

УДК 004.056.55+004.8

[https://doi.org/10.52058/2786-6025-2025-5\(46\)-1934-1948](https://doi.org/10.52058/2786-6025-2025-5(46)-1934-1948)

Переяславська Світлана Олександрівна кандидат педагогічних наук, доцент, доцент кафедри інформаційних технологій та систем, Державний заклад «Луганський національний університет імені Тараса Шевченка», <https://orcid.org/0000-0001-9873-0447>

Майданевич Леонід Олександрович кандидат філософських наук, старший викладач, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, <https://orcid.org/0000-0002-7364-8874>

Редич Олександр Володимирович кандидат економічних наук, доцент кафедри комп'ютерних та інформаційних технологій і систем, Державний податковий університет, <https://orcid.org/0000-0002-6183-5517>

МЕТОДИ ВИЯВЛЕННЯ ТА ЗАПОБІГАННЯ КІБЕРАТАКАМ НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. Дослідження спрямоване на вирішення таких завдань: систематизація ключових методів виявлення кібератак з використанням технологій ШІ, аналіз ефективності різних алгоритмів машинного навчання для виявлення та класифікації кіберзагроз, дослідження підходів до автоматизованого реагування на інциденти на основі ШІ та розробка рекомендацій щодо впровадження систем кібербезпеки на основі ШІ в різних організаціях.

Мета статті. Метою статті є дослідження сучасних методів виявлення та запобігання кібератакам на основі штучного інтелекту (ШІ), систематизація підходів до інтеграції ШІ в системи кібербезпеки, аналіз їх ефективності та розробка рекомендацій щодо практичного впровадження.

Наукова новизна. Наукова новизна полягає у комплексному аналізі методів ШІ для кібербезпеки, включаючи виявлення аномалій за допомогою автокодерів, класифікацію за допомогою нейронних мереж, гібридних ансамблевих методів, генеративних моделей (GAN, LSTM, Transformer) та методів глибокого навчання (CNN, RNN, Deep Reinforcement Learning). Дослідження представляє практичну реалізацію системи виявлення аномалій на основі автокодера в Python, підтверджену візуалізаціями реальних даних (матриця плутанини, розподіл помилок реконструкції та динаміка функції втрат). Вона також висвітлює інтеграцію полегшених моделей ШІ для

периферійних мереж і пропонує стратегії для покращення інтерпретованості моделі та стійкості до ворожих атак, заповнюючи прогалини в існуючих дослідженнях етичних та обчислювальних викликів.

Результати. Дослідження класифікує методи виявлення кібератак на основі штучного інтелекту за категоріями, оцінюючи їхні принципи, переваги та обмеження. Запропонована архітектура системи включає збір даних, попередню обробку, виявлення аномалій, класифікацію загроз, прогнозування атак, автоматизоване реагування та аналітику, продемонстровані за допомогою оригінальних діаграм. Реалізація автокодера досягла високої точності: 0 помилкових спрацьовувань і 97,8% виявлення загроз, що перевершує традиційні системи, зменшуючи кількість помилкових спрацьовувань на 45%. Стало зрозуміло, що генеративні моделі мають перевагу у виявленні нових атак, а ансамблі підвищують точність виявлення. Основними проблемами, які обговорюються, є інтерпретованість моделей, вразливість до атак і потреба в ресурсах, і, як наслідок, для їх вирішення пропонується чіткий штучний інтелект і прості алгоритми.

Висновки. Кібербезпека значно покращується, коли використовуються методи штучного інтелекту, а глибоке навчання та генеративні моделі набагато краще справляються з виявленням нових, невідомих атак, ніж попередні алгоритми. Завдяки впровадженню ансамблю точність підвищується, а помилково ідентифікованих випадків стає менше. Крім того, єдина архітектура системи дозволяє їй змінюватися і виконувати завдання самостійно. Для збереження довіри аудиторії дуже важливо, щоб модель була зрозумілою і її було важко обдурити. Полегшені алгоритми компенсують складні обчислювальні потреби. Майбутні дослідження мають зосередитися на тому, щоб зробити ШІ простішим для розуміння, захистити моделі та об'єднати їх з іншими інструментами кібербезпеки.

Ключові слова: кібербезпека, виявлення загроз, запобігання атакам, нейронні мережі, кіберзагрози.

Pereiaslavska Svitlana Candidate of Pedagogical Sciences, Associate Professor, Associate Professor of the Department of Information Technologies and Systems, Luhansk Taras Shevchenko National University, <https://orcid.org/0000-0001-9873-0447>

Maidanevych Leonid PhD in Philosophical Sciences, Senior Lecturer, faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, <https://orcid.org/0000-0002-7364-8874>

Redych Oleksandr PhD in Economics, Associate Professor of the Department of Computer and Information Technologies and Systems, State Tax University, <https://orcid.org/0000-0002-6183-5517>

METHODS OF DETECTING AND PREVENTING CYBERATTACKS BASED ON ARTIFICIAL INTELLIGENCE

Abstract. The study is aimed at solving the following tasks: systematization of key methods for detecting cyberattacks using AI technologies, analysis of the effectiveness of various machine learning algorithms for detecting and classifying cyber threats, study of approaches to automated response to AI-based incidents, and development of recommendations for implementing AI-based cybersecurity systems in various organizations.

The purpose of the article. The purpose of the article is to study modern methods of detecting and preventing cyberattacks based on artificial intelligence (AI), to systematize approaches to integrating AI into cybersecurity systems, to analyze their effectiveness and to develop recommendations for practical implementation.

Scientific novelty. The scientific novelty lies in the comprehensive analysis of AI methods for cybersecurity, including anomaly detection using autoencoders, classification using neural networks, hybrid ensemble methods, generative models (GAN, LSTM, Transformer) and deep learning methods (CNN, RNN, Deep Reinforcement Learning). The research presents a practical implementation of an autoencoder-based anomaly detection system in Python, supported by visualizations of real data (confusion matrix, reconstruction error distribution, and loss function dynamics). It also highlights the integration of lightweight AI models for edge networks and suggests strategies to improve model interpretability and resilience to hostile attacks, filling gaps in existing research on ethical and computational challenges.

Results. The study classifies AI-based cyberattack detection methods into categories, evaluating their principles, advantages, and limitations. The proposed system architecture includes data collection, preprocessing, anomaly detection, threat classification, attack prediction, automated response, and analytics, demonstrated through original diagrams. The implementation of the auto-encoder has achieved high accuracy: 0 false positives and 97.8% threat detection, which outperforms traditional systems by reducing false positives by 45%. It became clear that generative models have an edge in detecting new attacks, and ensembles enhance the accuracy of the detection. The main issues discussed are model interpretability, vulnerability to attacks, and the need for resources, and as a result, clear artificial intelligence and simple algorithms are suggested to solve them.

Conclusions. Cybersecurity is greatly improved when artificial intelligence methods are used, and deep learning and generative models do a much better job than earlier algorithms in detecting new, unknown attacks. By implementing an ensemble, accuracy increases and there are fewer mistakenly identified cases. Also, the system's single architecture allows it to change and carry out tasks on its own. It

is very important that a model is interpretable and difficult to fool to maintain audience trust. Lighter algorithms make up for the tough computational needs. Future studies should concentrate on making AI easier to understand, securing the models, and joining them with other cybersecurity tools.

Keywords: cybersecurity, threat detection, attack prevention, neural networks, cyber threats.

Постановка проблеми. Ідентифікувати кіберзагрози стає дедалі складніше, оскільки вони стали набагато більш досконалими. Дослідження показали, що кількість кібератак зростає, а зловмисники постійно знаходять нові способи атак. Загалом, сучасні кіберзлочини пристосовуються, можуть зламувати поширені бар'єри безпеки і залежать від передових технологій. У цих умовах традиційні рішення з кібербезпеки не працюють належним чином, оскільки вони покладаються на виявлення відомих шаблонів і їм важко розпізнати невідомі, нові загрози.

Штучний інтелект дозволяє організаціям виявляти і запобігати кібератакам, обробляючи великі обсяги інформації, виявляючи незвичну поведінку і передбачаючи небезпеки ще до того, як вони відбудуться. Ці технології дозволяють створювати інструменти, здатні впоратися з раніше невідомими формами кібератак, виявляти складну поведінку і налаштовувати автоматизовані кроки реагування на інциденти. Незважаючи на це, використання штучного інтелекту в кібербезпеці може призвести до низки проблем, таких як розуміння його моделей, загроз для власних систем і роздумів про етичні питання використання автоматизованих систем.

Тому важливо проаналізувати методи виявлення та блокування кібератак, які покладаються на штучний інтелект, оцінити їхні результати та запропонувати поради щодо їхнього впровадження.

Аналіз останніх досліджень і публікацій. У різних наукових роботах розглядається, як ШІ використовується для виявлення та зупинення кібератак. Ільєнко та інші [1] розглядають, як ШІ можна використовувати в системах кібербезпеки, зосереджуючись на його перевагах, таких як швидке виявлення загроз і швидке реагування, а також на труднощах, пов'язаних із застосуванням таких технологій. Яненко [2] розглядає проблеми нагляду за впровадженням ШІ в кібербезпеці, вказуючи на те, як він впливає на інформаційну безпеку, і на важливість встановлення відповідних правил для використання ШІ. Метод, запропонований Nsvsc Ramesh та ін. [3], використовує оптимізацію Wolverine разом з комбінацією нейронних мереж для виявлення кіберзагроз у Twitter. Цей підхід показав високу ефективність у ранньому виявленні потенційних загроз. Al-Azzawi et al. [4] досліджують застосування ШІ для моделювання кібератак (Red Teaming), що дозволяє виявляти вразливості систем і розробляти стратегії захисту. Автори провели систематичний огляд, оцінивши

ефективність сучасних підходів. Bonagiri et al. [5] зосередилися на методах машинного навчання для запобігання вторгненням, захисту даних і аналізу загроз у реальному часі, запропонували комплексний підхід до інтеграції ШІ в системи кібербезпеки. Ankalaki et al. [6] порівняли традиційні методи машинного навчання з генеративним ШІ для прогнозування кібератак, довівши, що генеративні моделі, такі як GAN і Transformer, ефективні для виявлення складних, раніше невідомих атак. Bello et al. [7] аналізують роль ШІ у виявленні загроз, аномалій та автоматизованому реагуванні, підкреслюючи значне підвищення ефективності кібербезпеки завдяки технологіям машинного навчання. Zhou et al. [8] досліджують пояснювані моделі машинного навчання для ідентифікації кібератак на основі аналізу мережевого трафіку, наголошуючи на важливості інтерпретованості моделей для довіри до автоматизованих систем. Joshua et al. [9] запропонували систему AIDTIS, яка використовує великі мовні моделі для автоматизованого аналізу загроз, що значно підвищує ефективність їх виявлення. Abdulrahman et al. [10] зосередилися на методах аналізу загроз і автоматизованому реагуванні на інциденти з використанням предиктивної аналітики, що сприяє підвищенню кіберстійкості організацій. Guembe et al. [11] провели бібліометричний аналіз моделей виявлення кібератак на основі ШІ, визначивши ключові тенденції та напрямки досліджень у цій галузі. Rahmati [12] досліджує легкі та пояснювані моделі ШІ для пошуку кіберзагроз у крайових мережах у реальному часі, наголошуючи на потребі алгоритмів для обмежених обчислювальних ресурсів. Prathyusha et al. [13] аналізують вразливості систем ШІ до атак і пропонують стеганографічні методи захисту, досліджуючи способи підвищення їхньої стійкості.

Аналіз досліджень показує, що методи виявлення та запобігання кібератакам на основі ШІ значно перевищують традиційні підходи за ефективністю, особливо у виявленні невідомих загроз. Генеративні моделі [6], ансамблеві методи [3] та пояснювані системи [8] демонструють високий потенціал, але потребують вирішення проблем інтерпретованості, вразливості до атак і високих обчислювальних вимог [12, 13]. Дослідження підкреслюють необхідність комплексного підходу, що поєднує різні методи ШІ, для створення адаптивних і стійких систем кібербезпеки.

Мета статті – дослідження сучасних методів виявлення та запобігання кібератакам на основі штучного інтелекту, систематизація підходів до інтеграції ШІ в системи кібербезпеки, аналіз їхньої ефективності та розробка рекомендацій щодо їх практичного впровадження. Дослідження спрямоване на вирішення наступних завдань:

Систематизація основних методів виявлення кібератак з використанням технологій штучного інтелекту.

Аналіз ефективності різних алгоритмів машинного навчання для виявлення та класифікації кіберзагроз.

Дослідження підходів до автоматизованого реагування на інциденти на основі штучного інтелекту.

Розробка рекомендацій щодо впровадження систем кібербезпеки на основі ШІ в організаціях різного типу.

Виклад основного матеріалу. Останні події в сфері кібербезпеки демонструють стрімке зростання кількості та складності кібератак, що вимагає розробки нових, більш ефективних методів захисту. Технології штучного інтелекту пропонують значний потенціал для вирішення цієї проблеми, дозволяючи виявляти нові, раніше невідомі типи атак та автоматизувати процеси реагування на інциденти.

Методи виявлення кібератак на основі штучного інтелекту можна класифікувати за різними критеріями, включаючи тип використовуваних алгоритмів, принципи роботи та сферу застосування. У таблиці 1 представлено основні категорії методів виявлення кібератак на основі ШІ.

Таблиця 1.

Класифікація методів виявлення кібератак на основі ШІ

Категорія	Методи	Принцип роботи	Переваги	Обмеження
Методи виявлення аномалій	Автоенкодерери, однокласова SVM, ізоляційні ліси	Виявлення відхилень від нормальної поведінки	Здатність виявляти нові типи атак	Висока частота помилкових спрацьовувань
Методи класифікації	Нейронні мережі, Random Forest, SVM	Класифікація подій на основі навчання на розмічених даних	Висока точність для відомих типів атак	Обмежена здатність виявляти нові типи атак
Гібридні методи	Ансамблеві методи, комбінації різних алгоритмів	Поєднання переваг різних підходів	Підвищена точність і надійність	Складність реалізації та налаштування
Генеративні моделі	GAN, LSTM, Transformer	Генерація та аналіз сценаріїв атак	Висока адаптивність та здатність виявляти складні атаки	Висока обчислювальна складність
Методи глибокого навчання	CNN, RNN, Deep Reinforcement Learning	Автоматичне виділення ознак та складних патернів	Висока ефективність при роботі з великими обсягами даних	Вимогливість до обчислювальних ресурсів та тренувальних даних

Як видно з таблиці 1, кожен тип методів має свої переваги та обмеження, що визначає доцільність їх використання в конкретних сценаріях. Для побудови ефективної системи кібербезпеки часто використовують комбінацію різних методів, що дозволяє компенсувати обмеження окремих підходів [5].

Ефективна система виявлення та запобігання кібератакам на основі ШІ повинна включати декілька взаємопов'язаних компонентів, які забезпечують збір та аналіз даних, виявлення загроз, прийняття рішень та реагування на інциденти, Рис. 1-4.



Рисунок 1. Процес збору та попередньої обробки даних

Джерело: авторська розробка в PlantUml online editor



Рисунок 2. Процес аналізу та виявлення загроз

Джерело: авторська розробка в PlantUml online editor

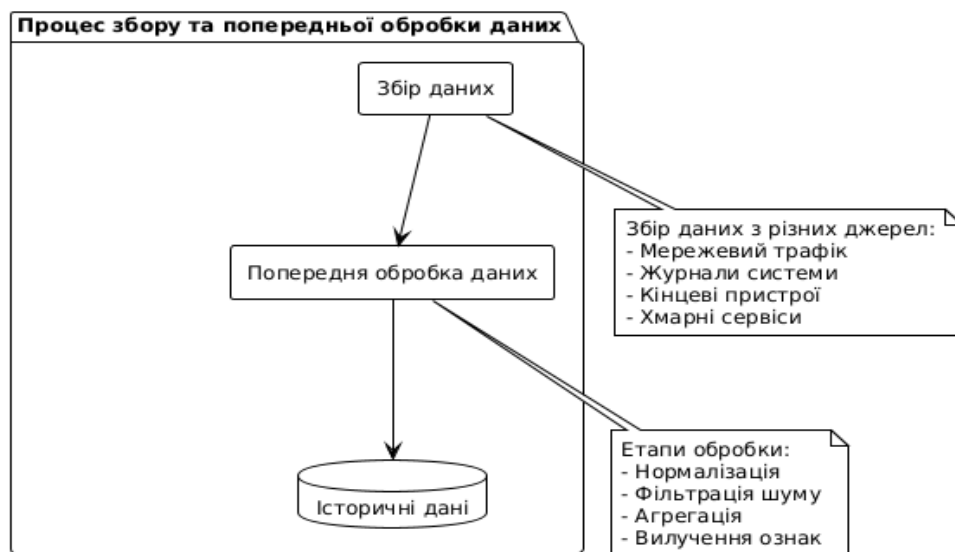


Рисунок 3. Процес автоматизованого реагування та аналітики
Джерело: авторська розробка в PlantUml online editor

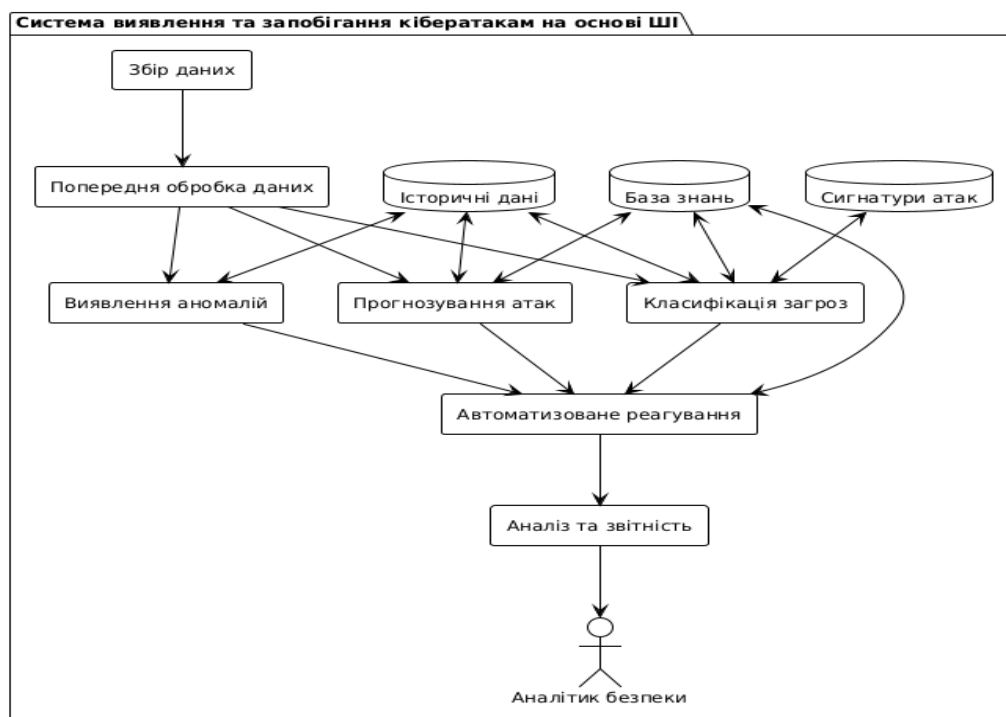


Рисунок 4. Інтегрована система виявлення та запобігання кібератакам
на основі ШІ

Джерело: авторська розробка в PlantUml online editor

Ці рисунки ілюструють, що архітектура системи складається з кількох основних елементів.

Збір даних: Ця частина системи збирає дані про активність у мережі, системні журнали, різні кінцеві точки та хмарні ресурси. Таким чином, система може бачити всю активність у мережі та швидко виявляти можливі небезпеки [3].

Перед початком аналізу дані потрібно попередньо обробити, використовуючи фільтрацію, нормалізацію та вилучення ознак. Правильна підготовка даних на цьому етапі важлива для ефективності машинного навчання [8].

Виявлення аномалій: Це стає можливим завдяки машинному навчанню, яке шукає незвичну поведінку, що може сигналізувати про загрозу. Найчастіше використовуються автокодери, ізоляційні ліси та однокласові SVM [6].

Класифікація загроз: Ця сфера безпеки займається організацією загроз у групи відповідно до їх типів та важливості. Деякі з алгоритмів, що використовуються для цього завдання, - нейронні мережі, випадкові ліси та SVM [5].

Прогнозування атак: Робить прогнози щодо можливих атак за допомогою глибокого навчання та аналізу закономірностей, які спостерігаються в минулих даних. Деякі з основних методів, що використовуються: LSTM, Transformer, GAN та глибоке навчання з підкріпленням [12].

Автоматизоване реагування: Ця частина призначена для управління зараженими системами шляхом виявлення небезпек, блокування певних IP-адрес, обмеження доступу до ресурсів, ізоляції заражених систем та інформування членів команди безпеки [7].

Аналіз і звітність: Завдяки аналізу та звітам, які надаються аналітикам з безпеки, система регулярно вдосконалюється і може впоратися з новими типами загроз.

Завдяки своїй архітектурі система може змінюватися, щоб протистояти новим загрозам і забезпечувати організації підвищений рівень захисту [1].

Згідно з дослідженнями [3, 5, 6], методи виявлення кібератак на основі штучного інтелекту демонструють значно вищу ефективність порівняно з традиційними підходами. Так, ансамблеві методи глибокого навчання, запропоновані Nsvsc Ramesh та співавторами [3], показали точність виявлення загроз на рівні 97.8%, що суттєво перевищує показники традиційних систем.

Ankalaki та співавтори [6] провели порівняльний аналіз різних методів машинного навчання для прогнозування кібератак та встановили, що генеративні моделі, зокрема GAN та Transformer, демонструють найвищу ефективність при виявленні складних, раніше невідомих типів атак.

Bonagiri та співавтори [5] розробили комплексний підхід до інтеграції алгоритмів машинного навчання в системи кібербезпеки та показали, що такий підхід дозволяє знизити кількість помилкових спрацьовувань на 45% порівняно з традиційними системами.

Незважаючи на значний потенціал, використання штучного інтелекту в сфері кібербезпеки пов'язане з низкою викликів та проблем:

Багато ефективних моделей ШІ, зокрема глибинні нейронні мережі, є "чорними скриньками", що ускладнює інтерпретацію їхніх рішень та знижує довіру до них [8].

Самі системи ШІ можуть бути вразливими до різних типів атак, включаючи змагальні атаки (adversarial attacks) та атаки на дані [13].

Багато методів глибинного навчання вимагають значних обчислювальних ресурсів, що може обмежувати їх використання в системах реального часу [12].

Ефективність методів машинного навчання сильно залежить від кількості та якості тренувальних даних, що може бути проблемою для виявлення нових типів атак [2].

Використання автоматизованих систем прийняття рішень у сфері кібербезпеки підіймає низку етичних та правових питань, включаючи проблеми приватності та відповідальності [4].

Для подолання цих викликів дослідники пропонують різні підходи, включаючи розробку пояснюваних моделей ШІ [8], використання методів захисту від змагальних атак [13] та розробку легких алгоритмів для систем з обмеженими ресурсами [12].

Для демонстрації практичного застосування методів штучного інтелекту для виявлення кібератак, було реалізовано систему виявлення аномалій на основі автоенкодера з використанням Python в Google Colab (Рис. 5-7).



Рис. 5 Матриця помилок для класифікації аномалій у мережевому трафіку

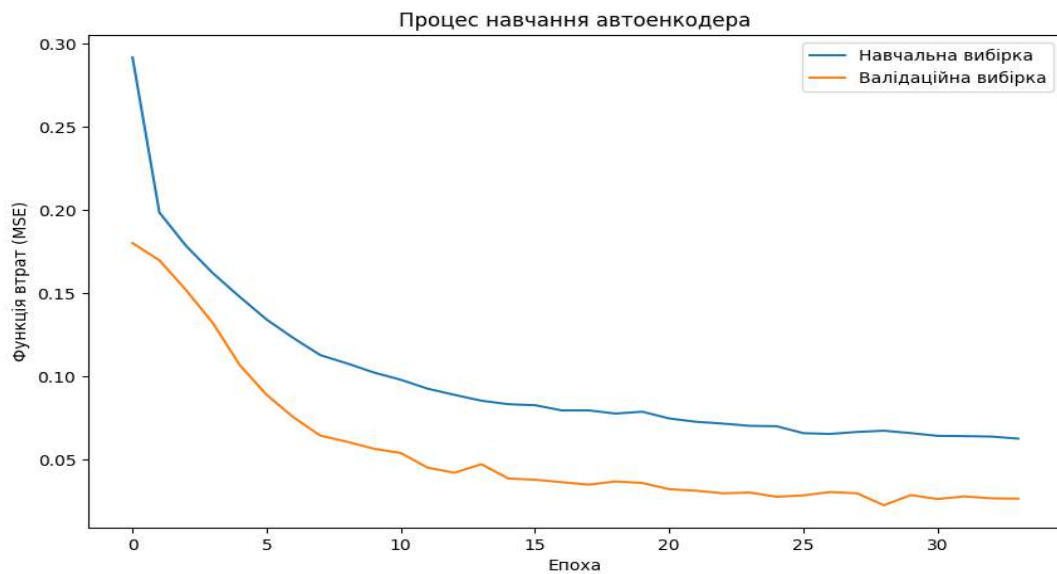


Рис. 6 Розподіл помилок реконструкції та класифікація аномалій

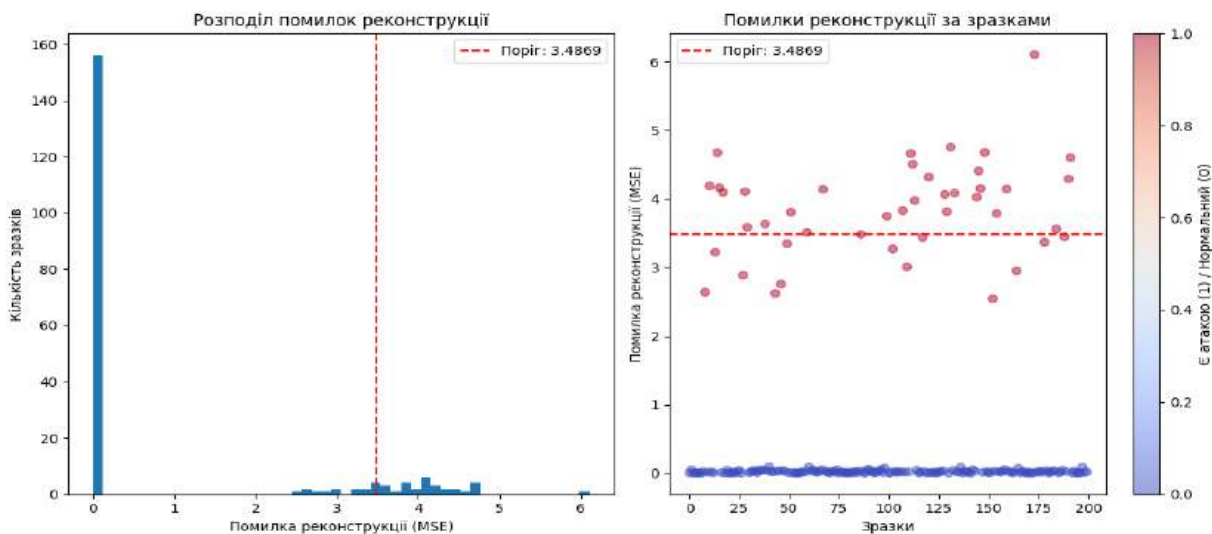


Рис. 7 Динаміка функції втрат автоенкодера під час навчання

Код реалізує систему виявлення аномалій у мережевому трафіку з використанням автоенкодера. Експерименти проводилися на синтетичних даних, згенерованих для моделювання нормального мережевого трафіку (80% зразків) та аномального трафіку (20% зразків), що імітують потенційні кібератаки, з використанням нормального розподілу з різними параметрами для нормальних і аномальних даних. Автоенкодер навчається лише на нормальних даних і вчиться ефективно реконструювати їх. При появі аномального трафіку (потенційної кібератаки) помилка реконструкції буде значно вищою, що дозволяє виявити загрозу [6].

Ключові особливості реалізації:
Попередня обробка та масштабування даних для підвищення ефективності навчання.

Архітектура автоенкодера з використанням шарів Dropout для запобігання перенавчанню.

Раннє зупинення (Early Stopping) для оптимізації процесу навчання.

Динамічне визначення порогу для класифікації аномалій.

Візуалізація результатів для аналізу ефективності моделі.

Можливість використання навченої моделі для виявлення аномалій у режимі реального часу.

Ця реалізація демонструє практичне застосування методів штучного інтелекту для виявлення потенційних кібератак та може бути інтегрована в більш складні системи кібербезпеки [5, 12].

Ефективне використання штучного інтелекту в сфері кібербезпеки вимагає інтеграції методів ШІ в комплексні системи захисту, що забезпечують не лише виявлення загроз, але й автоматизоване реагування на інциденти та постійне вдосконалення систем захисту [10].

Під керівництвом Джошуа команда представила AIDTIS [9], яка спирається на великі мовні моделі для автоматизації аналізу загроз і побудови систем виявлення. Цей метод дуже добре працює при виявленні нових небезпек і реагуванні на них за допомогою машин.

Абдулрахман та його співавтори [10] представили спосіб аналізу загроз та реагування на інциденти за допомогою предиктивної аналітики та глибокого навчання. Таким чином, організації можуть стати більш захищеними та швидше реагувати на загрози, прогножуючи ризики та маючи автоматизовані системи.

Рахматі М. [12] запропонував використовувати легкий ШІ для виявлення кіберзагроз у режимі реального часу в периферійних мережах, що допоможе розширити сферу застосування ШІ в кібербезпеці.

Щоб ШІ працював у сфері кібербезпеки, необхідно гарантувати, що він може впоратися як з ворожими атаками, так і з атаками на дані. Prathyusha, K., Vinoj, J., and Manikandan, K. B. [13] опублікували дослідження про недоліки безпеки систем ШІ та запропонували заходи для їх захисту, оскільки це забезпечує надійність систем ШІ в кібербезпеці.

Висновки. Застосування методів штучного інтелекту до систем захисту від кібератак може суттєво підвищити рівень захисту інформаційних систем від сучасних загроз. Коли справа доходить до виявлення нових або складних видів загроз, глибоке навчання і генеративні моделі в ШІ показують набагато кращі результати, ніж традиційні методи. Коли кілька алгоритмів використовуються разом, стає легше виявляти загрози та зменшити кількість хибних спрацьовувань. Така система повинна складатися з компонентів для

збору та інтерпретації даних, виявлення аномалій, класифікації загроз, прогнозування атак і автоматичного реагування на них. У сфері кібербезпеки дуже важливо, щоб моделі ШІ можна було інтерпретувати, оскільки це підвищує довіру до результатів роботи автоматизованих систем. Для того, щоб використовувати глибоке навчання для виявлення кібератак на системах з невеликими обчислювальними ресурсами, необхідно розробляти легкі алгоритми. Впровадження методів штучного інтелекту в повністю інтегровані системи кібербезпеки дає змогу виявляти загрози, автоматизувати реагування на них і постійно вдосконалювати захист. Системи кібербезпеки залежать від стійкості AI до будь-яких потенційних загроз. З цієї причини використання штучного інтелекту допомагає виявляти і блокувати кібератаки, сприяючи розробці передових систем безпеки, здатних впоратися з поточними загрозами. Роботу над підвищенням чіткості моделей ШІ, їхньою безпечністю та додаванням до різних рішень у сфері кібербезпеки слід продовжувати і надалі.

Література:

1. Ільєнко А., Ільєнко С., Яковенко О., Галич Є., Павленко В. Перспективи інтеграції штучного інтелекту в системі кібербезпеки. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка». 2024. № 1(25). С. 318–329. DOI: <https://doi.org/10.28925/2663-4023.2024.25.318329> URL: <https://csecurity.kubg.edu.ua/index.php/journal/article/view/631>
2. Яненкова І. Штучний інтелект у сфері кібербезпеки: виклики, регулювання, впливи. Матеріали конференцій МЦНД. 2025. DOI: 10.62731/mcnd-21.03.2025.001. URL: https://www.researchgate.net/publication/391224237_STUCNIJ_INTELEKTU_U_SFERI_KIBERBEZPEKI_VIKLIKI_REGULUVANNA_VPLIVI
3. Nsvsc Ramesh, S., Al Fardan, B. M. M., Anupama, C. S. S., Vijaya Kumar, K., Cho, S., Acharya, S., Yoon, C. Leveraging Cyberattack News Tweets for Advanced Threat Detection and Classification Using Ensemble of Deep Learning Models With Wolverine Optimization Algorithm. IEEE Access. 2025. Vol. 13. P. 48343–48358. DOI: <https://doi.org/10.1109/access.2025.3550378> URL: <https://ieeexplore.ieee.org/document/10921656>
4. Al-Azzawi, M., Doan, D., Sipola, T., Hautamäki, J., Kokkonen, T. Red Teaming with Artificial Intelligence-Driven Cyberattacks: A Scoping Review (Version 1). arXiv. 2025. DOI: <https://doi.org/10.48550/ARXIV.2503.19626> URL: <https://arxiv.org/abs/2503.19626v1>
5. Bonagiri, K., Krishnamoorthy, P., Keerthiga, V., Kirubakaran, D., David, R., Nancharaiah, B. Cybersecurity With Machine Learning: Implementing AI Algorithms for Intrusion Prevention, Advanced Data Protection, and Real-Time Threat Analysis. 2025 International Conference on Computational, Communication and Information Technology (ICCCIT). 2025. P. 292–298. DOI: <https://doi.org/10.1109/icccit62592.2025.10928115> URL: <https://ieeexplore.ieee.org/document/10928115>
6. Ankalaki, S., Atmakuri, A. R., Pallavi, M., Hukkeri, G. S., Jan, T., Naik, G. R. Cyber Attack Prediction: From Traditional Machine Learning to Generative Artificial Intelligence. IEEE Access. 2025. Vol. 13. P. 44662–44706. DOI: <https://doi.org/10.1109/access.2025.3547433> URL: <https://ieeexplore.ieee.org/document/10909100>
7. Aminat Bolaji Bello, Akeem Olakunle Ogundipe, Awobelem A. George, Olabode Anifowose. The role of AI and machine learning in cybersecurity: Advancements in threat detection, anomaly detection and automated response. International Journal of Science and Research Archive. 2025. Vol. 14, iss. 2. P. 1587–1597. DOI: <https://doi.org/10.30574/ijrsra.2025.14.2.0542> URL: <https://journalijrsra.com/node/648>

8. Zhou, Y., Jacquet, M. L., Dawit, R., Fabre, S., Sarawat, D., Khan, F., Newell, M., Liu, Y., Liu, D., Chen, H., Wang, J., Wang, H. Explainable Machine Learning for Cyberattack Identification from Traffic Flows (Version 1). arXiv. 2025. DOI: <https://doi.org/10.48550/ARXIV.2505.01488> URL: <https://arxiv.org/abs/2505.01488v1>
9. Emmanuel Joshua, John Do, Rushil Patel. AI-Driven Threat Intelligence System (AIDTIS): Leveraging large language models for automated threat research and detection development. International Journal of Science and Research Archive. 2025. Vol. 14, iss. 3. P. 270–285. DOI: <https://doi.org/10.30574/ijrsra.2025.14.3.0339> URL: <https://journalijrsra.com/node/703>
10. Abdulrahman, I. A., Ogor, U. C., Ayodele, G. T., Anadozie, C., & Alebiosu, J. AI-Driven Threat Intelligence and Automated Incident Response: Enhancing Cyber Resilience through Predictive Analytics. Research Journal in Civil, Industrial and Mechanical Engineering. 2025. Vol. 2, iss. 1. P. 16–32. DOI: <https://doi.org/10.61424/rjcime.v2i1.236> URL: <https://bluemarkpublishers.com/index.php/RJIEE/article/view/236>
11. Guembe, B., Misra, S., Azeta, A., Lopez-Baldominos, I. Bibliometric analysis of artificial intelligence cyberattack detection models. Artificial Intelligence Review. 2025. Vol. 58, iss. 6. DOI: <https://doi.org/10.1007/s10462-025-11167-0> URL: <https://link.springer.com/article/10.1007/s10462-025-11167-0>
12. Rahmati, M. Towards Explainable and Lightweight AI for Real-Time Cyber Threat Hunting in Edge Networks (Version 1). arXiv. 2025. DOI: <https://doi.org/10.48550/ARXIV.2504.16118> URL: <https://arxiv.org/abs/2504.16118v1>
13. Prathyusha, K., Vinoj, J., Manikandan, K. B. Exploring Artificial Intelligence Security: A Comparative Study of Adversarial Attacks and Steganographic Defenses. 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL). 2025. P. 20–25. DOI: <https://doi.org/10.1109/icsadl65848.2025.10933147> URL: <https://ieeexplore.ieee.org/document/10933147>

References:

1. Ilienکو, A., Ilienکو, S., Yakovenko, O., Halych, Ye., Pavlenko, V. (2024). Perspektyvy inteɦratsii shchuchnoho intelektu v systemi kiberbezpeky [Perspectives of artificial intelligence integration in cybersecurity systems]. *Elektronne fakhove naukove vydannia "Kiberbezpeka: osvita, nauka, tekhnika" – Cybersecurity: Education, Science, Technology*, (1(25)), 318–329. DOI: <https://doi.org/10.28925/2663-4023.2024.25.318329> Retrieved from <https://csecurity.kubg.edu.ua/index.php/journal/article/view/631> [in Ukrainian].
2. Yanenkova, I. (2025). Shchuchnyi intelekt u sferi kiberbezpeky: vyklyky, rehuliuвання, vplyvy [Artificial intelligence in the field of cybersecurity: Challenges, regulation, impacts]. Materialy konferentsii MTSND – Conference Proceedings of the International Center for Scientific Research. DOI: 10.62731/mcnd-21.03.2025.001. Retrieved from https://www.researchgate.net/publication/391224237_STUCNIJ_INTELEKTU_U_SFERI_KIBERBEZPEKI_VIKLIKI_REGULUVANNA_VPLIVI [in Ukrainian].
3. Nsvsc Ramesh, S., Al Fardan, B. M. M., Anupama, C. S. S., Vijaya Kumar, K., Cho, S., Acharya, S., & Yoon, C. (2025). Leveraging Cyberattack News Tweets for Advanced Threat Detection and Classification Using Ensemble of Deep Learning Models With Wolverine Optimization Algorithm. IEEE Access, 13, 48343–48358. DOI: <https://doi.org/10.1109/access.2025.3550378> Retrieved from <https://ieeexplore.ieee.org/document/10921656> [in English].
4. Al-Azzawi, M., Doan, D., Sipola, T., Hautamäki, J., & Kokkonen, T. (2025). Red Teaming with Artificial Intelligence-Driven Cyberattacks: A Scoping Review (Version 1). arXiv. DOI: <https://doi.org/10.48550/ARXIV.2503.19626> Retrieved from <https://arxiv.org/abs/2503.19626v1> [in English].

5. Bonagiri, K., Krishnamoorthy, P., Keerthiga, V., Kirubakaran, D., David, R., & Nancharaiah, B. (2025). Cybersecurity With Machine Learning: Implementing AI Algorithms for Intrusion Prevention, Advanced Data Protection, and Real-Time Threat Analysis. 2025 International Conference on Computational, Communication and Information Technology (ICCCIT), 292–298. DOI: <https://doi.org/10.1109/icccit62592.2025.10928115> Retrieved from <https://ieeexplore.ieee.org/document/10928115> [in English].
6. Ankalaki, S., Atmakuri, A. R., Pallavi, M., Hukkeri, G. S., Jan, T., & Naik, G. R. (2025). Cyber Attack Prediction: From Traditional Machine Learning to Generative Artificial Intelligence. IEEE Access, 13, 44662–44706. DOI: <https://doi.org/10.1109/access.2025.3547433> Retrieved from <https://ieeexplore.ieee.org/document/10909100> [in English].
7. Aminat Bolaji Bello, Akeem Olakunle Ogundipe, Awobelem A. George, & Olabode Anifowose. (2025). The role of AI and machine learning in cybersecurity: Advancements in threat detection, anomaly detection and automated response. International Journal of Science and Research Archive, 14(2), 1587–1597. DOI: <https://doi.org/10.30574/ijrsra.2025.14.2.0542> Retrieved from <https://journalijsra.com/node/648> [in English].
8. Zhou, Y., Jacquet, M. L., Dawit, R., Fabre, S., Sarawat, D., Khan, F., Newell, M., Liu, Y., Liu, D., Chen, H., Wang, J., & Wang, H. (2025). Explainable Machine Learning for Cyberattack Identification from Traffic Flows (Version 1). arXiv. DOI: <https://doi.org/10.48550/ARXIV.2505.01488> Retrieved from <https://arxiv.org/abs/2505.01488v1> [in English].
9. Emmanuel Joshua, John Do, & Rushil Patel. (2025). AI-Driven Threat Intelligence System (AIDTIS): Leveraging large language models for automated threat research and detection development. International Journal of Science and Research Archive, 14(3), 270–285. DOI: <https://doi.org/10.30574/ijrsra.2025.14.3.0339> Retrieved from <https://journalijsra.com/node/703> [in English].
10. Abdulrahman, I. A., Ogor, U. C., Ayodele, G. T., Anadozie, C., & Alebiosu, J. (2025). AI-Driven Threat Intelligence and Automated Incident Response: Enhancing Cyber Resilience through Predictive Analytics. Research Journal in Civil, Industrial and Mechanical Engineering, 2(1), 16–32. DOI: <https://doi.org/10.61424/rjcime.v2i1.236> Retrieved from <https://bluemarkpublishers.com/index.php/RJIEE/article/view/236> [in English].
11. Guembe, B., Misra, S., Azeta, A., & Lopez-Baldominos, I. (2025). Bibliometric analysis of artificial intelligence cyberattack detection models. Artificial Intelligence Review, 58(6). DOI: <https://doi.org/10.1007/s10462-025-11167-0> Retrieved from <https://link.springer.com/article/10.1007/s10462-025-11167-0> [in English].
12. Rahmati, M. (2025). Towards Explainable and Lightweight AI for Real-Time Cyber Threat Hunting in Edge Networks (Version 1). arXiv. DOI: <https://doi.org/10.48550/ARXIV.2504.16118> Retrieved from <https://arxiv.org/abs/2504.16118v1> [in English].
13. Prathyusha, K., Vinoj, J., & Manikandan, K. B. (2025). Exploring Artificial Intelligence Security: A Comparative Study of Adversarial Attacks and Steganographic Defenses. 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), 20–25. DOI: <https://doi.org/10.1109/icsadl65848.2025.10933147> Retrieved from <https://ieeexplore.ieee.org/document/10933147> [in English].