутрішні етичні комітети з розгляду рішень на основі ШІ, включаючи аудити моделей перед їхнім впровадженням.

Таким чином, хоча впровадження штучного інтелекту у сферу кібербезпеки й відкриває значні можливості для автоматизації, масштабування та ефективного реагування, воно супроводжується комплексом викликів, що вимагають міждисциплінарного підходу до вирішення. Ефективна імплементація ШІ передбачає не лише вдосконалення алгоритмів, а й розвиток нормативної, етичної та організаційної інфраструктури, що забезпечує відповідальне використання цих технологій у критично сферах.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Подальший розвиток ШІ буде базуватись на створенні гібридних систем, у яких автоматизовані рішення ШІ доповнюються людською оцінкою, особливо у випадках з високою вартістю помилки. Даний підхід дозволить досягти більш надійного та етичного результату, знижуючи ризик прийняття непрозорих або небажаних рішень.

Одним із векторів розвитоку є розробка пояснюваного штучного інтелекту (XAI). У контексті аудитів кібербезпеки це має вирішальне значення, оскільки дозволяє операторам та аудиторам розуміти причини спрацювання моделей, перевіряти логіку рішень та виявляти потенційні збої або упередження. Наприклад, у випадках, коли система блокує підозрілий трафік або ідентифікує внутрішнього зловмисника, необхідно чітко пояснити, які саме ознаки були критичними для класифікації.

Іншим перспективним напрямом є використання федеративного навчання (federated learning), яке дозволяє навчати моделі без передачі самих даних до централізованих серверів. У сфері кібербезпеки це має особливе значення, адже дані журналів подій, телеметрії чи користувацької активності часто є конфіденційними. Федеративне навчання зменшує ризики витоку інформації та дозволяє об’єднувати знання з різних джерел без порушення нормативних вимог, таких як GDPR. Нещодавні дослідження Google Research демонструють ефективність цього підходу в задачах виявлення аномалій у розподілених середовищах.

З наближенням епохи практичного квантового обчислення зростає потреба в алгоритмах, здатних функціонувати в умовах постквантового криптографічного середовища. Це стосується не лише захисту даних, які обробляє ШІ, а й моделювання ризиків, пов’язаних із використанням квантових технологій для злому систем. Сучасні ініціативи, зокрема NIST Post-Quantum Cryptography Project, вже окреслили набір алгоритмів, які мають бути інтегровані в системи безпеки, включно з тими, що базуються на ШІ.

Також важливою тенденцією є побудова гібридних систем, які поєднують переваги автоматизованого аналізу з людською експертизою. У таких системах ШІ відповідає за рутинні перевірки, первинну фільтрацію інцидентів та формування звітів, у той час як спеціалісти з кібербезпеки здійснюють стратегічний аналіз, інтерпретують складні кейси та ухвалюють фінальні рішення. Такий підхід забезпечує баланс між швидкістю машинного аналізу та контекстною гнучкістю людського мислення.

Додатково перспективними виглядають напрямки розробки адаптивних моделей на основі онлайн-навчання, які можуть оновлювати свої знання в режимі реального часу, реагуючи на зміну векторів атак. Це особливо актуально для Zero-Day атак, які неможливо передбачити заздалегідь. Такі підходи вже тестуються у дослідженнях DARPA та MITRE, з метою інтеграції в системи оперативного реагування.

Не менш важливо також формування етико-правової інфраструктури для впровадження ШІ в кібербезпеку. Поряд із технічними інноваціями, необхідні чіткі протоколи валідації моделей, методи контролю прийняття рішень, а також незалежні органи оцінювання ризиків, що дозволить уникнути зловживань і зберегти довіру до технологій в суспільстві.

Загалом, майбутні дослідження мають бути міждисциплінарними: технічні рішення мають супроводжуватись етичними, юридичними та соціальними оцінками, що дозволить створити стійку й прозору екосистему застосування ШІ в кібербезпеці.

Інтеграція штучного інтелекту в аудити та управління кіберризиками відкриває нові горизонти для ефективного, масштабованого та проактивного забезпечення кібербезпеки. Хоча технологія ще стикається з етичними та технічними обмеженнями, вона вже сьогодні демонструє значні переваги в автоматизації аналізу, швидкому реагуванні та підтримці прийняття стратегічних рішень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Vasan, D., Alazab, M., Wassan, S., et al. (2020). IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture. *Computer Networks, 171*, Article 107138. https://doi.org/10.1016/j.comnet.2020.107138
2. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502
3. Authorea. (n.d.). *AI-driven cyber risk assessment: Predicting and preventing data breaches with machine learning*. https://www.authorea.com/users/898703/articles/1274531-ai-driven-cyber-risk-assessment-predicting-and-preventing-data-breaches-with-machine-learning
4. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy* (pp. 305–316). IEEE. https://doi.org/10.1109/SP.2010.25
5. https://www.sciencedirect.com/science/article/abs/pii/S016740481930118X?via%3Dihub
6. Anjum, N., & Chowdhury, M. R. (2024). Revolutionizing cybersecurity audit through artificial intelligence automation: A comprehensive exploration. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*.
7. Phanishlakarasu. (n.d.). *AI for cybersecurity audits: Enhancing transparency and accountability*. Medium. https://medium.com/@phanishlakarasu/ai-for-cybersecurity-audits-enhancing-transparency-and-accountability-a4572a59b436
8. Expert.com.ua. (2024). *Meta планує автоматизувати багато оцінок ризиків продуктів*. https://expert.com.ua/200293-meta-planue-avtomatyzuvaty-bahato-ocinok-ryzykiv-produktiv.html
9. Chen, T., Wang, Z., & Zhang, C. (2021). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications, 58*, 102726. https://doi.org/10.1016/j.jisa.2021.102726
10. Roy, A., Dey, N., & Ashour, A. S. (Eds.). (2022). *Cyber security and digital forensics: Challenges and future trends*. Springer Nature. https://doi.org/10.1007/978-981-19-2591-3
11. Xu, W., Wang, L., & Zhao, Y. (2020). Intrusion detection system based on deep belief network and probabilistic neural network. *Neural Computing and Applications, 32*, 11265–11273. https://doi.org/10.1007/s00521-019-04552-2
12. Zhou, Y., & Sharma, A. (2022). A survey of NLP techniques for cybersecurity applications. *ACM Computing Surveys, 55*(3), Article 50. https://doi.org/10.1145/3491200
13. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint*. https://arxiv.org/abs/1702.08608
14. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
15. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* (pp. 1273–1282). PMLR. https://proceedings.mlr.press/v54/mcmahan17a.html

16. Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042
17. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint*. https://arxiv.org/abs/1802.07228
18. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org. https://fairmlbook.org
19. Islam, M. R. (2024). *Generative AI, cybersecurity, and ethics*. Wiley.
20. Petrenko, S. A., & Smirnov, V. I. (2023). Threat models and recommendations for protecting information systems based on AI. In *Proceedings of the Conference on Information Security and Cyber Defense* (pp. 764–770).
21. Lungol, O. M. (2024). Review of methods and strategies of cybersecurity using AI tools. In *Proceedings of the 2nd All-Ukrainian Scientific and Practical Conference "Digital Transformations in the Context of Security Challenges"* (pp. 379–389). Kyiv: National Academy of the Security Service of Ukraine.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (29), 2025
ISSN 2663 - 4023

Viktor Obodiak
Candidate of Technical Sciences, Associate Professor
Associate Professor of the Department of Cybersecurity
Sumy State University, Sumy, Ukraine
ORCID ID: 0000-0002-8539-1252
v.obodyak@cs.sumdu.edu.ua

Mykhailo Otroshchenko
Postgraduate Student, Department of Computer Science
Sumy State University, Sumy, Ukraine
ORCID ID: 0000-0001-5064-6780
m.otroshenko@ias.sumdu.edu.ua

Volodymyr Liubchak
Candidate of Physical and Mathematical Sciences, Associate Professor
Head of the Department of Cybersecurity
Sumy State University, Sumy, Ukraine
ORCID ID: 0000-0002-7335-6716
v.liubchak@dcs.sumdu.edu.ua

OPPORTUNITIES OF ARTIFICIAL INTELLIGENCE FOR CYBERSECURITY AUDIT AND RISK MANAGEMENT

Abstract. This article explores the potential of artificial intelligence (AI) in cybersecurity auditing and risk management within the context of ongoing digital transformation. Traditional approaches to information security auditing—based on manual data collection and periodic assessments—are increasingly insufficient for dynamic and large-scale digital ecosystems. They are limited in scalability, prone to human error, and lack the capacity for continuous monitoring. The integration of AI technologies allows for automated anomaly detection, proactive risk assessment, real-time decision support, and analysis of vast volumes of both structured and unstructured data, including event logs, network traffic, and audit reports. The study examines the application of machine learning and deep learning models in audit practices, including recurrent and convolutional neural networks, clustering algorithms, and natural language processing (NLP) techniques for detecting security policy violations. Particular attention is given to the concept of Network Situation Awareness, which enables the prediction of system behavior and potential threats based on historical and real-time behavioral data. In addition to technical achievements, the research addresses the ethical challenges associated with AI deployment in audits: algorithmic opacity, bias risks, privacy concerns, and difficulties in delegating decision-making to automated systems. The need for explainable AI (XAI) and the development of ethical guidelines for responsible AI use in cybersecurity audits is emphasized. AI is highlighted as a dual-use technology—capable of both defending against and facilitating cyberattacks. The article refers to real-world incidents, such as the use of generative models in social engineering and voice-based fraud. The aim of the study is to identify both the benefits and limitations of AI-powered cybersecurity auditing and to provide recommendations for the ethical and effective implementation of intelligent systems. The paper concludes that a hybrid model—combining AI automation with human expertise—is the most promising strategy for enhancing the accuracy, efficiency, and adaptability of cybersecurity risk assessment. This integrated approach is essential to improving cyber resilience in today's volatile digital environment.

Keywords: artificial intelligence; cybersecurity audit; risk management; anomaly detection; machine learning.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Vasan, D., Alazab, M., Wassan, S., et al. (2020). IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture. *Computer Networks, 171*, Article 107138. https://doi.org/10.1016/j.comnet.2020.107138

2. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502
3. Authorea. (n.d.). *AI-driven cyber risk assessment: Predicting and preventing data breaches with machine learning*. https://www.authorea.com/users/898703/articles/1274531-ai-driven-cyber-risk-assessment-predicting-and-preventing-data-breaches-with-machine-learning
4. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy* (pp. 305–316). IEEE. https://doi.org/10.1109/SP.2010.25
5. https://www.sciencedirect.com/science/article/abs/pii/S016740481930118X?via%3Dihub
6. Anjum, N., & Chowdhury, M. R. (2024). Revolutionizing cybersecurity audit through artificial intelligence automation: A comprehensive exploration. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*.
7. Phanishlakarasu. (n.d.). *AI for cybersecurity audits: Enhancing transparency and accountability*. Medium. https://medium.com/@phanishlakarasu/ai-for-cybersecurity-audits-enhancing-transparency-and-accountability-a4572a59b436
8. Expert.com.ua. (2024). *Meta планує автоматизувати багато оцінок ризиків продуктів*. https://expert.com.ua/200293-meta-planue-avtomatyzuvaty-bahato-ocinok-ryzykiv-produktiv.html
9. Chen, T., Wang, Z., & Zhang, C. (2021). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications, 58*, 102726. https://doi.org/10.1016/j.jisa.2021.102726
10. Roy, A., Dey, N., & Ashour, A. S. (Eds.). (2022). *Cyber security and digital forensics: Challenges and future trends*. Springer Nature. https://doi.org/10.1007/978-981-19-2591-3
11. Xu, W., Wang, L., & Zhao, Y. (2020). Intrusion detection system based on deep belief network and probabilistic neural network. *Neural Computing and Applications, 32*, 11265–11273. https://doi.org/10.1007/s00521-019-04552-2
12. Zhou, Y., & Sharma, A. (2022). A survey of NLP techniques for cybersecurity applications. *ACM Computing Surveys, 55*(3), Article 50. https://doi.org/10.1145/3491200
13. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint*. https://arxiv.org/abs/1702.08608
14. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2
15. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics* (pp. 1273–1282). PMLR. https://proceedings.mlr.press/v54/mcmahan17a.html
16. Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042
17. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint*. https://arxiv.org/abs/1802.07228
18. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org. https://fairmlbook.org
19. Islam, M. R. (2024). *Generative AI, cybersecurity, and ethics*. Wiley.
20. Petrenko, S. A., & Smirnov, V. I. (2023). Threat models and recommendations for protecting information systems based on AI. In *Proceedings of the Conference on Information Security and Cyber Defense* (pp. 764–770).
21. Lungol, O. M. (2024). Review of methods and strategies of cybersecurity using AI tools. In *Proceedings of the 2nd All-Ukrainian Scientific and Practical Conference "Digital Transformations in the Context of Security Challenges"* (pp. 379–389). Kyiv: National Academy of the Security Service of Ukraine.