Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
Факультет комп'ютерно-інформаційних систем і програмної інженерії
(назва факультету )
Кафедра кібербезпеки
(повна назва кафедри)

# КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр
(освітній рівень)

на тему: «Автоматизація виявлення кіберзагроз із застосуванням штучного інтелекту»

Виконала: студентка

Спеціальності:
125 «Кібербезпека та захист інформації»
(шифр і назва напряму підготовки, спеціальності)
Лунгол О.М.
підпис (прізвище та ініціали)

Керівник Муж В.В.
підпис (прізвище та ініціали)

Нормоконтроль Тимощук Д.І.
підпис (прізвище та ініціали)

Завідувач кафедри Загородна Н.В.
підпис (прізвище та ініціали)

Рецензент
підпис (прізвище та ініціали)

м. Тернопіль – 2024

Міністерство освіти і науки України
**Тернопільський національний технічний університет імені Івана Пулюя**

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра кібербезпеки
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Загородна Н.В.
(підпис) (прізвище та ініціали)
«__» _____ 2024 р.

# ЗАВДАННЯ
## НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 125 Кібербезпека та захист інформації
(шифр і назва спеціальності)

Студенту Лунгол Ользі Миколаївні
(прізвище, ім'я, по батькові)

1. Тема роботи Автоматизація виявлення кіберзагроз із застосуванням штучного інтелекту

Керівник роботи Муж Валерій Вікторович, кандидат юридичних наук, доцент кафедри кібербезпеки
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «22» 11 2024 року № 4/7-1120

2. Термін подання студентом завершеної роботи 12.12.2024

3. Вихідні дані до роботи Автоматизація цифрової розвідки засобами штучного інтелекту. Розробка моделей автоматизації виявлення кіберзагроз із застосуванням штучного інтелекту.

4. Зміст роботи (перелік питань, які потрібно розробити)
Проаналізувати роль штучного інтелекту в забезпеченні кібербезпеки.
Розглянути сучасні алгоритми машинного навчання для класифікації та аналізу кіберзагроз.
Розробити алгоритми обробки та аналізу даних із відкритих джерел засобами ШІ.
Розробити моделі автоматизованого виявлення кіберзагроз засобами штучного інтелекту.
Описати охорону праці та безпеку в надзвичайних ситуаціях.
Розробити стратегії захисту систем штучного інтелекту від кібератак. Висновки.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)
Тема, мета, задачі. Наукова новизна та практичне значення роботи. Теоретичні аспекти застосування ШІ у виявленні кіберзагроз. Концептуальні засади та роль ШІ у забезпеченні кібербезпеки. Огляд сучасних алгоритмів машинного навчання для класифікації та аналізу кіберзагроз. Теоретичні основи цифрової розвідки у виявленні кіберзагроз. Інтеграція методів ШІ для автоматизації цифрової розвідки у виявленні та аналізі кіберзагроз. Методологічні підходи та принципи цифрової розвідки до ідентифікації та аналізу кіберзагроз. Удосконалення та автоматизація цифрової розвідки у виявленні кіберзагроз із застосуванням технологій ШІ. Удосконалення методів OSINT для виявлення кіберзагроз за допомогою ШІ. Автоматизація методів SOCMINT для моніторингу соціальних мереж на предмет кіберзагроз. Удосконалення методів GEOINT для виявлення локалізованих кіберзагроз. Розробка алгоритмів обробки та аналізу даних із відкритих джерел засобами ШІ для ідентифікації потенційних кіберзагроз. Розробка моделей автоматизованого виявлення кіберзагроз засобами ШІ. Модель гібридного виявлення аномалій у мережевому трафіку. Модель глибинної нейронної мережі для виявлення вторгнень у мережевий трафік. Використання трансформерів для аналізу кіберзагроз у текстових даних. Модель автоматизованої роботи DI для виявлення кіберзагроз. Стратегії захисту систем штучного інтелекту від кібератак. Висновки.

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---|---|---|---|
| | | завдання видав | завдання прийняв |
| Охорона праці | Осухівська Г.М., к.т.н., доцент | | |
| Безпека в надзвичайних ситуаціях | Клепчик В.М., проректор з адміністративно-господарської роботи та будівництва | | |

7. Дата видачі завдання 22.11.2024 р.

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів роботи | Термін виконання етапів роботи | Примітка |
|---|---|---|---|
| 1. | Ознайомлення із завданням до кваліфікаційної роботи | 22.11 | Виконано |
| 2. | Визначення теми та мети кваліфікаційної роботи | 22.11 | Виконано |
| 3. | Проведення аналізу літературних джерел з питань застосування штучного інтелекту у забезпеченні кібербезпеки | 23.11 – 24.11 | Виконано |
| 4. | Огляд та класифікація сучасних алгоритмів машинного та глибокого навчання для виявлення кіберзагроз | 25.11 – 27.11 | Виконано |
| 5. | Розробка методологічних підходів та принципів цифрової розвідки до автоматизації процесів виявлення кіберзагроз | 28.11 – 30.11 | Виконано |
| 6. | Розробка та опис моделей автоматизованого виявлення кіберзагроз засобами штучного інтелекту | 01.12 – 02.12 | Виконано |
| 7. | Проведення експериментального дослідження автоматизації процесів виявлення кіберзагроз з використанням алгоритмів штучного інтелекту | 03.12 – 04.12 | Виконано |
| 8. | Аналіз небезпек та розробка стратегій захисту систем штучного інтелекту у моделях автоматизованого виявлення кіберзагроз | 05.12 – 06.12 | Виконано |
| 10. | Оформлення розділів кваліфікаційної роботи згідно з вимогами | 06.12 – 08.12 | Виконано |
| 12. | Нормоконтроль | 08.12 – 12.12 | Виконано |
| 13. | Перевірка на плагіат | 12.12 – 14.12 | Виконано |
| 14. | Попередній захист кваліфікаційної роботи | 20.12.2024 | Виконано |
| 15. | Захист кваліфікаційної роботи | 28.12.2024 | |
| | | | |
| | | | |

Студент _____ Лунгол О.М.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Муж В.В.
(підпис) (прізвище та ініціали)

# АНОТАЦІЯ

Автоматизація виявлення кіберзагроз із застосуванням штучного інтелекту // ОР «Магістр» // Лунгол Ольга Миколаївна // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра кібербезпеки, група СБмд-61 // Тернопіль, 2024 // С. 102, рис. – 5, табл. – 1 , кресл. – -, додат. – 8.

Ключові слова: штучний інтелект, кібербезпека, цифрова розвідка, модель, превентивні заходи, шкідливе програмне забезпечення.

Актуальність роботи пов'язана із зростанням кількості та складності кіберзагроз у сучасному цифровому світі.

Магістерська робота спрямована на дослідження і розробку методів автоматизації виявлення кіберзагроз із застосуванням штучного інтелекту. Робота включає аналіз ролі штучного інтелекту в забезпеченні кібербезпеки, дослідження сучасних алгоритмів машинного навчання для класифікації та аналізу кіберзагроз, інтеграцію методів штучного інтелекту в автоматизацію цифрової розвідки, розробку моделей автоматизованого виявлення кіберзагроз. Окрема увага приділена розробці стратегій захисту систем штучного інтелекту від кібератак.

Результати кваліфікаційної роботи можуть бути використані для впровадження в системи кібербезпеки, розробки нових програмних продуктів і рішень у галузі кібербезпеки, підвищення ефективності існуючих методів кіберзахисту, навчання та підготовки фахівців в галузі кібербезпеки та захисту інформації, а також для проведення подальших наукових досліджень.

# ABSTRACT

Automation of Cyber Threat Detection Using Artificial Intelligence // Thesis of educational level "Master"// Olha Lunhol // Ternopil Ivan Puluj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Cybersecurity, group СБмд-61 // Ternopil, 2024 // P. 102, figs. – 5, tables – 1, chair. – -, added. – 8.

Keywords: artificial intelligence, cybersecurity, digital intelligence, model, preventive measures, malware.

The relevance of this work is linked to the increasing number and complexity of cyber threats in today's digital world.

This master's thesis aims to research and develop methods for automating the detection of cyber threats using artificial intelligence. The work includes analyzing the role of artificial intelligence in ensuring cybersecurity, studying modern machine learning algorithms for classifying and analyzing cyber threats, integrating artificial intelligence methods into the automation of digital intelligence, and developing models for automated threat detection. Special attention is paid to developing strategies for protecting artificial intelligence systems from cyber attacks.

The results of this qualification work can be used for implementation in cybersecurity systems, developing new software products and solutions in the field of cybersecurity, improving the efficiency of existing cyber protection methods, training and preparing specialists in the field of cybersecurity and information protection, and conducting further scientific research.

ЗМІСТ

# ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ML — машинне навчання (Machine Learning)

DL — глибоке навчання (Deep Learning)

ШІ — штучний інтелект (Artificial Intelligence)

NLP — обробка природної мови (Natural Language Processing)

CNN — згорткові нейронні мережі (Convolutional Neural Networks)

RNN — рекурентні нейронні мережі (Recurrent Neural Networks)

OSINT — розвідка на основі відкритих джерел (Open Source Intelligence)

RL — методи навчання з підкріпленням (Reinforcement Learning)

ШПЗ — шкідливе програмне забезпечення

SIEM — системи управління інформацією та подіями безпеки (Security Information and Event Management)

DDoS-атака — розподілена атака на відмову в обслуговуванні, яка спрямована на перевантаження мережевих ресурсів або сервісів з метою зробити їх недоступними для користувачів (Distributed Denial of Service)

URL — уніфікований ідентифікатор ресурсу, що використовується для вказівки адреси ресурсу в Інтернеті (Uniform Resource Locator)

IP-адреса — унікальний числовий ідентифікатор пристрою в мережі, що використовується для його адресації і зв’язку з іншими пристроями (Internet Protocol Address)

APT — тип кібератак «Складні постійні загрози» (Advanced Persistent Threats)

DI — цифрова розвідка (Digital Intelligence)

SOCMINT — розвідка із соціальних мереж (Social Media Intelligence)

GEOINT — геопросторова розвідка (Geospatial Intelligence)

# ВСТУП

Актуальність теми. Постійне зростання обсягів даних і збільшення кількості кіберзагроз вимагає нових підходів до їх виявлення та нейтралізації. Традиційні методи кібербезпеки не здатні забезпечити належний рівень захисту через їхню обмежену ефективність і необхідність значних ресурсів для аналізу великих обсягів даних. Застосування штучного інтелекту (ШІ) у виявленні кіберзагроз відкриває нові можливості для автоматизації процесів і підвищення ефективності кіберзахисту. ШІ дозволяє автоматично аналізувати великі масиви даних, виявляти аномалії та передбачати потенційні загрози, що значно скорочує час реакції на загрози і знижує ризики для інформаційної безпеки. Інтеграція методів ШІ в системи кібербезпеки дає змогу забезпечити проактивний підхід до виявлення та нейтралізації загроз.

Метою дослідження є розробка моделей штучного інтелекту для автоматизації процесів виявлення кіберзагроз.

Завданнями дослідження є:

1. дослідити концептуальні засади та роль штучного інтелекту у забезпеченні кібербезпеки;
2. провести огляд сучасних алгоритмів машинного навчання для класифікації та аналізу кіберзагроз;
3. визначити методологічні підходи до ідентифікації та аналізу кіберзагроз;
4. скласти алгоритми обробки та аналізу даних із відкритих джерел для виявлення потенційних кіберзагроз;
5. розробити моделі автоматизованого виявлення кіберзагроз засобами штучного інтелекту;
6. запропонувати стратегії захисту систем штучного інтелекту від кібератак.

Об’єктом дослідження є процеси виявлення кіберзагроз у цифровому середовищі.

Предметом дослідження кваліфікаційної роботи є методи та алгоритми штучного інтелекту, що використовуються для автоматизації виявлення кіберзагроз.

Наукова новизна одержаних результатів кваліфікаційної роботи. Наукова новизна полягає в розробці інноваційних методів автоматизації процесів виявлення кіберзагроз з використанням алгоритмів штучного інтелекту. Запропоновані методи вдосконалюють існуючі підходи та дозволяють більш ефективно ідентифікувати та аналізувати кіберзагрози, що підвищує рівень кібербезпеки в інформаційних системах.

Практичне значення одержаних результатів. Дослідження має важливе практичне значення для підвищення кібербезпеки за рахунок автоматизації виявлення загроз засобами штучного інтелекту. Впровадження розроблених моделей дозволить ефективно аналізувати великі обсяги даних, швидко і точно ідентифікувати кіберзагрози, а також зменшити час реакції на них. Результати дослідження можуть бути застосовані в різних галузях, зокрема у фінансовому секторі, державних установах та критичній інфраструктурі.

Апробація результатів магістерської роботи. Основні результати проведених досліджень обговорювались на: Всеукраїнській науково-практичній конференції «Безпекова ситуація в Україні в умовах війни: стан, загрози, напрями забезпечення безпеки» (м. Київ, 27.09.2024 р.) [13], II Міжнародній науково-практичній конференції «Topical aspects of modern scientific research» (Tokyo, Japan, 26-28.10.2023) [50], Міжнародній науково-практичній конференції «Кібербезпека в Україні: правові та організаційні питання» (м. Одеса, 17.11.2023) [51], II Міжнародній науково-практичній Інтернет-конференції «Інновації та перспективні шляхи розвитку інформаційних технологій» (ІПШРІТ-2023) (м. Черкаси, 06.12.2023 р.), VII Міжнародній науково-практичній конференції «Реалії та перспективи розбудови правової держави в Україні та світі» (м. Суми, 31.05.2024 р. – 01.06.2024 р.), Міжнародній науково-практичній конференції «Актуальні питання підготовки фахівців для сектору безпеки і оборони в умовах війни» (м. Кропивницький, 19.04.2024 р.) [52], Всеукраїнській науково-практичній конференції «Сучасні пріоритети розвитку України: економічна та інформаційна безпека» (м. Дніпро, 10.10.2023 р.) [53], Всеукраїнській науково-практичній конференції «Сучасні інформаційні

технології в діяльності Національної поліції України» (м. Дніпро, 02.11.2023 р.), ІІІ Всеукраїнській науково-практичній конференції «Регіональні особливості злочинності: сучасні тенденції та стратегії протидії» (м. Кривий Ріг, 02.06.2023 р.) [54], ІІІ Всеукраїнській науково-практичній конференції «Безпека дітей в Інтернеті: попередження, освіта, взаємодія» (м. Кропивницький, 05 – 09.02.2024 р.) [55].

Апробація результатів магістерської роботи також проведена на базі Донецького державного університету внутрішніх справ: Акт впровадження результатів науково-дослідної роботи в освітній процес Донецького державного університету внутрішніх справ від 13.12.2024 та Акт впровадження результатів науково-дослідної роботи в науково-дослідну діяльність Донецького державного університету внутрішніх справ від 13.12.2024.

Публікації. Основні результати кваліфікаційної роботи опубліковано у фаховій статті категорії Б (див. дод. А).

# РОЗДІЛ 1. ТЕОРЕТИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ВИЯВЛЕННІ КІБЕРЗАГРОЗ

## 1.1 Концептуальні засади та роль штучного інтелекту у забезпеченні кібербезпеки

Штучний інтелект (ШІ) є однією з найважливіших сучасних технологій, яка змінює підходи до забезпечення кібербезпеки, особливо у сфері виявлення кіберзагроз. За своєю суттю, ШІ – це комплекс методів і алгоритмів, які дозволяють комп'ютерним системам імітувати людські когнітивні процеси, такі як навчання, розпізнавання образів, аналіз великих масивів даних та прийняття рішень. Використання ШІ в контексті кібербезпеки є актуальним через стрімке зростання обсягів даних, що обробляються в інформаційних системах, а також через еволюцію кіберзагроз, які стають дедалі складнішими та більш витонченими.

Для більш детального дослідження концептуальних засад та ролі ШІ у забезпеченні кібербезпеки ми проаналізували роботи зарубіжних та вітчизняних науковців за тематикою дослідження, а саме: Чжан Чж., Нін Х., Ши Ф., Фарха Ф., Сюй Я., Чжан Ф., Фархін А., Бібху Д., Павankумар Ш., Нікіта Я., Менгідіс Н., Цікрика Т., Врочідіс С., Компаціаріс І., Панайоту П., Балантрапу С., Хілгурт С., Давиденко А., Матовка Т. та ін. [1 – 6].

Так, китайські вчені Чжан Чж., Нін Х., Ши Ф., Фарха Ф., Сюй Я., Чжан Ф. [1] провели детальне дослідження щодо використання методів ШІ у широкому діапазоні програм кібербезпеки. Дослідники вивчили 54 наукові роботи з 2016 по 2020 рр. щодо застосування ШІ в автентифікації доступу користувачів, обізнаності про ситуацію в мережі, моніторингу небезпечної поведінки та ідентифікації аномального трафіку. Науковці, на основі отриманих результатів, представили концептуальну модель кібербезпеки «людини в циклі розвідки» [1].

Вчені Камберлендського університету Фархін А., Бібху Д., Павankумар Ш. та Нікіта Я. [2] серед напрямків використання ШІ в кібербезпеці виділяють: захист паролів та аутентифікацію, управління вразливостями, виявлення

фішингу та контроль його запобігання, поведінкову аналітику та захист мережі. Науковці також розглядають ШІ у кібербезпеці за принципом «come and rescue technology», та зазначають, що навчання на основі досвіду є властивістю алгоритмів ШІ, завдяки якій системи можуть навчатися з подій, що відбувалися раніше. Ці алгоритми застосовуються в технологіях і методах кібербезпеки, щоб запобігти повторенню помилок. Атаки таким чином інтегруються в систему, де алгоритм ШІ виявляють їх та навчаються на них. Науковці стверджують, що технологія ШІ вже зараз забезпечує більш швидке виявлення збоїв у системах, що позитивно впливає на убезпечення від проникнення в систему неавторизованих осіб. Моніторинг трафіку в реальному часі дозволяє технології ШІ ідентифікувати будь-яку активність і відповідно реагувати на неї. Що стосується заходів безпеки даних і протоколів, технологія ШІ на даний момент є однією з найкращих технологій, зазначають Фархін А., Бібху Д., Павankумар Ш. та Нікіта Я. [2].

Дослідники Менгідіс Н., Цікрика Т., Врочідіс С. та Компаціаріс І. [3] також стверджують, що впровадження систем навчання ШІ в кібербезпеку допомагає запобігти атакам на систему. В такому випадку система вивчає дії зловмисників і налаштовується на захист інформації та унеможливлює доступ зловмисників до даних. Наявність системи, яка постійно коригується та навчається, є одним із факторів, які зробили дану технологію дуже ефективною [3].

Удосконалення техніки захисту на основі сигнатур, як однієї з визначальних характеристик технології ШІ у сфері кібербезпеки, описують у своїх дослідженнях науковці Панайоту П., Менгідіс Н., Цикріка Т., Врочідіс С., Компаціаріс І., Балантрапу С., Гільгурт С., Давиденко А., Матовка Т., Пригара М. [4 – 6]. Метод сигнатур полягає у виявленні шкідливих програм та кіберзагроз шляхом аналізу унікальних сигнатур або поведінкових закономірностей характерних для певних типів атак. Використання ШІ дозволяє не лише швидко зіставляти наявні сигнатури з базами даних відомих загроз, а й виявляти закономірності та тенденції, що можуть вказувати на потенційно нові загрози. Це досягається завдяки високій швидкості обробки інформації та здатності алгоритмів ШІ навчатися на прикладах попередніх атак, що також знижує час

реакції на нові кібератаки. Дослідники [4 – 6] зазначають, що вдосконалення техніки кіберзахисту на основі сигнатур є важливим етапом у розвитку превентивних заходів у сфері кібербезпеки, оскільки вона дозволяє зменшити кількість помилкових спрацьовувань та підвищити точність виявлення загроз.

Серед робіт вітчизняних науковців особливої уваги заслуговують дослідження Саєнко Д., Болілого В., Колісника Т., Антоненко А., Бенедіко І., Вічкарука А., Лисенко К., Сижко О., Гладкої Ю., Назаренко Є. та ін. [7 – 10].

Дослідники Саєнко Д. та Колісник Т. [7] підкреслюють, що в сфері кібербезпеки вже активно використовуються різні програми на базі ШІ, такі як SIEM-системи [8], різні спам-фільтри, засоби захищеної автентифікації та інструменти прогнозування атак. Ці програми навчаються на основі попередніх даних про поведінку і можуть визначати, чи є конкретна поведінка шкідливою. За словами вчених [7], відповідно до даних компанії IBM, збитки від порушення даних у всьому світі зменшаться, якщо організації впроваджуватимуть автоматизовані засоби безпеки, включаючи ті, що використовують технології ШІ.

Гладка Ю. та Назаренко Є. [9] стверджують, що ШІ дозволяє практично повністю виключити людей з процесів забезпечення захисту і залишає їм допоміжні функції моніторингу та корекції. Водночас, це підходить до забезпечення швидкості й ефективності виявлення загроз, особливо коли мова йде про великі обсяги даних або складні атаки, які важко відстежити традиційними методами. ШІ здатний не тільки автоматично обробляти дані, але й аналізувати поведінкові закономірності користувачів та пристроїв, виявляючи аномалії, що можуть свідчити про загрози. У свою чергу, це дозволяє системі швидко адаптуватися до нових типів атак та розпізнавати кіберзагрози навіть у випадках, коли сигнатури загроз ще не включені до баз даних. Як відзначають дослідники [9], алгоритми ШІ створюють додаткові можливості для випереджаючого захисту, оскільки здатні передбачати можливі сценарії розвитку подій на основі аналізу попередніх інцидентів та поточних умов.

Антоненко А., Бенедіко І., Вічкарук А., Лисенко К. та Сижко О. [10] аналізують взаємозв’язок між кібербезпекою та ШІ, звертаючи увагу на

можливості як атак, так і захисту з використанням ШІ, а також на захист систем машинного навчання (ML) і створення контенту за допомогою ML. Вони відзначають, що системи ML забезпечують вищу ефективність порівняно з дискримінантними моделями, хоча проблема атак залишається невирішеною. Системи ШІ досягають значних результатів у створенні контенту, включаючи дипфейки, що робить сертифікацію контенту єдиним надійним методом боротьби з ними. У сфері кібербезпеки атаки на основі ШІ поки що переважають над захисними рішеннями, причому реальна кількість атак значно менша за потенційно можливу. Інтелектуальна автоматизація, здійснювана ШІ, сприяє організації та управлінню атаками. Найбільший прогрес у використанні ML спостерігається в аналізі атак, де нейронні мережі використовуються для аналізу даних і розпізнавання шаблонів.

Антоненко А., Бенедіко І., Вічкарук А., Лисенко К. та Сижко О. [10] виділяють наступні напрямки використання ШІ для попередження атак:

– розстановку пріоритетів для попереджень про потенційні атаки;

– виявлення численних спроб злому з плином часу, які є частиною більших і тривалих кампаній зі злому;

– виявлення слідів дій шкідливих програм, як усередині комп’ютера, так і у мережі;

– ідентифікацію потоку шкідливого програмного забезпечення, що запроваджується через конкретну організацію;

– визначення автоматизованих підходів до пом’якшення наслідків атак, коли потрібне швидке реагування, щоб запобігти поширенню атаки.

Проаналізуємо більш детально кожен із видів. Якщо говорити про розстановку пріоритетів для попереджень про потенційні атаки, то ML дозволяє автоматизувати процеси визначення важливості різних попереджень про потенційні кіберзагрози, що знижує ймовірність помилкових спрацьовувань і дає змогу зосередитись на найбільш критичних загрозах. Для цього використовуються алгоритми, що навчаються на історичних даних і можуть класифікувати загрози за різними критеріями, такими як: тип загрози (визначення, чи є попередження про вірус, хакерську атаку, або фішинг);

історичну значущість (якщо попередження стосується події, яка раніше була пов’язана з масштабною атакою, система надає йому вищий пріоритет); тимчасовий фактор (атаки, які здійснюються в критичні періоди або в момент високої активності, можуть мати вищий рівень пріоритету); складність атаки (виявлення складних багатоступеневих атак вимагає більш високого рівня уваги).

Виявлення численних спроб злому з плином часу, які є частиною більших і тривалих кампаній зі злому – це напрямок застосування ML для виявлення постійних або прихованих атак, які зазвичай не виявляються при одиничних вторгненнях. Атаки типу APT часто включають серії малопомітних спроб злому протягом тривалого часу, метою яких є поступове проникнення в систему, збирання даних і знищення або викрадення важливої інформації. ML, зокрема алгоритми кластеризації та виявлення аномалій, дозволяють системам безпеки ідентифікувати спроби злому, що відбуваються поетапно і мають ознаки відсутності явних зовнішніх сплесків активності. Наприклад, методи на основі нейронних мереж можуть аналізувати поведінкові закономірності атакуючих, виявляючи аномалії у послідовності доступу до мережі або зміни у файлах протягом певного часу. Вони також можуть ідентифікувати зв’язок між різними незначними спробами, які можуть бути частиною більшої, організованої кампанії.

Методи ML активно використовуються для виявлення ШПЗ через аналіз поведінки, замість традиційного порівняння з базами сигнатур. Шкідливі програми часто намагаються уникати виявлення, маскуючись або змінюючи свою поведінку. Для цього в системах кібербезпеки застосовують такі методи, як аналіз поведінки (наприклад, при запуску програми її дії можуть аналізуватися в реальному часі. Якщо програма починає виконувати підозрілі операції, такі як зміни в реєстрі або запис у мережу, алгоритми ML можуть класифікувати її як шкідливу, навіть якщо програма досі не була зареєстрована в базі даних антивірусів); ідентифікацію закономірностей атак (наприклад, при виявленні підозрілих закономірностей в мережевих транзакціях або внутрішніх процесах системи, які можуть вказувати на наявність шкідливого програмного

забезпечення); контекстуальну класифікацію (визначення контексту дій програми, що дозволяє відрізняти легітимні процеси від шкідливих. Наприклад, використання глибинних нейронних мереж для автоматичного аналізу дій програм у складних багаторівневих мережах) тощо.

Ідентифікація потоку шкідливого програмного забезпечення, що запроваджується через конкретну організацію, або Living off the Land (LotL) описує стратегію, при якій зловмисники використовують легітимні інструменти або програмне забезпечення організації для здійснення своїх атак. Це можуть бути різноманітні сценарії, коли атака не передбачає використання шкідливих програм, що активно запускаються на комп'ютерах, а, натомість, використовуються вже наявні інструменти в організації для виконання зловмисних дій. Для виявлення таких атак застосовуються:

- аналіз мережевого трафіку, що дозволяє виявляти підозрілі зв'язки між віддаленими пристроями та внутрішніми ресурсами;

- аналіз використання легітимних інструментів. Наприклад, використання алгоритмів для моніторингу команд, які виконуються в рамках адміністративних утиліт, може допомогти ідентифікувати підозрілі операції, що виконуються за допомогою легітимного програмного забезпечення;

- моделі, що вивчають взаємозв'язки між різними компонентами системи. Наприклад, зловмисники можуть використовувати легітимні утиліти для сканування мережі або зміни налаштувань безпеки. ML при цьому допомагає визначити, чи ці дії є частиною звичайної адміністративної роботи або насправді являються шкідливими.

Одним із найважливіших напрямків ML в кібербезпеці є автоматизація реагування на інциденти. В умовах швидкої еволюції кіберзагроз важливо мати системи, які можуть вчасно реагувати, щоб мінімізувати шкоду, див. рис. 1.1.

Автоматичне блокування пристроїв може спрацювати, наприклад, коли система виявляє підозрілу активність, типу спроби зловмисного доступу до важливих файлів, автоматизована система може відключити відповідні пристрої від мережі для запобігання подальшому розповсюдженню атаки.

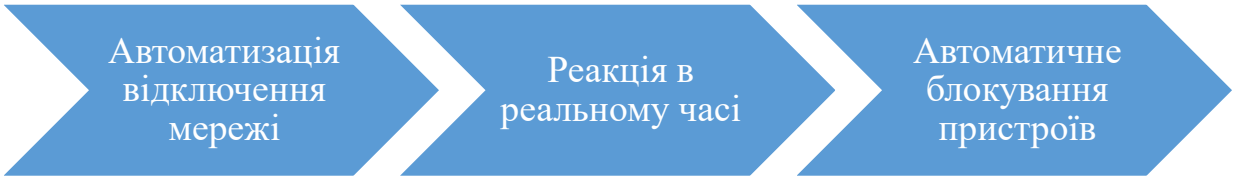


Рисунок 1.1 – Складові автоматизації реагування на інциденти

Автоматизація відключення мережі актуальна у випадках, коли за допомогою алгоритмів виявлено загрози, наприклад, пов'язані з програмами-вимагачами. Негайне відключення мережі дозволить зупинити поширення шкідливого коду.

Підсумовуючи вищезазначене, застосування ШІ в контексті кібербезпеки дозволяє створювати інтелектуальні системи для виявлення, аналізу та реагування на кіберзагрози в режимі реального часу. Це сприяє зменшенню ризиків, підвищенню ефективності заходів кіберзахисту та забезпеченню своєчасного реагування на інциденти. Методи ШІ є важливим компонентом сучасних систем кіберзахисту, оскільки вони вже зараз дозволяють підвищити ефективність виявлення, запобігання та реагування на кіберзагрози, що постійно еволюціонують.

## 1.2 Огляд сучасних алгоритмів машинного навчання для виявлення кіберзагроз

Застосування ШІ для підвищення рівня кібербезпеки є однією з найбільш прогресивних та перспективних галузей сьогодення. Зокрема, у кібербезпеці технології ШІ застосовуються для автоматизації процесів, які раніше вимагали значних людських ресурсів. Наприклад, для виявлення аномалій у мережевому трафіку, аналізу поведінки користувачів, ідентифікації закономірностей, які можуть свідчити про наявність кіберзагроз тощо. Основними методами ШІ, які використовуються для вищезазначених цілей, є машинне навчання (ML), глибоке навчання (DL) та обробка природної мови (NLP). Кожен з цих методів має свою специфіку застосування і може бути ефективним у різних аспектах забезпечення кібербезпеки.

ML, як підгалузь ШІ, особливо важливе у виявленні кіберзагроз, оскільки воно дозволяє системам самостійно навчатися на основі великих масивів даних і покращувати точність виявлення загроз без втручання людини. При цьому, алгоритми ML можна класифікувати на такі основні групи: контрольоване навчання, неконтрольоване навчання та навчання з підкріпленням. У контексті кібербезпеки особливу увагу приділяють контрольованому і неконтрольованому навчанням. Контрольоване навчання використовується для ідентифікації відомих кіберзагроз, де існують великі обсяги навчальних даних, тоді як неконтрольоване навчання допомагає виявляти нові, невідомі поведінки загроз на основі кластеризації та пошуку аномалій.

Основна перевага ML у кібербезпеці полягає в здатності визначати атаки, аналізувати шаблони та закономірності, характерні для вторгнень, швидко оцінювати та пріоритизувати загрози, а також використовувати накопичену інформацію для удосконалення методів виявлення загроз. Згідно з даними Бюро статистики праці США (U.S. Bureau of Labor Statistics), зайнятість у сфері кібербезпеки зросте на 33% у період з 2020 по 2030 рік. Аналогічна тенденція спостерігається і в інших країнах. Водночас, за даними звіту ISC (Internet Systems Consortium), опублікованого в жовтні 2021 року, у світі бракує 2,72 мільйона спеціалістів з кібербезпеки. Таким чином, автоматизація процесів кібербезпеки стає необхідною умовою.

DL, як більш спеціалізована галузь ML, також знаходить широке застосування у кібербезпеці, особливо для обробки неструктурованих даних, таких як мережевий трафік, лог-файли або дані поведінкової аналітики. Алгоритми DL, такі як згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN), є ефективними для виявлення складних залежностей та прихованих закономірностей у великих масивах даних, що дозволяє прогнозувати та ідентифікувати аномалії з високою точністю.

NLP, як ще одна підгалузь ШІ, забезпечує аналіз текстових даних, які можуть містити важливу інформацію про потенційні загрози у кіберпросторі. Інформація з відкритих джерел (OSINT), соціальних мереж або інших текстових потоків може бути автоматично проаналізована для виявлення індикаторів загроз

та прогнозування інцидентів кібербезпеки. Методи NLP дозволяють структурувати і обробляти великі обсяги текстових даних, що є важливим у розробці прогнозних моделей кібербезпеки.

Основні завдання, що стоять у сфері кібербезпеки, включають декілька важливих напрямів, серед яких: запобігання атакам, їхнє виявлення, розслідування інцидентів, класифікацію та аналіз загроз, моделювання та навчання систем кібербезпеки. Кожен із цих напрямів має критичне значення для забезпечення цілісності, конфіденційності та доступності інформаційних ресурсів, і на сьогоднішній день усі вони значною мірою вдосконалюються за допомогою алгоритмів ML.

Запобігання кіберзагрозам є першочерговою метою для будь-якої системи захисту інформації. Завдяки алгоритмам ML можна створювати автоматизовані системи, які здатні аналізувати величезні масиви даних для виявлення потенційних вразливостей ще до того, як вони будуть використані зловмисниками. Наприклад, такі методи, як передбачувальна аналітика та DL, дозволяють прогнозувати ймовірність та вид атак на основі історичних даних і поведінкових шаблонів, що мінімізує ризики та запобігає інцидентам до їхнього виникнення.

Виявлення кіберзагроз є однією з найбільш затребуваних функцій систем кібербезпеки, оскільки здатність виявити атаку на ранніх етапах є вирішальним фактором для мінімізації шкоди. Алгоритми ML, зокрема методи виявлення аномалій та класифікаційні алгоритми, здатні розрізняти нормальну активність та підозрілу поведінку, яка може свідчити про загрозу. Наприклад, методи кластеризації та нейронні мережі допомагають ідентифікувати нетипові зразки поведінки в мережі, що є сигналом можливого кібернападу.

Після виявлення загрози важливим завданням є проведення детального розслідування, щоб визначити джерело атаки, її характер і ступінь завданої шкоди. Використання алгоритмів ML дозволяє автоматизувати аналіз великих обсягів даних журналів подій, звітів про активність і мережевого трафіку. Завдяки цьому фахівці з кібербезпеки можуть швидше ідентифікувати й

візуалізувати ланцюжок дій зловмисників, що дозволяє скоротити час на реагування та запобігти подібним інцидентам у майбутньому.

Різноманіття кіберзагроз, що постійно еволюціонують, потребує надійної класифікації для розуміння природи загрози та подальшої її нейтралізації. Тут алгоритми ML, зокрема дерева рішень, методи байєсівської класифікації та CNN, можуть бути використані для автоматичної класифікації загроз за типами: ШПЗ, фішинг, DDoS-атаки тощо. Класифікація допомагає виявляти й аналізувати особливості конкретних видів атак, а також оцінювати потенційну небезпеку, що дозволяє розробити дієву стратегію захисту.

Дерева рішень є одним із найпоширеніших алгоритмів для класифікації в кібербезпеці. Вони функціонують як структуровані схеми, що допомагають приймати рішення, розбиваючи дані на вузли, кожен з яких представляє певний критерій або атрибут. Кожне відгалуження в дереві відповідає певному рішенню, що дозволяє моделі «розрізняти» між безпечними та потенційно небезпечними діями. Завдяки можливості легкої інтерпретації, дерева рішень дозволяють системам кібербезпеки швидко класифікувати загрози, використовуючи прості критерії, як-от IP-адресу, країну походження трафіку чи тип файлу. Цей метод особливо корисний у ситуаціях, коли необхідно приймати швидкі рішення з мінімальною кількістю обчислень, наприклад, при блокуванні шкідливих IP-адрес у реальному часі.

Байєсівські методи базуються на теорії ймовірностей і дозволяють системам оцінювати ймовірність певної події, спираючись на наявні дані. У контексті кібербезпеки, байєсівські класифікатори використовуються для передбачення ймовірності того, що певний файл, URL або IP-адреса є шкідливими, спираючись на історичні дані та контекст. Наприклад, якщо зафіксовано, що певні типи файлів частіше використовуються для поширення ШПЗ, байєсівський класифікатор надасть таким файлам вищу ймовірність загрози. Байєсівська класифікація є ефективною для аналізу фішингових листів, спаму та інших загроз, що мають багато схожих характеристик із раніше відомими атаками, оскільки вона швидко навчається з минулих випадків і забезпечує високу точність класифікації.

CNN, або згорткові нейронні мережі, є потужним інструментом в задачах, де присутні складні та об’ємні дані, як-от зображення, але в кібербезпеці вони також знайшли застосування для аналізу мережевого трафіку, поведінкових шаблонів і навіть аналізу зловмисного коду. Завдяки багатошаровій структурі, CNN здатні виявляти глибокі, приховані зв’язки в даних. Наприклад, у задачах виявлення ШПЗ CNN можуть аналізувати шкідливий код, виявляючи унікальні поведінки та характеристики, які властиві лише для ШПЗ, що дозволяє відрізняти його від легітимного коду. У контексті мережевого трафіку CNN здатні розпізнавати аномальні шаблони трафіку, які можуть свідчити про DDoS-атаки або інші типи вторгнень.

Постійне вдосконалення та адаптація систем кібербезпеки до нових загроз вимагає створення моделей, які здатні «навчатися» на нових даних і автоматично покращувати свої захисні функції. Алгоритми глибокого навчання, зокрема RNN та методи навчання з підкріпленням, допомагають системам безперервно адаптуватися, аналізуючи дані в реальному часі. Використовуючи такі моделі можна вдосконалити алгоритми виявлення загроз, підвищуючи їхню здатність реагувати на нові типи атак, а також забезпечити безперервний процес оптимізації систем безпеки. Рекурентні нейронні мережі, завдяки своїй здатності зберігати інформацію про попередні стани, є потужними інструментами для аналізу послідовних даних, таких як мережевий трафік або поведінкові закономірності користувачів. Вони дозволяють моделювати тимчасові залежності в даних і можуть ефективно виявляти аномалії, які часто передують кібернападам, таким як DDoS-атаки чи зловмисні дії в системах. Водночас методи навчання з підкріпленням (Reinforcement Learning, RL) допомагають системам адаптуватися до нових умов, покращуючи стратегії реагування на кіберзагрози. Вони базуються на принципі «нагорода-штраф», де система вчиться на основі своїх дій у реальному часі, що дозволяє автоматично оптимізувати процеси виявлення та запобігання атакам. Наприклад, в контексті систем моніторингу кібербезпеки, RL може використовуватися для покращення фільтрації мережевого трафіку, сприяючи виявленню нових, невідомих типів атак, навіть якщо вони ще не були визначені в базах даних сигнатур.

Оскільки кіберзагрози постійно еволюціонують, використання таких алгоритмів дає змогу забезпечити безперервну адаптацію систем безпеки до нових умов. Моделі DL можуть обробляти великі обсяги даних у реальному часі, що дозволяє знижувати затримки у виявленні загроз і зменшувати час реагування. Такі методи особливо корисні для виявлення складних атак, наприклад, тих, що використовують методи соціальної інженерії або вектори «нульового дня», де традиційні методи виявлення можуть бути неефективними.

Завдяки постійному навчанню та адаптації, алгоритми DL не тільки покращують існуючі системи безпеки, а й дозволяють передбачати нові загрози, ще до того як вони проявляться в мережі. Це робить їх незамінними інструментами, наприклад, для систем цифрової розвідки, особливо в умовах складних і швидко змінюваних кіберзагроз, з якими стикаються організації та держави під час гібридних воєн чи масштабних кібератак [11 – 15].

Таким чином, інтеграція алгоритмів ML на всіх етапах – від запобігання до розслідування – створює фундамент для динамічної та адаптивної системи кіберзахисту, що значно підвищує ефективність реагування на сучасні кіберзагрози.

# РОЗДІЛ 2 ІНТЕГРАЦІЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ АВТОМАТИЗАЦІЇ ЦИФРОВОЇ РОЗВІДКИ У ВИЯВЛЕННІ ТА АНАЛІЗІ КІБЕРЗАГРОЗ

## 2.1 Теоретичні основи та методологічні принципи цифрової розвідки у виявленні кіберзагроз

Цифрова розвідка (Digital Intelligence, DI) – це процес збору, аналізу та інтерпретації інформації з цифрових джерел для виявлення та оцінки загроз у кіберпросторі. DI включає різноманітні інструменти та методи, які дозволяють отримувати дані з відкритих і закритих джерел, аналізувати їх та приймати рішення щодо кібербезпеки на основі отриманих результатів. DI опрацьовує широкий спектр інформації, що може варіюватися від даних з Інтернет-ресурсів та соціальних мереж до аналізу геопросторових даних і поведінкових моделей користувачів у мережі [14].

Застосування методів DI у виявленні кіберзагроз є особливо важливим в умовах активної фази гібридної війни, яку зараз переживає Україна [12; 16]. Кіберзлочинність виступає одним із ключових інструментів гібридної війни, що використовуються для підриву національної безпеки та створення нестабільності в Україні [17]. Атаки на критичну інфраструктуру, маніпулювання суспільною свідомістю через соціальні мережі, злам інформаційних ресурсів та крадіжка конфіденційної інформації – все це є складовими сучасних кіберзагроз, які вимагають ефективних методів виявлення та нейтралізації. Використання методів DI дозволяє не лише вчасно ідентифікувати ці загрози, але й аналізувати відповідні джерела, методи та потенційний вплив. Це дає змогу приймати обґрунтовані рішення, які сприяють забезпеченню кібербезпеки та захисту стратегічної інформації, а також підвищують стійкість держави до кібернетичних атак. Таким чином, DI стає невід’ємною складовою сучасних стратегій протидії гібридним загрозам у цифровому середовищі.

До джерел інформації у DI ми відносимо:

- відкриті джерела [11; 18], тобто інформацію, що доступна у відкритому доступі через Інтернет. Наприклад, новинні ресурси, блоги, публічні форуми, соціальні мережі, державні реєстри та бази даних, наукові статті та інші публікації. OSINT є базовим компонентом DI, оскільки дозволяє отримати велику кількість даних без порушення законодавства;

- закриті джерела, тобто дані, доступ до яких обмежений або контролюється (наприклад, внутрішні бази даних організацій, інформація про мережевий трафік, логи серверів та інші внутрішні джерела, які вимагають спеціального доступу).

Процес збору та аналізу інформації із вказаних джерел зазвичай включає етапи, представлені на рис. 2.1.


Рисунок 2.1 – Процес збору та аналізу джерел у цифровій розвідці

Наприклад, є автоматизовані системи для збору даних з соціальних мереж, які аналізують ключові слова або обговорення, що можуть вказувати на потенційні загрози. Аналіз даних з використанням аналітичних методів та інструментів, алгоритмів ML, статистичного аналізу, візуалізації даних тощо необхідний для виявлення закономірностей, аномалій та інших ознак кіберзагроз. Інтерпретація результатів фахівцями, важлива для визначення

потенційних загроз та надання рекомендацій щодо подальших дій для забезпечення кібербезпеки.

До основних методів DI, які більш детально дослідимо у подальших пунктах даної наукової роботи, ми відносимо:

- OSINT (Open Source Intelligence) [11], який дозволяє збирати інформацію з відкритих джерел, відстежувати активність кіберзлочинців, отримувати корисні інсайти щодо їхньої діяльності;

- SOCMINT (Social Media Intelligence), який фокусується на зборі та аналізі даних з соціальних мереж. Інструменти та методологія SOCMINT дозволяє відслідковувати обговорення, групи, коментарі та інші активності, які можуть бути пов’язані з кіберзагрозами, поширенням дезінформації або плануванням кібернападів;

- GEOINT (Geospatial Intelligence) [11; 18], суть якого полягає у активному досліженні та аналізі географічної інформації, такої як супутникові знімки та геолокаційні дані, для виявлення загроз, які можуть бути пов’язані з конкретними місцями або регіонами. Дана інформація є особливо корисною для виявлення атак, спрямованих на інфраструктурні об’єкти.

Менш популярними, але не менш важливими у певних галузях, є такі методи DI, як CYBINT (Cyber Intelligence), HUMINT (Human Intelligence) [19], TECHINT (Technical Intelligence) [20], SIGINT (Signals Intelligence) [21], MASINT (Measurement and Signature Intelligence) [22], FININT (Financial Intelligence) [23], IMINT (Imagery Intelligence) [24 – 25]. Більш детальна інформація по зазначеним видам DI представлена у дод. Б. Кожен з цих методів доповнює один одного, дозволяючи створювати комплексний підхід до виявлення кіберзагроз та управління кіберризиками. Використання комбінації методів забезпечує більш повний аналіз ситуації та допомагає швидше реагувати на загрози у цифровому середовищі.

Отже, сутність DI полягає в її здатності надавати актуальну та релевантну інформацію, необхідну для прийняття обґрунтованих рішень у сфері кібербезпеки. В гармонійному поєднанні вище зазначених методів, DI дозволяє проактивно виявляти загрози на основі аналізу поведінки користувачів та

зловмисників у мережі; оцінювати ризики для інформаційних систем та мереж, враховуючи потенційні вектори атак; сприяти розробці стратегій реагування на інциденти та вдосконалювати наявні системи захисту від загроз. DI також сприяє оптимізації ресурсів організацій, які займаються кібербезпекою, оскільки дозволяє зосередити зусилля на найбільш актуальних загрозах [11; 17; 26]. Використання автоматизованих інструментів та систем аналізу дає змогу суттєво зменшити обсяг рутинної роботи та сконцентруватися на стратегічному аналізі [27; 28].

Інтенсивні дослідження в галузі DI створюють підґрунтя для нових підходів до моніторингу кіберзагроз, які включають в себе використання ШІ, ML та великих даних. Вони дозволяють автоматизувати процеси збору інформації та ідентифікації аномалій, зокрема з відкритих джерел, соціальних мереж і Dark Web-ресурсів. Даний процес сприяє швидкій реакції на загрози та знижує ризики для кібербезпеки організацій, оскільки дослідники постійно вдосконалюють інструменти для роботи в умовах мінливих кіберризиків.

Актуальність вивчення сучасного стану наукових досліджень у DI також пов’язана з необхідністю розробки нових підходів до захисту приватної інформації та конфіденційних даних. У світлі посилення регулювання кіберпростору та підвищення рівня цифрової загрози, для користувачів важливо впроваджувати інноваційні методи захисту, що базуються на науково обґрунтованих рішеннях. Таким чином, DI стає необхідним інструментом для забезпечення стійкості систем кібербезпеки в умовах зростаючих викликів інформаційного суспільства.

Використанню сучасних інструментів DI присвячені наукові дослідження вітчизняних науковців Живила Є.О. [29], Дамяна М.Ю. [29], Сторчака А.С. [30], Самойлова І.В. [30] та ін. Дослідники підкреслюють, що ефективне застосування DI не лише сприяє оперативному реагуванню на загрози, а й забезпечує систематичний підхід до моніторингу інформаційного середовища.

На особливості використання технології OSINT для збору, узагальнення та аналізу інформації на основі різних соціальних мереж акцентує увагу в своїх наукових дослідженнях Горбач М. [31]. Науковець підкреслює важливість

технології OSINT для збору відкритих даних, що дозволяє виявляти приховані зв’язки між користувачами, аналізувати контент і визначати потенційні загрози. Горбач М. [31] відзначає, що ефективне використання OSINT-інструментів дає змогу отримати достовірну інформацію з різних платформ, таких як Twitter (мережа X), Facebook, Instagram, і здійснювати якісний аналіз з урахуванням поведінкових закономірностей, геолокаційних даних та трендів, що суттєво підвищує ефективність аналітичних досліджень.

Горбач М. описує інструмент, який він створив для відстеження активності користувачів у соціальних мережах та проведення OSINT розвідки. Цей інструмент, написаний мовою програмування Python, використовує API сервісу Social Searcher для пошуку користувачів за електронною адресою. Горбач М. також зазначає, що для OSINT розвідки використовувалися такі соціальні мережі, як Facebook, Twitter (мережа X), Instagram, LinkedIn, а також інструменти Google Dorking, PimEyes, FaceCheck, Seon та Have I Been Pwned..

Окремої уваги заслуговують дослідження використання DI як засобу боротьби в російсько-українській війні [31 – 35]. Науковці Рябцева Д.О., Гудзь Т.І., Цуркан І.О., Шимко А.О., Верзілов М.Р., Стецик Р.М., Комісарчук Ю.А., Черевко В.В., Віннічук Б.В., Стрелков В.В. [31 – 35] у своїх наукових роботах зазначають, що DI дозволяє оперативно збирати та аналізувати інформацію про військові дії, переміщення техніки та особового складу, а також виявляти дезінформацію, що активно використовується ворогом. Вони підкреслюють важливість інтеграції інструментів OSINT у стратегічне планування та прийняття рішень, що суттєво підвищує ефективність інформаційної безпеки України в умовах війни.

Низка вітчизняних та зарубіжних науковців й фахівців наголошують на важливості застосування методів DI для кібербезпеки [36 – 43], та на тому, що цей процес набуває особливого значення в умовах зростання обсягів і складності кібератак. DI виступає важливим інструментом для моніторингу, аналізу та запобігання кіберзагрозам, оскільки вона дозволяє виявляти потенційні загрози на ранніх етапах завдяки збору та обробці даних з відкритих джерел. Через розширення можливостей обробки великих даних і ШІ, DI не лише збільшує

швидкість реагування на загрози, але й підвищує точність ідентифікації потенційних ризиків. Наукові дослідження показують, що комплексний підхід до DI дозволяє кіберфахівцям здійснювати моніторинг кіберпростору в реальному часі, швидко виявляти аномалії, аналізувати джерела атак і оцінювати їхній вплив на інфраструктуру. Таке стратегічне застосування DI не тільки захищає критичну інфраструктуру, але й допомагає організаціям розробляти комплексні плани кіберзахисту та зміцнювати інформаційну безпеку, що дозволяє створити екосистему захисту, поєднуючи технології і методи, здатні забезпечити повноцінний контроль над інформаційними потоками.

Так, Запорожченко М. [38] пояснює етапи життєвого циклу кібератаки та підкреслює необхідність використання інструментів OSINT під час аудиту інформаційної безпеки організації та проведення тестів на проникнення. Науковець зазначає, що застосування методів та інструментів OSINT при проведенні тестів на проникнення дозволяє захистити будь-яку галузь, підвищити анонімність в Інтернеті, захистити співробітників компанії від соціальної інженерії тощо. Знання про те, яка інформація про компанію, що створює ризики для її інформаційної безпеки, доступна у відкритих джерелах, є важливим етапом для ефективного впровадження інструментів OSINT, що дозволяє не лише посилити заходи захисту на етапі розвідки, а й мінімізувати ймовірність успішного проведення кібератак.

Пучков О., Ланде Д., Субач І., Болюх М. та Нагоний Д. [39] представляють методику розслідування та прогнозування кіберінцидентів, засновану на використанні відкритих джерел інформації та безкоштовного програмного забезпечення з відкритим кодом. Ця методика відноситься до OSINT і базується на моніторингу сучасного Інтернет-простору, аналізі великих даних (Big Data), складних мереж (Complex Networks) і добуванні знань з текстових масивів (Text Mining). Науковці детально описують компоненти технології, такі як виявлення ключових слів (NLTK), понять (SpaCy, NLP), а також системи візуалізації та аналізу графів. Основна ідея методики полягає в застосуванні методів збирання даних через глобальні пошукові системи, агрегування інформаційних потоків та інтелектуального аналізу зібраних даних.

Методика включає функції збору релевантної інформації з визначених ресурсів, автоматичного сканування та первинної обробки інформації з вебсайтів, формування повнотекстових масивів інформації, аналізу текстових повідомлень, визначення їх тональності, створення аналітичних звітів, інтеграції з географічними інформаційними системами, візуалізації інформації, дослідження динаміки інформаційних потоків та прогнозування подій на основі аналізу динаміки публікацій.

Також методика передбачає інструменти для графічного представлення динаміки даних, відображення кількості повідомлень за добу, перегляд сюжетів з повідомлень про кіберінциденти та кластерів, згрупованих за алгоритмом кластерного аналізу. Вона дозволяє формувати мережі з концептів, що відповідають персонам, організаціям та інформаційним джерелам, для вивчення взаємозв'язків між ними. Серед робіт закордонних науковців щодо використання DI для забезпечення кібербезпеки, виділяємо дослідження Рібе Т., Бізеллі Т., Кауфхольд М.-А. Ройтер К., Галіндо Х. [41 – 43].

Так, німецький науковець Рібе Т. [41] зазначає, що використання DI для моніторингу та виявлення загроз кібербезпеці набуває популярності серед багатьох організацій й особливу актуальність даного процесу для моніторингу кіберзагроз для критично важливих інфраструктурних служб. Науковець наголошує, що більшість систем використовують загальнодоступні дані, зібрані в соціальних мережах та інших відкритих джерелах. Таким чином, прийняття OSINT-систем користувачами, а також умови, які впливають на прийняття, є актуальними для розробки OSINT-систем для кібербезпеки. Рібе Т., Бізеллі Т., Кауфхольд М.-А. та Ройтер К. [42] наводять результати опитування (N=1093), під час якого запитали учасників їх сприйняття OSINT-систем, а також про їхні проблеми щодо конфіденційності. Дослідники надають рекомендації для подальших досліджень та використання систем OSINT для кібербезпеки органами влади. Оскільки OSINT є радше методологією, аніж окремою технологією, підходи можуть обиратися та комбінуватися, щоб дотримуватися принципів мінімізації даних та анонімізації, а також використовувати інновації в галузі обчислень із захистом конфіденційності та ML.

Іспанський вчений Галіндо Х. [43] у своєму дисертаційному дослідженні обґрунтовує та демонструє OSINT як референсну методологію, яка має доповнити сучасні та майбутні рішення з кібербезпеки й стратегії кіберзахисту на національному і міжнародному рівнях. Науковець виділяє 4 завдання:

1) критичний аналіз OSINT у світлі цифрової революції та розвитку Big Data й ШІ;

2) класифікацію ризиків безпеки та конфіденційності, пов'язаних із використанням OSINT;

3) практичні випадки застосування OSINT для протидії загрозам, зокрема через соціальні мережі;

4) дослідження Dark Net з точки зору OSINT, ідентифікацію та аналіз методів виявлення Tor-адрес.

Для досягнення цих цілей Галіндо Х. [43] наводить методологію, що починається з огляду літератури поточного стану OSINT, його застосування та викликів. Наступним етапом є аналіз потенційних загроз від використання OSINT у кіберзлочинах. Далі проводиться практичне дослідження у соціальних мережах (науковець наводить приклад Twitter та вибори у Іспанії 2019 року для виявлення соціальних ботів). Останній етап спрямований на дослідження Dark Net для ідентифікації анонімних сайтів. Галіндо Х. [43] описує високу ефективність OSINT як інструменту для вирішення проблем у різних сферах. У добу Big Data та ШІ OSINT допомагає здобувати дані з таких джерел, як соціальні мережі, онлайн-документи, веб-сторінки й навіть Dark Net. Використання SOCMINT дозволяє аналізувати соціальних ботів і вплив дезінформації, тоді як DARKINT сприяє ідентифікації анонімних сайтів у Dark Net.

Узагальнюючи проведене дослідження щодо теоретичних аспектів DI у сфері виявлення кіберзагроз, можна зробити висновок, що сучасний стан наукових досліджень у галузі DI та її застосування для забезпечення кібербезпеки свідчить про необхідність розвитку нових підходів та інструментів. В умовах зростаючих кіберзагроз, DI стала критично важливою складовою комплексної системи захисту. Дослідження в цій сфері акцентують на

автоматизації збору й аналізу даних, що надходять з відкритих джерел, соціальних мереж, Dark Web та інших платформ, завдяки чому вдається виявляти загрози на ранніх стадіях.

Залучення ШІ, ML та технологій обробки великих даних значно підвищує ефективність DI, зокрема у контексті моніторингу, прогнозування та нейтралізації кіберзагроз. Інтеграція інструментів OSINT у процеси кібербезпеки сприяє швидкому реагуванню на інциденти та підтримці стійкості критичної інфраструктури. Водночас в умовах воєнного стану DI відіграє важливу роль у забезпеченні інформаційної безпеки та боротьбі з дезінформацією.

Таким чином, з огляду на динамічність і складність кіберсередовища, DI розвивається в напрямку побудови адаптивних, автоматизованих та інтелектуальних систем захисту, що дозволяють вчасно реагувати на нові загрози. Вона є важливою складовою сучасної кібербезпеки та надає можливість створювати стійку екосистему захисту, яка сприятиме зниженню ризиків та підвищенню безпеки у цифровому світі.

## 2.2 Методологічні підходи та принципи цифрової розвідки до ідентифікації та аналізу кіберзагроз

DI є одним із ключових напрямів у сучасній кібербезпеці, спрямованим на збір, обробку та аналіз даних з відкритих джерел для виявлення та запобігання кіберзагрозам. Основою DI є використання широкого спектра джерел, таких як соціальні мережі, Інтернет-сайти, блоги, форуми, різного роду бази даних тощо. Основна мета DI у кібербезпеці – це забезпечення об’єктивної інформації для виявлення загроз, прогнозування потенційних атак і мінімізації ризиків у кіберпросторі.

Методологія DI є науково обґрунтованим підходом, який передбачає систематичне застосування технологій і методів для збору, аналізу та інтерпретації даних, що можуть використовуватися для виявлення, прогнозування та попередження загроз у цифровому середовищі. Як

міждисциплінарна галузь, DI поєднує принципи кібербезпеки, інформаційних технологій, криміналістики, психології, соціальних наук і управління ризиками.

Методологічно DI базується на принципах збору та аналізу даних, кореляції між різними типами інформації та побудові профілів кіберзагроз. У сучасних умовах, коли обсяг даних постійно зростає, використання технологій ШІ та ML є важливим аспектом DI, оскільки саме ці технології здатні обробляти великі масиви даних, виявляти аномалії та ідентифікувати потенційно небезпечні закономірності.

Методологія DI включає 7 основних етапів, які детально описані в дод. В, а саме: постановку мети та завдань, збір даних, попередню обробку даних, аналіз даних, візуалізацію та інтерпретацію, прогнозування та рекомендації, оцінку ефективності та оптимізацію.

Для ефективного виконання задач DI використовуються спеціалізовані програмні засоби. Серед них виділяємо ПЗ, представлені у табл. 2.1:

Таблиця 2.1

| *Спеціалізований програмний засіб* | *Призначення* |
|---|---|
| Maltego | платформу для збору, аналізу та візуалізації розвідувальної інформації, яка дозволяє відстежувати зв'язки між різними об'єктами (персонами, організаціями, IP-адресами тощо). |
| Shodan | пошукову система для знаходження підключених до Інтернету пристроїв та вразливостей у них |
| Recon-ng | інструмент, призначений для OSINT розвідки, що дозволяє автоматизувати збір інформації з відкритих джерел |
| SpiderFoot | інструмент для автоматизації DI, який обробляє великі масиви даних та виявляє потенційні кіберзагрози |
| FOCA | інструмент для аналізу метаданих файлів, які можуть містити конфіденційну інформацію, таку як IP-адреси або дані користувачів |

Більш детально робота із зазначеними спеціалізованими програмними засобами представлена нами у публікаціях [11; 45; 47 – 50].

Сучасний довідник інформації про різні джерела та ресурси, які використовуються для відкритого збору даних представлений у «Компендіумі OSINT-інструментів» [46]. Даний ресурс створила команда Victory Drones (БФ «Дігнітас») за підтримки Європейського Союзу та Міжнародного Фонду

«Відродження» в рамках спільної ініціативи «Європейського Відродження України». Проєкт Victory Drones був створений у 2022 році та займається технологічною освітою цивільних та військових, акцентуючи увагу на технологічній мілітаризації українського суспільства. Ресурс [46] містить інформацію про актуальні до використанні в Україні інструменти для проведення DI. Розробники зазначають платформи інструментів з DI, надають стислий опис основних можливостей, вартість інструментів, а також лаконічно зазначають переваги та недоліки. Для пошуку інформації засобами DI рекомендовано до використання: Amnesty YouTubeDataviever, Geogramint, HowToFind_Bot, Intelligence X`s TG, InVID verification plugin, Paliscopes`s Discovery, Social Searcher, TelegramDB, Telemetrio, Telepathy, TGstart, TwitterAudit, Twitter Advanced Search (in-built), YouTube Geofind, YouTubeMetadata та ін.

До зручних інструментів з пошуку людей відносять: AnyMailFinder, Artellence, Betaface, CallApp, CyberSec Search, DataLeak Bot, Email finder leeadstall, Face Check.ID, Findclone, GetContact, OSINT Industries, Paliscope`s Yose, PimEyes, Rocket Reach Search, Search4Faces, Social Catfish, X-Ray.

Окремо виділені інструменти по роботі із супутниками та мапами: Airbus, Bellingcat OpenStreet Map Search, Bellingcat`s map on Ukraine, Bing Maps, CIR`s map on Ukraine, Copernicus Open Access Hub, Dual Maps, EOS Landviewer, F4Map, Geo-confirmed, GeoSpy AI, Google Earth Pro, Image Hunter, LiveUAMap, Map of attacks on healthcare of Ukraine, Maxar, OpenAerielMap, OpenInfraMap, Overpass Turbo, Photo-Map, Planet, Satellogic, Sentitel-2, Shadow Map, Soar.Earth, SunCalc, Umbra, Wikimapia, World.

Систематизовано ресурси для анонімізації особи, яка здійснює цифрову розвідку: ClearVPN, HTTPS guide та ін.

Окрема увага приділяється роботі з архівними даними за допомогою інструментів: Archive Today, Archiving YouTube videos, Bellingcat`s guide, Hunchly, Twitter Video Downloader, Wayback Machine, WITNESS та ін.

Зважаючи на складні умови сьогодення, обумовлені повномасштабним вторгненням рф та територію нашої держави, важливими і актуальними є

рекомендації щодо інструментів пошуку підрозділів армії рф та інформації щодо окремих осіб: InformNapalm`s Map, JominiWest, Oryx, Pouletvolant, Russo-Ukrainian Warspotting, Scramble Dutch Aviation Society, UAControl Map, Книга катів українського народу, Реєстр російських військових від ГУР, Російські воєнні злочинці та ін.

Інструменти, представлені в «Компендіумі OSINT-інструментів» [46], є більш адаптованими до сучасних реалій України та зручнішими для місцевих користувачів в порівнянні з OSINT Framework. Вони враховують регіональні особливості інформаційного простору України, що робить їх особливо корисними в умовах інформаційних викликів сьогодення. На відміну від OSINT Framework, який є загальним ресурсом, компендіум акцентує увагу на практичних прикладах та локалізованих методах для розслідувань і моніторингу, що підвищує ефективність використання інструментів в українському контексті.

В останні роки значна увага приділяється інтеграції ШІ у процеси DI. ШІ використовується для автоматизації збору та аналізу великих обсягів даних, що дозволяє швидше ідентифікувати кіберзагрози та аномалії. Наприклад, алгоритми ML можуть аналізувати мережевий трафік, виявляти аномальні закономірності та прогнозувати можливі атаки. Такі підходи підвищують ефективність та швидкість розвідки, знижуючи ризики, пов’язані з затримками в реагуванні на загрози.

Загострення кіберзагроз [17; 26] і збільшення кількості атак на державні та приватні організації підкреслюють важливість DI як інструмента проактивного захисту. DI дозволяє передбачити можливі загрози, підготуватися до них заздалегідь та мінімізувати потенційні ризики. Застосування сучасних методів та інструментів DI є невід’ємною частиною створення надійних систем захисту даних і забезпечення безпеки інформаційних систем в умовах цифрової трансформації.

## 2.3 Розробка алгоритмів обробки та аналізу даних із відкритих джерел засобами штучного інтелекту для ідентифікації потенційних кіберзагроз

Аналіз даних із відкритих джерел для виявлення загроз є важливою складовою сучасної кібербезпеки. Сьогодні загрози стають дедалі витонченішими й здебільшого обговорюються на форумах, в соціальних мережах або навіть ховаються у вигляді невеликих згадок в новинних статтях. У таких умовах завдання пошуку та аналізу цих сигналів стає комплексним, а ефективні алгоритми – незамінними.

В основі процесу обробки відкритих даних лежить динаміка та гнучкість алгоритмів, які дозволяють швидко збирати, відсіювати та аналізувати великі обсяги інформації. Наприклад, автоматизовані інструменти, що «скрейплять» вебсайти чи застосовують API для доступу до соціальних мереж, дають змогу витягувати дані на різних рівнях – від коротких згадок до поглиблених тематичних обговорень. Але сам збір даних – це лише початок. Великий масив інформації зазвичай містить безліч «шуму», що компенсується алгоритмами фільтрації. Їх завдання – знайти точку балансу між видаленням нерелевантних даних і збереженням потрібного контексту.

Далі, після фільтрації, йде ще більш складний процес класифікації та аналізу. Дослідники з’ясовують, як групувати та класифікувати інформацію, використовуючи, наприклад, техніки ML [40 – 42], які дають змогу зрозуміти, чи є певна активність простою інформаційною хвилею або потенційною загрозою. Даний процес може варіюватися від аналізу специфічних ключових слів до виявлення аномальної поведінки в інформаційних потоках.

Ще один важливий момент – пріоритезація інформації, адже не всі виявлені загрози становлять однакову небезпеку. Тут критично важливим є вміння «зрізати» зайве, фокусуючись на потенційно найбільш небезпечних ситуаціях, і одночасно не пропустити можливий сигнал тривоги. Алгоритми на цьому етапі працюють над тим, щоб формувати «рівні загрози», що дозволяє командам безпеки оперативно спрямовувати свої зусилля на дійсно важливі аспекти.

Інший ключовий аспект – перевірка інформації на достовірність. В епоху фейкових новин та масових дезінформаційних кампаній важливо розрізняти факти та фікцію. Алгоритми повинні вміти не тільки ідентифікувати джерело, а й оцінювати його надійність, враховуючи можливі упередження, маніпуляції чи дезінформацію. Тут часто застосовуються комбінації ML з аналітичними інструментами, що перевіряють контекст згадок та авторитетність джерела.

Отже, алгоритми обробки даних із відкритих джерел – це комплексний, багатогранний інструмент, що вимагає ретельного підходу та виваженого налаштування. Ефективність цих алгоритмів залежить не тільки від їхньої точності, а й від здатності враховувати зміни в інформаційних потоках і адаптуватися до нових типів загроз. Тому подальші дослідження в цій галузі будуть спрямовані на вдосконалення адаптивності та масштабованості алгоритмів, що дозволить їм ще швидше та точніше реагувати на загрози.

Автоматизація процесів DI сьогодні перетворюється на невід'ємну частину стратегії забезпечення безпеки програмного забезпечення. Цей підхід дозволяє значно прискорити реагування на потенційні загрози та дає змогу досліджувати інформацію з відкритих джерел практично в реальному часі. Автоматизовані алгоритми аналізують великі обсяги даних з різноманітних платформ, від соціальних мереж до форумів і навіть новинних ресурсів, що значно зменшує ризик пропустити важливі сигнали про кіберзагрози.

Наприклад, алгоритми, які використовуються для OSINT, мають вбудовані механізми збору даних з відкритих джерел за допомогою веб-скрейпінгу або API-запитів. Вони часто включають налаштування для фільтрації контенту за ключовими словами та обмеженнями за географічною зоною чи мовою. Це дозволяє зосередити увагу саме на тій інформації, яка є релевантною для певного типу загрози.

SOCMINT алгоритми, у свою чергу, створені для аналізу текстових даних та соціальної активності, спираючись на NLP для виділення потенційних загрозливих висловлювань або аномальної поведінки. Основна ідея таких алгоритмів полягає в тому, щоб розпізнати аномалії або підозрілі закономірності в соціальній активності користувачів. NLP при цьому дозволяє автоматично

виділяти й інтерпретувати семантику тексту, емоційні забарвлення, а також виявляти приховані зв’язки.

Алгоритм обробки та аналізу даних із відкритих джерел за допомогою SOCMINT та NLP складається з наступних кроків:

1. SOCMINT-алгоритми починають роботу з автоматизованого збору даних, використовуючи API платформ соціальних мереж (наприклад, мережа X, API, Reddit API) або методи веб-скрейпінгу для отримання доступу до відкритих даних. Джерела можуть включати твіти, коментарі, публікації, відповіді на форумах, блоги та навіть відеоконтент (у вигляді текстових описів чи підписів). На цьому етапі важливо забезпечити комплексність вибірки, щоб вона містила достатню кількість даних з різних типів джерел для повнішого аналізу загроз. Наприклад, коментарі до новинних статей можуть вказувати на певні події чи настрої, пов’язані з потенційними загрозами, які будуть корисними для SOCMINT-аналізу.

2. Зібрані дані «очищуються» й готуються до подальшої обробки. Основні етапи попередньої обробки включають: усунення зайвих символів, стоп-слів, спеціальних символів, HTML-тегів, щоб уникнути перевантаження алгоритмів непотрібною інформацією; приведення тексту до єдиної форми (наприклад, до нижнього регістру) для уніфікації та покращення точності обробки; розбивка тексту на окремі слова чи фрази (токени), що полегшує обробку; зведення слів до їх базової форми для подальшого точнішого аналізу. Наприклад, «атаки» стають «атака», що дозволяє алгоритмам ефективніше знаходити зв’язки між схожими термінами.

3. Крок аналізу семантики та емоцій, коли на основі підготовлених текстових даних SOCMINT-алгоритми використовують NLP для виявлення тональності висловлювань. Тональність тексту може допомогти визначити потенційні загрозливі висловлювання, зокрема: негативна тональність може свідчити про виражену неприязнь, обурення або гнів, що часто супроводжують загрозливі висловлювання; сильний емоційний відгук чи тривожний контент можуть вказувати на події чи ситуації, пов’язані з потенційними загрозами; аналіз емоцій також може включати визначення таких настроїв, як страх, агресія

або паніка, що є індикаторами аномальної поведінки або спроб соціального впливу.

4. Виявлення загрозливих фраз та закономірностей, коли SOCMINT-алгоритми використовують методи ML, зокрема класифікацію на основі попередньо навчених моделей. Моделі можуть бути навчені на спеціально зібраних наборах даних, що містять приклади загрозливих фраз. Наприклад: ключові слова або фрази, пов'язані з небезпекою, такими як «вибух», «злочин», «атака», допомагають автоматично виявити текст з потенційно шкідливим наміром. Набір предиктивних моделей, таких як Support Vector Machines (SVM), нейронні мережі або Decision Trees, дозволяє класифікувати виявлені фрази як такі, що становлять ризик або є нейтральними. Більш просунуті моделі, наприклад, глибокі нейронні мережі (DNN) або рекурентні нейронні мережі (RNN), здатні виявляти складніші закономірності, враховуючи попередні висловлювання та події в контексті конкретної соціальної мережі.

5. Аналіз соціальної поведінки та виявлення аномалій. Наприклад, аномальне зростання активності користувача може сигналізувати про підготовку до масової кампанії, такої як дезінформація чи соціальна інженерія; зміна в типових темах чи тоні постів користувача може свідчити про потенційне залучення до загрозливих дій або маніпуляцій.

Для цього алгоритми використовують кластеризацію поведінки на основі нормальних закономірностей та порівняння з новими даними. Наприклад, модель може відслідковувати середню кількість постів користувача, тематичний діапазон або частоту взаємодії з певними темами. Якщо спостерігається значне відхилення, система може ідентифікувати активність як підозрілу.

6. Після збору та обробки даних SOCMINT-алгоритми можуть представляти результати у вигляді графіків, таблиць або теплових карт. Наочність дозволяє аналітикам кібербезпеки швидко ідентифікувати потенційні ризики, локалізувати джерела загроз і прийняти відповідні заходи.

Розглянемо прикладний сценарій з виявлення терористичних загроз. Наприклад, SOCMINT-алгоритм може здійснювати моніторинг соціальних мереж на предмет згадок конкретних ключових слів (наприклад, «бомба»,

«атака»), і в разі високої негативної тональності та тривожної поведінки ініціювати тривогу. Виявлені закономірності допоможуть ідентифікувати учасників, час і можливі цілі загрози, що надає фахівцям змогу швидше та точніше реагувати на ситуацію.

Таким чином, автоматизовані процеси SOCMINT дозволяють здійснювати ефективний моніторинг і реагування на загрози шляхом аналізу величезних обсягів соціальних даних. Даний процес забезпечує потужний інструмент для швидкого реагування на кіберзагрози та зниження ризиків, пов’язаних із деструктивною поведінкою в мережі. Такі алгоритми не тільки покращують швидкість та ефективність роботи з великими масивами даних, але й дозволяють оперативно виявляти нові загрози, зокрема ті, що можуть поширюватися в реальному часі. Іншими словами, автоматизація дозволяє DI не лише відповідати на загрози, але й проактивно ідентифікувати їх на ранніх етапах, що, в підсумку, значно підвищує рівень загальної безпеки програмного забезпечення.

Інтеграція алгоритмів DI у системи моніторингу кібербезпеки вимагає як адаптації сучасних методів обробки даних, так і належного врахування особливостей кожної з систем. Це процес, який має на меті глибше злиття автоматизованих аналітичних можливостей з конкретними безпековими потребами організацій, що прагнуть підвищити свої рівні захисту. Зараз DI розглядається не просто як набір окремих алгоритмів, але як комплексний, інтегрований підхід до виявлення загроз, в якому кожна частина взаємодіє з іншими, створюючи ефективну систему попередження та виявлення загроз.

Перший крок у цьому напрямі – забезпечення тісного взаємозв’язку між SOCMINT, GEOINT, OSINT, і навіть SIGINT-алгоритмами, щоб комплексно покривати різні рівні аналізу. SOCMINT здатен фільтрувати текстові дані та визначати тональність у спільнотах, GEOINT допомагає прив’язати інформацію до географічних локацій, а OSINT відкриває доступ до загальнодоступних веб-ресурсів, що разом дозволяє більш точно оцінити потенціал загрози. Ключовим аспектом є створення єдиної платформи, яка може збирати й автоматично агрегувати дані з різних джерел, щоб уникати дублювання та неузгодженостей, які могли б зашкодити оперативності рішень.

Інший важливий крок полягає у розвитку та впровадженні алгоритмів ML, орієнтованих на виявлення невідомих раніше загроз шляхом аналізу закономірностей, що ще не були описані в існуючих базах. Використовуючи алгоритми, що здатні адаптивно навчатися на реальних даних, інтеграція DI забезпечує не лише точність у розпізнаванні загроз, але й своєчасне оновлення та корекцію методів, що робить моніторинг більш живим та динамічним.

Третій момент полягає у тісній інтеграції з методами автоматизації попереджень та оповіщень, коли система не просто повідомляє про загрозу, а й генерує чіткі рекомендації щодо дій. Це можливо завдяки заздалегідь визначеним алгоритмам дій у відповідь на виявлені закономірності та аномалії. Наприклад, якщо виявлено підозрілу активність у соціальних мережах, яка може свідчити про сплановану атаку або деструктивну подію, SOCMINT-технологія може не лише попередити операторів, а й автоматично обмежити доступ до ключових ресурсів чи заборонити певні типи взаємодій. Це підвищує загальний рівень захищеності системи, зводячи до мінімуму можливість людської помилки в умовах інформаційного перенавантаження.

У довгостроковій перспективі, для вдосконалення інтеграції, важливо приділяти увагу й аспектам адаптивного налаштування під вимоги конкретного користувача або системи. Це означає, що алгоритми повинні враховувати специфіку кожного середовища, зокрема різні правові обмеження, політики конфіденційності та унікальні вимоги організацій. Наприклад, у галузі правоохоронних органів алгоритми можуть бути налаштовані для більш чутливого виявлення загроз безпеці у певних географічних районах або сферах діяльності, у той час як комерційні організації можуть вимагати більш орієнтованого на клієнта підходу.

Проаналізуємо конкретні рекомендації щодо інтеграції алгоритмів DI в системи моніторингу кібербезпеки.

Наприклад, для виявлення загрозливих тенденцій або підозрілих повідомлень у соціальних мережах, алгоритми NLP можуть бути налаштовані на виявлення специфічних ключових слів, фраз або закономірностей мови, що часто асоціюються з кібератаками або іншими загрозами. Якщо зловмисники

обговорюють цілі для атак або поширюють зловмисні інструкції, NLP може виділити ці повідомлення за допомогою моделей, які аналізують тональність, семантику та частоту використання певних слів. Впровадження моделей NLP, таких як BERT або GPT-4, дозволяє не лише ідентифікувати загрозливі фрази, але й розуміти контекст, що підвищує точність аналізу. Рекомендації можуть включати необхідність створення спеціальних «словників загроз» для різних категорій кібератак, які регулярно оновлюються на основі даних з останніх інцидентів. Така практика дасть змогу системі миттєво реагувати на нові загрози або тенденції, виявляючи навіть приховані або зашифровані повідомлення.

Методи OSINT можуть бути корисними для моніторингу витоків даних і несанкціонованого поширення інформації. Наприклад, за допомогою алгоритмів автоматизації можна налаштувати моніторинг Dark Web ресурсів для виявлення викрадених облікових даних, конфіденційних даних компанії або персональної інформації. Інструменти, такі як Shodan чи Censys, дозволяють сканувати відкриті порти та аналізувати інфраструктуру в Інтернеті, що дає змогу отримати додаткову інформацію про потенційні вразливості.

Рекомендації можуть включати розробку автоматичної системи, яка здатна сканувати визначені ресурси і виявляти підозрілу інформацію про організацію. Така система може регулярно перевіряти ринки DarkNet та форуми для виявлення витоків паролів чи фінансових даних. Окрім того, корисно використовувати API для швидкого доступу до даних з відкритих ресурсів.

GEOINT дозволяє аналізувати геопросторові дані, що можуть вказувати на географічну прив’язку кібератак. Так, при виявленні IP-адрес, з яких надходять потенційні загрози, можна використовувати сервіси геолокації, як-от MaxMind або IP2Location, для встановлення їхнього фізичного розташування. Це корисно для відстеження атак на інфраструктурні об’єкти, адже часто подібні атаки мають локальну прив’язку.

Рекомендації включають можливість інтегрувати функцію визначення геолокації в систему SOC, що дасть можливість швидко отримувати інформацію про географічне розташування джерел загроз. Це буде особливо актуально для

моніторингу інфраструктурних атак, де швидкість і точність локалізації має вирішальне значення.

Для виявлення аномальної активності можна застосувати алгоритми кластеризації або DL, що автоматично ідентифікують нетипові закономірності в мережевому трафіку. Зокрема, LSTM (Long Short-Term Memory) або GRU (Gated Recurrent Unit) мережі добре підходять для аналізу часових рядів, оскільки дозволяють передбачати, коли в мережі відбуваються незвичні збої або несанкціоновані спроби доступу. Рекомендація полягає у можливості налаштувати періодичне перенавчання моделей ML на нових даних, що дозволить адаптуватися до змін у поведінці трафіку. Додавання шарів постійного навчання та регулярне оновлення датасетів дозволить алгоритмам швидше пристосовуватися до змін у шаблонах атак.

З метою моніторингу соціальних мереж та відкритих ресурсів, доцільно створити алгоритм, що буде об’єднувати аналітику SOCMINT та OSINT. Об’єднання цих методів може допомогти відслідковувати взаємодії зловмисників на відкритих платформах та в соціальних мережах, виявляючи сигнали про майбутні атаки. Рекомендація полягає у можливості встановити регулярні запити до соціальних платформ та налаштувати фільтри для відбору підозрілих обговорень. Це дозволить створити додатковий рівень попередження перед тим, як потенційні загрози перетворяться у реальні атаки.

Загалом, інтеграція цих алгоритмів у систему кібербезпеки не лише розширює можливості моніторингу, але й дозволяє забезпечити більш оперативну реакцію на різні загрози. Комбінуючи інструменти аналізу даних, ML і багаторівневу аналітику, організація може досягти більш високого рівня захисту від кіберзагроз та мінімізувати ризики, пов’язані з потенційними атаками. Розробка алгоритмів обробки та аналізу даних із відкритих джерел із використанням ШІ є важливим етапом у зміцненні кібербезпеки й на загальнодержавному рівні. Ці алгоритми дозволяють автоматизувати виявлення потенційних загроз, забезпечуючи швидкість та точність, які неможливо досягти традиційними методами. Завдяки інтеграції методів ML, DL текстів, візуалізації даних та класифікації аномалій, сучасні системи можуть оперативно виявляти

підозрілу активність, ідентифікувати вразливі точки у мережах та передбачати можливі сценарії атак. Результати впровадження таких алгоритмів підвищують ефективність моніторингу кіберзагроз, скорочують час реагування на інциденти та мінімізують людський фактор у процесах аналізу великих обсягів інформації. Крім того, наведені алгоритми дозволяють адаптуватися до постійно змінюваних умов кіберсередовища, навчаючись на нових даних і вдосконалюючи моделі ідентифікації загроз. Це створює значні передумови для більш глибокого використання ШІ в системах кіберзахисту, забезпечуючи комплексний підхід до управління ризиками та підвищення загального рівня безпеки.

# РОЗДІЛ 3 АВТОМАТИЗАЦІЯ ПРОЦЕСІВ ВИЯВЛЕННЯ КІБЕРЗАГРОЗ З ВИКОРИСТАННЯМ АЛГОРИТМІВ ШТУЧНОГО ІНТЕЛЕКТУ

## 3.1 Удосконалення та автоматизація методів цифрової розвідки у виявленні кіберзагроз із застосуванням технологій штучного інтелекту

Оцінка ефективності методів DI у виявленні різних типів кіберзагроз передбачає аналіз результативності описаних у попередніх пунктах основних підходів, таких як OSINT, SOCMINT і GEOINT, кожен з яких має свої переваги та обмеження в ідентифікації певних видів загроз. Так, OSINT виступає потужним методом для виявлення загроз, оскільки використовує відкриті дані з численних онлайн-джерел, таких як веб-сайти, форуми, соціальні мережі, новинні сайти та урядові ресурси. Даний метод DI дозволяє виявляти загрози, пов’язані з витоками даних, витоком інформації з публічних ресурсів та моніторингом підозрілих активностей. Проте OSINT не охоплює закриті канали комунікації, зокрема зашифровані чати, де злочинці обговорюють можливі загрози, а також є вразливим до дезінформації та маніпуляцій.

В той же час SOCMINT є ефективним методом для виявлення загроз через соціальні мережі, дозволяє відстежувати активність потенційних зловмисників, зокрема за рахунок аналізу поведінкових закономірностей і ключових слів. Інструментарій даного методу дозволяє швидко зреагувати на події, що набувають популярності в соціумі, та відстежувати обговорення загрозливих подій або потенційних кібератак. Основний недолік SOCMINT вбачаємо у необхідності комплексної обробки великих масивів даних і ймовірні обмеження політикою конфіденційності, налаштуваннями приватності користувачів у соціальних мережах тощо.

GEOINT дозволяє виявляти та локалізувати кіберзагрози, пов’язані з фізичними об’єктами, такими як центри обробки даних або інфраструктура критичної важливості, а також моніторити геопросторові аномалії. Особливість даного методу полягає в тому, що він надає додатковий рівень контексту, допомагаючи відстежувати джерела загроз, їхню географічну прив’язку та

можливий вплив на критичні інфраструктури. Проте, GEOINT обмежений фізичним місцем розташування, а також може мати труднощі з ідентифікацією анонімних кіберзагроз, які не мають прив'язки до конкретної локації.

Загалом, кожен метод DI по-різному ефективний для певних типів кіберзагроз. OSINT і SOCMINT краще підходять для моніторингу відкритої інформації та поведінкових тенденцій, тоді як GEOINT ефективний для локалізації загроз, пов'язаних із фізичними об'єктами. Для досягнення високої ефективності у виявленні різноманітних кіберзагроз доцільно поєднувати ці методи, що забезпечує всебічний підхід і знижує ризик пропуску загроз. Поєднання методів DI у рамках комплексного підходу дозволяє підвищити ефективність у виявленні кіберзагроз за рахунок взаємного доповнення переваг кожного з методів. Наприклад, OSINT можна застосувати для збору загальної інформації про загрози та моніторингу трендів у відкритих джерелах, тоді як SOCMINT може надати більш детальні відомості про обговорення потенційних атак у соціальних мережах, а GEOINT забезпечує розуміння фізичної локалізації загроз. Об'єднання даних із різних джерел допомагає створити повну картину загроз і зменшити ймовірність пропуску важливої інформації. Також, завдяки інтеграції даних із SOCMINT і GEOINT у рамках систем моніторингу можливо швидше ідентифікувати локалізацію та джерело загроз, що сприяє оперативності у виявленні та протидії кіберінцидентам.

Застосування різних методів DI ефективно працює для різних категорій кіберзагроз. OSINT та SOCMINT здатні виявляти кампанії з фішингу та соціальної інженерії, аналізуючи тренди поширення шкідливих повідомлень та обговорення потенційних жертв у соціальних мережах. Висока ефективність обумовлена наявністю відповідної інформації у відкритому доступі. GEOINT ефективно використовується для виявлення загроз, які можуть бути фізично прив'язані до географічної локації, наприклад, несанкціонований доступ до центрів обробки даних або збої у критичній інфраструктурі. Поєднання GEOINT із OSINT може надати додатковий контекст щодо поточної ситуації в регіоні. OSINT зазвичай є найменш ефективним для виявлення анонімних кіберзагроз, таких як атаки з використанням ботнетів. У цьому випадку застосування

спеціалізованих інструментів кібербезпеки для аналізу аномальної активності є більш доцільним, хоча SOCMINT може допомогти ідентифікувати загрозу на ранніх стадіях, наприклад, через сплеск обговорень на тематику зловмисних дій.

У випадку атак на мережеву інфраструктуру, OSINT у поєднанні з аналітичними інструментами кібербезпеки дозволяє отримати доступ до інформації про типові методи зловмисників. SOCMINT також надає важливі дані, оскільки інколи зловмисники обговорюють свої дії на закритих форумах або у специфічних спільнотах.

Отже, ефективність методів DI виявляє значну важливість інтегрованого підходу, коли OSINT, SOCMINT, GEOINT та інші методи (див. дод. Б) використовуються комплексно. Комбінування цих методів дає можливість отримати глибші знання про кіберзагрози, забезпечує всебічний моніторинг потенційних ризиків і створює базу для більш точної, швидкої та ефективної реакції на кіберінциденти. Таке поєднання сприяє підвищенню рівня кібербезпеки за рахунок швидшого й ефективнішого виявлення загроз та мінімізації потенційних ризиків.

### 3.1.1 Удосконалення методів OSINT (Open Source Intelligence) з виявлення кіберзагроз за допомогою штучного інтелекту

OSINT є ключовим компонентом сучасної DI, який дозволяє збирати, аналізувати та використовувати інформацію з відкритих джерел для ідентифікації кіберзагроз. Завдяки відкритому характеру джерел, таких як відкриті реєстри, бази даних, форуми, реєстри доменів, новинні портали тощо, OSINT забезпечує аналітиків актуальною інформацією для виявлення нових загроз, оцінки їхнього впливу та прогнозування подальших сценаріїв.

OSINT дає змогу розробляти проактивний підхід до кібербезпеки та забезпечувати глибоке розуміння потенційних загроз.

Процес OSINT включає кілька послідовних етапів: ідентифікацію джерел даних, збір даних, аналіз і кореляцію даних, візуалізацію і створення звітів.

На етапі ідентифікації джерел даних обираються найбільш релевантні джерела інформації для конкретного завдання. Наприклад, для виявлення загроз

можна використовувати платформи з інформацією про вразливості (CVE), форуми про кібербезпеку, бази даних WHOIS для аналізу доменів, а також соціальні мережі для моніторингу активності та потенційних загроз. Інформація збирається автоматично за допомогою спеціалізованих інструментів, таких як Maltego, Shodan, Recon-ng, theHarvester, або вручну, що дозволяє аналізувати індивідуальні профілі, коментарі та інші цифрові сліди, які залишають кіберзловмисники.

На етапі аналізу та кореляції дані обробляються, фільтруються та піддаються аналізу з метою виявлення закономірностей, аномалій і можливих загроз. Різноманітні методи аналізу, такі як кореляція подій і поведінковий аналіз, дозволяють ідентифікувати потенційно небезпечні активності та виявляти зв’язки між різними об’єктами. Результати OSINT часто представляють у вигляді звітів з діаграмами, схемами та графіками, що дозволяє аналітикам і керівникам оперативно приймати рішення про запобігання загрозам.

Для ефективного виявлення кіберзагроз у рамках OSINT виділяємо наступні методи: соціальну інженерію, моніторинг соціальних мереж, аналіз метаданих, WHOIS-аналіз та аналіз мережевого трафіку.

Використання технік соціальної інженерії дозволяє збирати інформацію про організації, структуру їхніх мереж та можливі слабкі місця через публічні канали. Соціальна інженерія може також допомогти виявляти підозрілу поведінку користувачів і потенційні інсайдерські загрози. OSINT активно використовує методи SOCMINT (Social Media Intelligence) для аналізу інформації з соціальних мереж. Даний процес включає пошук потенційних загроз, ворожих настроїв, обговорень про нові вразливості, а також відстеження можливих витоків даних.

Не менш важливим є метод OSINT – аналіз метаданих. Метадані, що зберігаються в документах, фотографіях та інших файлах, є важливим джерелом інформації. Вони можуть містити дані про автора, час створення, координати зйомки, що може допомогти в ідентифікації місця чи осіб, причетних до інциденту.

WHOIS-аналіз використовується для вивчення інформації про власників доменів, історії реєстрації доменів, їхніх IP-адрес. Даний метод дозволяє ідентифікувати зловмисні сайти або підозрілі доменні зони, що можуть використовуватися для фішингу або атак.

OSINT інструменти можуть використовуватися для моніторингу та аналізу мережевого трафіку, зокрема за допомогою Shodan або Censys, які дозволяють виявляти вразливі пристрої, що підключені до Інтернету.

Серед сучасних інструментів для реалізації OSINT особливою популярністю користуються: Maltego (інструмент для графічного відображення інформації, що дозволяє виявляти зв'язки між об'єктами), Shodan (пошукова система для виявлення підключених до Інтернету пристроїв, включаючи сервери, маршрутизатори, камери спостереження тощо), Recon-ng (фреймворк для автоматизації збору інформації з відкритих джерел), theHarvester (інструмент для збору інформації про домени, електронні адреси, IP-адреси тощо).

Залучення технологій ШІ значно посилює можливості OSINT, дозволяючи автоматизувати рутинні процеси, виявляти складні закономірності та адаптуватися до постійно змінюваного кіберсередовища. До основних аспектів удосконалення OSINT за допомогою ШІ ми відносимо: оптимізацію збору даних із відкритих джерел, покращення аналізу метаданих і кореляції даних, інтеграцію прогнозної аналітики та модернізацію інструментів OSINT.

Застосування методів ML для автоматизації процесу ідентифікації релевантних джерел інформації знижує вплив людського фактора та збільшує охоплення даних. ШІ моделі, такі як NLP, можуть ефективно аналізувати великий обсяг текстової інформації з різних джерел, розпізнавати ключові загрози та формувати зв'язки між подіями. Наприклад, автоматичний моніторинг соціальних мереж за допомогою класифікаторів тексту дозволяє виявляти підозрілу активність, а тональний аналіз дописів визначає потенційно небезпечні обговорення.

Метадані у файлах (зображення, документи) містять приховану інформацію, яка може бути використана для ідентифікації кіберзловмисників або локацій атак. Алгоритми комп'ютерного зору, інтегровані в OSINT, здатні

обробляти візуальні дані, такі як зображення чи відео, для виявлення прихованих деталей.

ШІ також підвищує ефективність аналізу кореляцій між різними джерелами. Завдяки технологіям DL можна знаходити невидимі для людини зв’язки, наприклад, між різними IP-адресами, доменами або користувачами, що дозволяє виявляти багаторівневі схеми кіберзагроз.

ШІ дозволяє реалізовувати прогнозну аналітику у рамках OSINT. Наприклад, алгоритми на основі навчання з підкріпленням здатні аналізувати історичні дані та створювати сценарії розвитку загроз, що допомагає не лише реагувати на існуючі інциденти, а й попереджати нові.

Інструменти, такі як Maltego, Shodan та theHarvester, інтегровані з модулями ШІ, отримують нові функціональні можливості, зокрема, динамічне оновлення алгоритмів пошуку, автоматизовану обробку та класифікацію великих масивів даних.

Інтеграція ШІ у методи OSINT має низку викликів, зокрема необхідність забезпечення точності алгоритмів, дотримання законодавства про конфіденційність, а також боротьбу з дезінформацією. Водночас потенціал застосування ШІ в OSINT відкриває нові горизонти для проактивного захисту цифрових активів.

Таким чином, удосконалення OSINT із використанням ШІ сприяє підвищенню швидкості та ефективності виявлення кіберзагроз, що є критично важливим у сучасному світі динамічних загроз і постійно змінюваних технологій.

### 3.1.2. Автоматизація методів SOCMINT (Social Media Intelligence) для моніторингу соціальних мереж на предмет кіберзагроз

SOCMINT (Social Media Intelligence) або розвідка у соціальних мережах є одним із новітніх напрямів у DI, який забезпечує моніторинг та аналіз даних з відкритих соціальних платформ. Зважаючи на величезну кількість даних, що генеруються користувачами щодня, SOCMINT є важливим інструментом для

своєчасного виявлення кіберзагроз, таких як атаки соціальної інженерії, фішинг, витік конфіденційної інформації та потенційні атаки на бренди й організації.

SOCMINT надає можливість моніторити та аналізувати поведінку користувачів, публікації, коментарі та групові обговорення на різних соціальних платформах. Даний метод моніторингу соціальних мереж на предмет кіберзагроз дозволяє їх виявляти та попереджати. Зазвичай, це пов’язано із тим, що хакери або групи зловмисників діляться інформацією про свої наміри чи навіть інструменти для атак на Dark Net або в прихованих групах. Аналіз подібних дискусій допомагає виявляти потенційні загрози до того, як вони стануть активними. SOCMINT також дозволяє ідентифікувати фішингові кампанії та шахрайські сайти, які активно рекламуються в соціальних мережах.

SOCMINT використовує ряд методів для моніторингу та аналізу соціальних мереж, що дозволяє ефективно виявляти кіберзагрози через аналіз лексики та мови, виявлення аномалій, геоаналіз, візуальний аналіз

Використання алгоритмів аналізу природної мови NLP дозволяє розпізнавати підозрілі слова, фрази або жаргон, які можуть бути пов’язані з кіберзагрозами. Наприклад, слова на кшталт «dump», «exploit», «leak», «zero-day» різними мовами можуть свідчити про обговорення вразливостей або зломів. Алгоритми NLP також можуть аналізувати емоційне забарвлення повідомлень, виявляючи тривожні тональності або агресивні висловлювання, що можуть свідчити про підготовку до атак. Наприклад, виявлення слів чи фраз, що вказують на готовність до здійснення атак, як-от «launch», «attack», «payload», «target», може сигналізувати про можливу небезпеку та потребує негайної перевірки.

Іншим прикладом є використання NLP для аналізу метаданих в соціальних мережах з метою виявлення координованих дій. Коли однакові або схожі повідомлення, що містять слова, пов’язані з кіберзагрозами, повторюються з різних акаунтів або в різний час, це може свідчити про організовану кібератаку чи поширення ШПЗ. Наприклад, слова типу «botnet», «command and control» можуть вказувати на активність бот-мереж або інфраструктури зловмисників.

Завдяки NLP також можна автоматично визначати терміни, які мають особливе значення в певному контексті або регіоні, наприклад, розпізнавати специфічний сленг, який може вказувати на незаконні активності, включаючи слова на кшталт «carding» або «phishing». Подібний аналіз дозволяє системам бути більш адаптивними до нових типів загроз та враховувати культурні та мовні особливості кіберпростору.

За допомогою алгоритмів ML можна ідентифікувати аномальну поведінку, наприклад, різке зростання кількості згадок про певну компанію чи сервіс у негативному контексті, що може вказувати на інформаційну атаку або поширення фейкових новин.

Соціальні платформи часто містять геолокаційні дані, які можна використовувати для відстеження джерел потенційних загроз. Наприклад, аналіз геолокаційних даних може допомогти ідентифікувати джерела дезінформації або шахрайських активностей у певному регіоні.

SOCMINT також охоплює аналіз зображень, що публікуються в соціальних мережах. Використання технологій розпізнавання облич, об’єктів та навіть тексту на зображеннях може допомогти ідентифікувати зловмисників або виявити інформацію, пов’язану з кіберзагрозами. Наприклад, зловмисники можуть випадково опублікувати зображення з конфіденційними даними.

Існує низка інструментів для реалізації SOCMINT, кожен з яких надає можливості для збору та аналізу даних з соціальних мереж, див. рис. 3.1 та дод. Д. Попри високу ефективність, використання SOCMINT для моніторингу соціальних мереж має певні обмеження, пов’язані із конфіденційністю та правовими обмеженнями, швидкою зміною контенту, помилковими даними тощо. Соціальні мережі регулюються законами про конфіденційність, і отримання певної інформації може бути обмежене. Кожна країна має власні правила щодо захисту особистих даних, що може ускладнювати доступ до необхідної інформації. В той же час, соціальні мережі дуже динамічні, і публікації можуть швидко зникати або змінюватися, що ускладнює трекінг певних загроз.

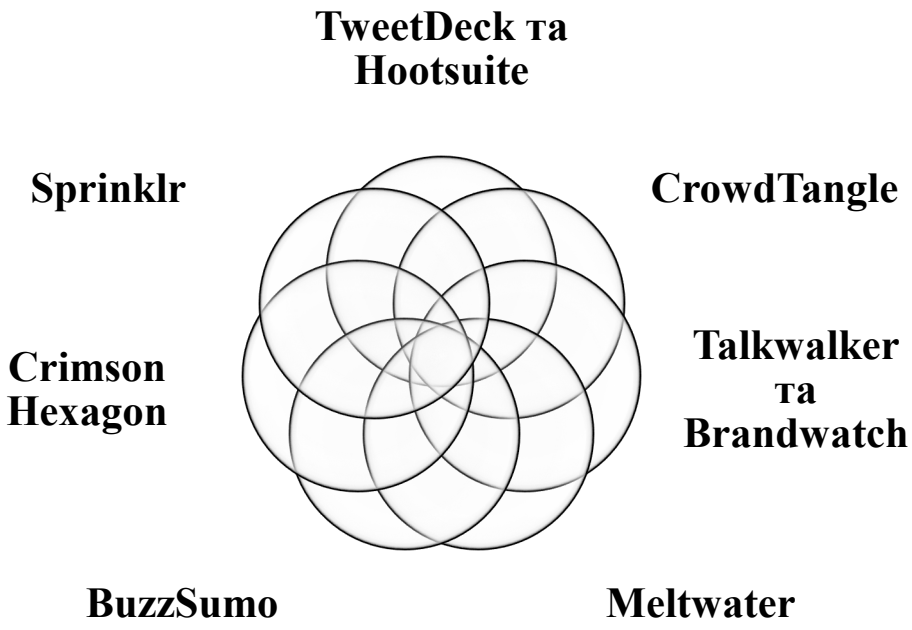


Рисунок 3.1 – Інструменти для реалізації SOCMINT

Висока кількість інформаційного шуму та фальшивих даних може призводити до помилкових висновків, тому важливо використовувати методи перевірки й фільтрації контенту.

Отже, автоматизація SOCMINT із застосуванням технологій ШІ дозволяє значно підвищити ефективність моніторингу соціальних мереж на предмет кіберзагроз. Використання алгоритмів NLP, геоаналізу, візуального аналізу та виявлення аномалій забезпечує швидку ідентифікацію потенційних ризиків, що дозволяє оперативно реагувати на загрози та мінімізувати їхній вплив.

У перспективі подальший розвиток SOCMINT сприятиме створенню більш надійної системи кіберзахисту, адаптованої до динамічних викликів сучасного інформаційного середовища.

### 3.1.3. Удосконалення методів GEOINT (Geospatial Intelligence) для виявлення локалізованих кіберзагроз

GEOINT (Geospatial Intelligence) або геопросторова розвідка – це метод збору, аналізу та інтерпретації даних про географічне розташування з метою виявлення загроз, визначення точних координат подій або об’єктів та їхнього

контексту [36 – 39]. У контексті кібербезпеки, GEOINT надає можливості для моніторингу та виявлення локалізованих кіберзагроз, таких як координовані атаки, пов’язані з певними регіонами, поширення ШПЗ або фізичні загрози, що можуть супроводжувати цифрові атаки.

Методи GEOINT дозволяють поєднувати геопросторову інформацію з даними з інших джерел, наприклад, з соціальних мереж (SOCMINT) або відкритих джерел (OSINT), для створення більш повної картини кіберзагроз. Виявлення географічних даних у кіберзагрозах може бути корисним у випадках визначення джерела кібератак, виявлення локалізованих загроз, ідентифікації зон ризику тощо. Тобто, за допомогою аналізу геопросторових даних можна виявити місця походження підозрілої активності, таких як поширення вірусів або координація ботнет-мереж. Знання точного географічного розташування зловмисників або постраждалих організацій дозволяє швидко реагувати на загрози, які можуть бути локально обмежені. Аналіз геопросторових даних допомагає виявити регіони з високою концентрацією кіберактивності, що може свідчити про підвищені ризики для організацій у цих регіонах.

GEOINT використовує низку інструментів та методів, які дозволяють аналізувати геопросторову інформацію в контексті кібербезпеки. Один із основних методів полягає у визначенні географічного місцезнаходження IP-адрес, з яких проводиться підозріла активність. Це може допомогти ідентифікувати кластери шкідливої активності, пов’язані з певними регіонами або країнами.

За допомогою SOCMINT, GEOINT дозволяє моніторити місця, звідки відбувається поширення шкідливого програмного забезпечення або інформаційних атак. Візуалізація поширення кіберзагроз на карті дозволяє швидко ідентифікувати осередки активності. Також, GEOINT допомагає відстежувати місця, де розташовані об’єкти критичної інфраструктури, такі як Data-центри або енергетичні мережі, які можуть стати мішенями кіберзагроз. Наприклад, моніторинг даних про фізичні атаки на об’єкти інфраструктури може вказувати на зростання кіберзагроз, пов’язаних із цими об’єктами.

Деякі загрози можуть залежати від зовнішніх факторів, наприклад, від екстремальних кліматичних умов або соціально-політичної нестабільності. GEOINT дозволяє інтегрувати ці дані в аналіз, щоб передбачати ризики для певних регіонів.

GEOINT застосовує різні програмні засоби та платформи для моніторингу кіберзагроз. Серед них виділяємо інструменти, зазначені у дод. Е.

Удосконалення методів GEOINT для виявлення локалізованих кіберзагроз із використанням технологій ШІ потребує багатошарового підходу, що поєднує сучасні аналітичні інструменти, обробку даних у реальному часі та адаптивні алгоритми. Одним із ключових напрямів є інтеграція геолокаційних даних із різних джерел, включаючи GPS, супутникові знімки, мобільні пристрої та IoT-системи. Важливо забезпечити доступ до цих даних у режимі реального часу, що дозволить оперативно ідентифікувати потенційні загрози в конкретних регіонах. Наприклад, алгоритми на основі ШІ можуть аналізувати аномалії у переміщенні пристроїв або збіги геолокацій із відомими кіберзагрозами.

Одним із перспективних напрямів є розвиток алгоритмів ML для аналізу супутникових та аерофотознімків. Такі алгоритми дозволяють розпізнавати закономірності, які можуть свідчити про підготовку до атак. Наприклад, незвичне розташування обладнання, зміни в інфраструктурі або підозрілі скупчення активності можуть вказувати на джерело потенційної загрози. Аналіз цих даних може доповнюватися інформацією з соціальних мереж або інших відкритих джерел для формування більш повної картини.

Особливу увагу слід приділити використанню алгоритмів для виявлення та аналізу аномальних поведінкових дій у кіберпросторі. Якщо, наприклад, відстежується значна кількість запитів до певного серверу з однієї геолокації, це може вказувати на DDoS-атаку. ШІ здатний швидко ідентифікувати такі відхилення, враховуючи при цьому локальні особливості, наприклад, час доби, мережеві налаштування та історичні дані про активність у регіоні.

Також варто впроваджувати технології розпізнавання географічних зв’язків у кіберзагрозах. Наприклад, під час аналізу кіберінцидентів алгоритми можуть встановлювати, чи пов’язана певна активність з іншими загрозами, які раніше

фіксувалися у тій самій геолокації. Це дозволить створювати динамічні моделі ризику для окремих регіонів, враховуючи як поточні дані, так і історичні тренди.

Не менш важливо розробляти адаптивні моделі, які зможуть враховувати географічні, культурні та мовні особливості регіонів, де виявляються кіберзагрози. Наприклад, специфічний контент, розміщений у певному регіоні, може мати локальні особливості, які стандартні алгоритми не завжди розпізнають. Використання ШІ для аналізу контексту таких даних дозволить уникнути помилкових висновків.

Важливим аспектом є розширення можливостей візуалізації геолокаційних даних. Інтерактивні карти, які в реальному часі показують осередки активності, дозволять аналітикам швидко ідентифікувати зони ризику. Інструменти доповненої реальності можуть бути корисними для аналізу складних даних, дозволяючи більш ефективно оцінювати ситуацію та приймати рішення.

Зрештою, для підвищення ефективності GEOINT необхідно розвивати систему автоматизації аналізу. Це передбачає створення платформ, здатних автоматично інтегрувати дані з різних джерел, проводити їх попередню обробку, виявляти аномалії та формувати рекомендації для аналітиків. У таких системах ШІ буде не лише інструментом аналізу, але й активним учасником у створенні стратегії кіберзахисту.

## 3.2 Розробка моделей автоматизованого виявлення кіберзагроз засобами штучного інтелекту

Автоматизоване виявлення кіберзагроз із використанням ШІ – це складний, але необхідний крок у забезпеченні безпеки інформаційних систем в умовах сьогодення. Такі моделі мають виявляти загрози не лише на основі заздалегідь відомих сигнатур, а й прогнозувати нові, ще невідомі типи атак за допомогою аналізу поведінкових моделей.

У рамках даного наукового дослідження розглянемо наступні моделі автоматизованого виявлення кіберзагроз засобами ШІ: гібридного виявлення аномалій у мережевому трафіку, глибинної нейронної мережі для виявлення

вторгнень у мережевий трафік, використання трансформерів для аналізу кіберзагроз у текстових даних, автоматизованої роботи DI для виявлення кіберзагроз.

Процес створення ***моделі гібридного виявлення аномалій у мережевому трафіку*** засобами ШІ полягає у необхідності поєднати методи сигнатурного та поведінкового аналізу для виявлення кіберзагроз у реальному часі. Сигнатурний аналіз дозволяє швидко ідентифікувати відомі загрози (наприклад, віруси чи троянські програми), а поведінковий аналіз на основі алгоритмів ML допоможе виявляти раніше невідомі аномалії.

Для розробки моделі необхідно створити датасет, який міститиме реальний мережевий трафік, включаючи дані про відомі атаки та нормальну активність. Далі застосувати нейронні мережі (наприклад, LSTM або GRU, див. рис. 3.2) для аналізу послідовностей мережевого трафіку та методи кластеризації (наприклад, K-Means або DBSCAN) для виявлення аномалій. Важливо також реалізувати комбіновану систему, яка зможе не лише класифікувати дані, але й прогнозувати ризики.

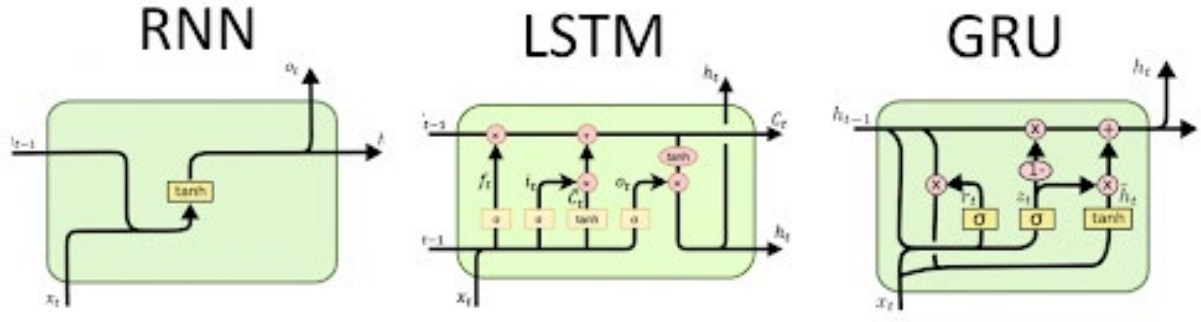


Рисунок 3.2 – Типи рекурентних нейронних мереж (RNN): RNN, LSTM, GRU

LSTM (Long Short-Term Memory) – це тип рекурентної нейронної мережі (RNN), розроблений для вирішення проблеми довгострокової залежності в послідовних даних. На відміну від звичайних RNN, LSTM здатна ефективно «запам’ятовувати» важливу інформацію на тривалих часових проміжках, що робить її придатною для аналізу послідовних даних, таких як мережевий трафік, (якраз для моделі, яку ми зараз розглядаємо), а також текст, аудіо чи відео. LSTM складається з окремих блоків (осередків), кожен із яких має наступні основні

компоненти: запам’ятовування інформації через *комірки пам’яті*, механізми управління потоком інформації через троє спеціальних «воріт» (вхідні ворота (*Input gate*), ворота забуття (*Forget gate*), вихідні ворота (*Output gate*)).

Комірка пам’яті є ключовим елементом LSTM. Вона відповідає за збереження інформації протягом декількох часових кроків. На кожному часовому кроці мережа вирішує, яку інформацію додати до комірки, яку видалити і яку передати далі.

Вхідні ворота визначають, яку нову інформацію потрібно зберегти у комірці пам’яті. Спочатку обчислюється значення, яке необхідно додати до пам’яті, використовуючи тангенсоїдну функцію. Потім визначається ступінь важливості цієї інформації за допомогою сигмоїдної функції.

Ворота забуття при цьому визначають, яку частину попередньої інформації з комірки пам’яті необхідно забути. Вхідні дані та попередній стан передаються через сигмоїдну функцію, яка повертає значення між 0 і 1, що визначає, яку частину пам’яті видалити (0 – повне забуття, 1 – повне збереження).

Вихідні ворота визначають, яку інформацію з пам’яті необхідно передати на вихід. Інформація проходить через сигмоїдну функцію, а потім модулюється тангенсом.

LSTM ефективно вирішує проблему «зникнення градієнта» за рахунок використання комірок пам’яті. Модель може працювати з довгими послідовностями даних, наприклад, мережевим трафіком чи часовими рядами. LSTM використовується в багатьох завданнях, зокрема для аналізу тексту, розпізнавання мови, обробки відео та виявлення аномалій.

Застосування LSTM для виявлення аномалій у мережевому трафіку включає збір даних про трафік (пакети, IP-адреси, порти, протоколи), нормалізацію та кодування вхідних даних, навчання моделі із використанням нормального трафіку для моделювання «нормальної поведінки», використання аномального трафіку для класифікації загроз. Виявлення аномалій відбувається через аналіз кожного пакету у потоці даних та ідентифікацію відхилень від нормальних шаблонів.

В той же час присутні певні обмеження використання LSTM, такі як: обчислювальна складність, оскільки LSTM вимагає більше ресурсів, ніж інші методи; ефективність моделі залежить від якісного та великого набору даних; чутливість до параметрів. Проте LSTM залишається одним із найпотужніших інструментів для обробки послідовних даних і має широкий спектр застосувань у задачах кібербезпеки.

Нейронна мережа GRU (Gated Recurrent Unit) – це спрощений тип рекурентної нейронної мережі (RNN), який було створено як альтернатива до LSTM. GRU зберігає ефективність у роботі з послідовними даними, зменшуючи складність моделі. На відміну від LSTM, GRU має лише два типи воріт, що знижує обчислювальні витрати. Архітектура GRU включає два ключові компоненти: ворота оновлення (*Update gate*) та ворота скидання (*Reset gate*).

Завдяки відсутності окремої комірки пам’яті, GRU швидше навчається і потребує менше ресурсів, ніж LSTM. GRU демонструє схожу продуктивність із LSTM у більшості задач. Завдяки меншій кількості воріт, GRU має менше параметрів, що знижує ризик перенавчання.

Проте через спрощену структуру GRU може бути менш ефективною для задач, які вимагають детальної обробки довготривалих залежностей. У деяких випадках, наприклад, при роботі з дуже довгими послідовностями, LSTM може перевершувати GRU.

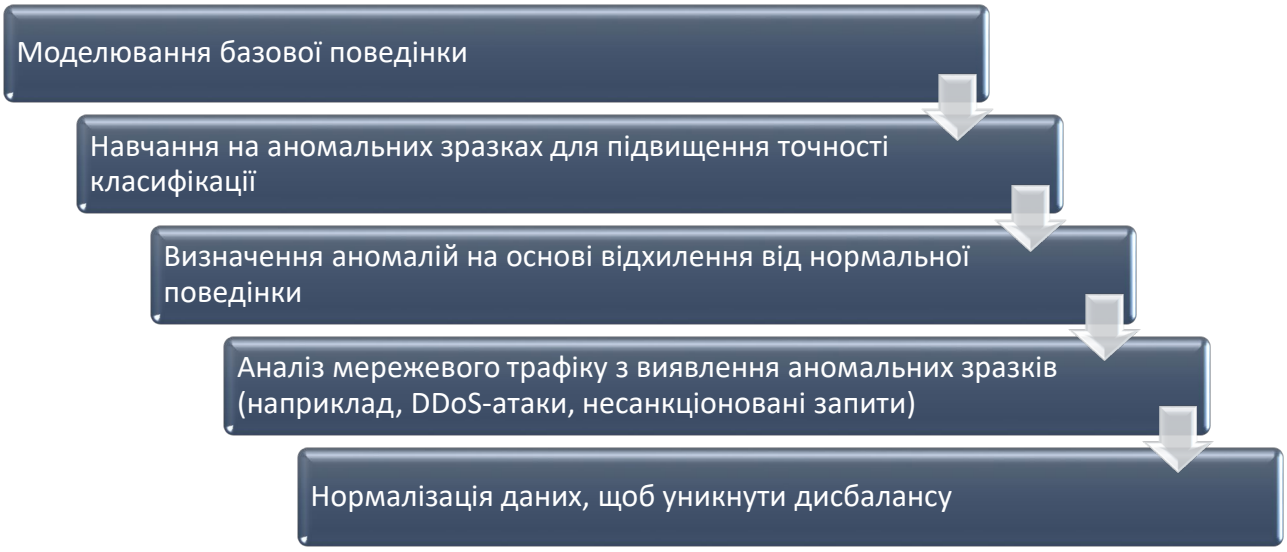


Рисунок 3.3 – Напрямки використання GRU В процесі виявлення аномалій у мережевому трафіку

Загалом, GRU є потужним інструментом для аналізу послідовних даних, особливо коли важливими є швидкість та ефективність. Данна модель підходить для завдань з виявлення аномалій, машинного перекладу, розпізнавання мовлення та інших задач, де потрібно працювати з часовими рядами.

Описана модель гібридного виявлення аномалій у мережевому трафіку буде здатна з високою точністю виявляти нові типи загроз та аномалії у реальному часі з мінімальними затримками (орієнтовно не більше 1 – 2 с). У дод. Ж наведено програмний код, який реалізує описану модель гібридного виявлення аномалій у мережевому трафіку, використовуючи LSTM та кластеризацію K-Means. Код включає обробку даних, побудову нейронної мережі для аналізу послідовностей трафіку та кластеризацію для ідентифікації аномалій.

***Модель глибинної нейронної мережі для виявлення вторгнень у мережевий трафік*** може включати багатошарову архітектуру, здатну знаходити складні закономірності у даних, що допоможе ідентифікувати аномалії або шкідливу активність. Вхідний шар архітектури глибинної нейронної мережі може включати кілька рівнів, кожен з яких виконує свою роль у процесі виявлення. Вхідний шар може містити IP-адреси (джерело, отримувач); час запиту; розмір пакету; порт протоколу; кількість з’єднань за одиницю часу; мітки типу трафіку (HTTP, FTP, SSH тощо); статистичні ознаки (середня затримка, частота помилок). Вхідний шар передає дані до наступних прихованих шарів.

Приховані шари складаються з кількох повнозв’язних шарів (Fully Connected Layers), кожен із яких включає нейрони, що застосовують нелінійні активаційні функції (ReLU, Tanh або Leaky ReLU). Основна мета – навчитися виділяти взаємозв’язки між ознаками та формувати високорівневі репрезентації даних.

Вихідний шар використовує функцію активації (Softmax для багатокласової класифікації або Sigmoid для двокласової) для визначення вірогідності того, що зразок є аномальним або нормальним. На основі вихідних ймовірностей модель класифікує мережевий трафік на категорії: «Нормальний» або «Вторгнення» (DoS, DDoS, Brute Force, SQL Injection, Malware, тощо).

Етапи розробки моделі включають збір і підготовку даних, розробку моделі, навчання моделі, оцінку й тестування. На етапі збору і підготовки даних використовуються набори даних, такі як NSL-KDD, CICIDS2017, або реальні логи мережевого трафіку, а також відбувається передобробка (наприклад, видалення нерелевантних ознак, нормалізація числових даних до діапазону [0, 1], кодування категоріальних ознак як от, One-Hot Encoding для протоколів). NSL-KDD та ICIDS2017 – це два популярних набори даних, які використовуються для досліджень у сфері кібербезпеки, зокрема для виявлення вторгнень (IDS – Intrusion Detection Systems). NSL-KDD є вдосконаленою версією набору даних KDD Cup 1999ь. Його основна мета – вирішення проблем перенасичення та надмірної кількості дублікатів, які були присутні в KDD Cup 1999. NSL-KDD став більш збалансованим і реалістичним для оцінки алгоритмів ML. Характеристики NSL-KDD включають 41 ознаку (кількість байтів, тип протоколу, кількість невдалих підключень і т.д.) та 1 мітку класу – нормальний трафік або один із кількох видів атак, а саме: DoS (Denial of Service, атаки на недоступність сервісу); R2L (Remote to Local, несанкціонований доступ до локального ресурсу); U2R (User to Root, підвищення привілеїв); Probe (сканування мережі).

ICIDS2017 створено Канадським інститутом кібербезпеки. Він призначений для дослідження складних загроз у мережах, включаючи сучасні типи атак. Набір даних моделює реальний трафік, зібраний у контрольованих умовах, з поділом на нормальний трафік і кілька видів атак. ICIDS2017 включає до 80 характеристик (кількість пакетів, середній час затримки, ентропія потоку і т.д.) та 1 мітку класу (нормальний трафік або тип атаки: Brute Force (HTTP, SSH); DoS/DDoS; Web Attacks (SQL Injection, XSS); Infiltration (вторгнення в мережу); Botnet; Port Scan, Heartbleed тощо). До переваг відносимо високу деталізацію характеристик мережевого трафіку. Проте великий обсяг даних може потребувати значних обчислювальних ресурсів.

Етап розробки моделі включає вибір глибини мережі (кількість шарів і нейронів), налаштування функції втрат (Binary Crossentropy для двокласової класифікації або Categorical Crossentropy для багатокласової), використання

оптимізаторів (наприклад, Adam – популярний вибір для більшості задач, або RMSprop для швидшого сходження).

На етапі навчання моделі відбувається поділ даних на навчальну, валідаційну та тестову вибірки. Далі – визначення кількості епох і розміру батчу, та застосування регуляризації (Dropout, L2-норма) для уникнення перенавчання.

Під час етапу оцінки та тестування можна використати метрики Accuracy, Precision, Recall, F1-Score, ROC-AUC (для аналізу продуктивності моделі). Або провести тестування на реальних даних, що не були використані під час навчання.

Програмний код моделі представлений у дод. И. Практичне застосування може полягати в інтеграції моделі у системи IDS (Intrusion Detection System), а також при використанні для моніторингу корпоративних мереж у режимі реального часу. Зазначена модель дозволить автоматизувати виявлення кіберзагроз, що суттєво підвищить ефективність роботи аналітиків і систем кібербезпеки.

***Модель використання трансформерів для аналізу кіберзагроз у текстових даних*** ґрунтується на застосуванні алгоритмів DL, здатних ефективно обробляти та інтерпретувати великі обсяги неструктурованої текстової інформації. Трансформери, такі як BERT (Bidirectional Encoder Representations from Transformers) і GPT (Generative Pre-trained Transformer), або їх спеціалізовані варіанти, адаптовані для аналізу даних з кібербезпеки, виступають основою цієї моделі. Їх використання дозволяє аналізувати текстову інформацію з різних джерел, зокрема звітів про інциденти, системних логів, комунікацій у соціальних мережах, обговорень на форумах у даркнеті тощо.

Основна ідея моделі полягає у навчанні трансформерів виявляти особливості у текстових даних, які можуть свідчити про потенційні загрози. Наприклад, аналіз звітів про інциденти допомагає визначати повторювані типи атак, їхні ключові характеристики та методи експлуатації вразливостей. Аналіз логів системи безпеки дозволяє виявляти аномалії або невідповідності, які можуть свідчити про наявність шкідливої активності. У соціальних мережах і даркнет-форумах такі моделі здатні виділяти загрозливий контент або

попереджати про підготовку до атак на основі характерних лексичних зворотів чи частотності ключових слів.

Завдяки здатності трансформерів працювати з контекстом у двох напрямах (у випадку з BERT) або генерувати новий текст (як GPT), вони можуть бути використані для виконання різних завдань. Наприклад, BERT може класифікувати тексти на основі типів кіберзагроз чи оцінювати ймовірність їх настання. Натомість GPT можна застосовувати для автоматизованого складання звітів про можливі ризики, що знижує навантаження на аналітиків.

Практичне значення даної моделі полягає у можливості раннього виявлення загроз, оскільки своєчасний аналіз текстових джерел дозволяє знаходити згадки про нові типи атак чи експлойти ще до їх масового поширення. Також стає доступною автоматизація аналізу великих обсягів даних, оскільки завдяки потужностям трансформерів, вдається обробляти гігабайти тексту за короткий час, що значно перевершує людські здібності.

Як результат – зменшення часу реагування на інциденти, тому що ідентифікуючи загрозливі події та кореляції, трансформери прискорюють прийняття рішень щодо протидії атакам.

Модель використання трансформерів для аналізу кіберзагроз у текстових даних є перспективним інструментом для забезпечення кібербезпеки, оскільки дозволяє не лише виявляти існуючі загрози, але й прогнозувати нові ризики, базуючись на аналізі текстових даних. Дана властивість є особливо важливою у сучасних умовах постійного збільшення обсягів інформації, яку потрібно аналізувати в режимі реального часу. У дод. К ми представили базовий приклад програмного коду для реалізації моделі використання трансформерів (BERT) з аналізу текстових даних, пов’язаних із кіберзагрозами. Наведений приклад використовує бібліотеку Hugging Face Transformers і демонструє кроки для класифікації текстів на основі кіберзагроз.

В сучасних умовах кіберзагрози еволюціонують з великою швидкістю, і традиційні підходи до їх виявлення часто не встигають за цими змінами. Використання ***автоматизованих моделей на основі DI для виявлення загроз*** дозволяє значно підвищити ефективність систем кіберзахисту, інтегруючи

величезні обсяги відкритої інформації у процес прийняття рішень. Нижче представлена розроблена модель автоматизованої роботи DI, яка базується на трьох ключових аспектах: зборі, аналізі та реакції на інформацію з відкритих джерел.

Модель починається із збору даних, який відбувається через численні джерела відкритої інформації. Соціальні мережі, такі як мережа X і LinkedIn, є важливими платформами, де часто з'являються попередження про нові кіберзагрози або інциденти. Ці джерела доповнюються форумами, у тому числі даркнет-майданчиками, де розповсюджується інформація про вразливості або нові види ШПЗ. Також активно використовуються офіційні звіти, публікації аналітичних компаній та дані DNS-аналізу для виявлення потенційно небезпечних доменів. Для забезпечення високого рівня автоматизації збір даних здійснюється за допомогою API доступу до сервісів, таких як Shodan або VirusTotal, а також веб-скрейпінгу для обробки контенту з відкритих джерел.

Модель автоматизованої роботи DI для виявлення кіберзагроз засобами ШІ поєднує аналітичні можливості відкритих джерел інформації із прогностичною силою алгоритмів ML. Основою цієї моделі є інтеграція засобів DI зі спеціалізованими платформами ШІ, які здатні обробляти великі обсяги неструктурованих даних та виявляти приховані закономірності, що вказують на потенційні кіберзагрози.

Ключовим елементом такої моделі є використання ШІ для автоматизації збору, аналізу та категоризації даних. Інструменти DI вилучають інформацію з відкритих джерел, таких як форуми, соціальні мережі, публічні реєстри, мапи, сайти новин, урядові портали тощо. Алгоритми NLP, засновані на ШІ, дозволяють аналізувати ці дані, визначати їх релевантність та виявляти ключові індикатори загроз, такі як IP-адреси, домени, методи атак або типи зловмисного програмного забезпечення.

Особливу роль у моделі відіграють алгоритми DL, які дозволяють аналізувати не лише текстову інформацію, а й мультимедійний контент. Наприклад, нейронні мережі можуть ідентифікувати зображення чи відео,

пов’язані з діяльністю хакерських груп, або розпізнавати фрагменти коду шкідливих програм, що поширюються у відкритих джерелах.

ШІ забезпечує можливість кореляції даних OSINT із внутрішніми системами кіберзахисту, такими як SIEM (Security Information and Event Management) [8]. Наприклад, якщо у відкритих джерелах виявлено підозрілу IP-адресу, система автоматично зіставляє її з подіями в мережі організації. У разі підтвердження загрози запускається механізм реагування: блокування підозрілого джерела, створення звітів та генерація правил для виявлення схожих атак у майбутньому.

Модель також включає прогностичну аналітику, яка базується на аналізі шаблонів поведінки зловмисників. Завдяки алгоритмам ML система може прогнозувати нові кіберзагрози, аналізуючи послідовності дій хакерів або типові сценарії атак. Наприклад, аналіз активності в соціальних мережах чи хакерських форумах може вказати на підготовку до DDoS-атаки чи розгортання ШПЗ.

Для підвищення ефективності система має механізми самонавчання. Зворотній зв’язок із фахівцями з кібербезпеки та автоматичний аналіз результатів виявлених інцидентів дозволяють алгоритмам ШІ адаптуватися до нових викликів, підвищуючи точність і швидкість аналізу.

Реалізація моделі вимагає сучасної інфраструктури, яка включає високопродуктивні сервери для обчислень, хмарні платформи для зберігання даних та інструменти для обробки Big Data.

Таким чином, модель автоматизованої роботи з DI для виявлення кіберзагроз засобами ШІ забезпечує не лише швидке виявлення поточних загроз, а й формування проактивної стратегії кіберзахисту. Інтеграція OSINT й інших видів DI з ШІ створює умови для побудови інтелектуальної системи, здатної ефективно реагувати на сучасні кіберзагрози та адаптуватися до динамічних змін у глобальному кіберпросторі.

Наведені у даному підрозділі моделі спрямовані на підвищення ефективності виявлення кіберзагроз та аналізу безпекових даних. Кожна з описаних моделей демонструє інноваційний підхід до використання ШІ та аналітичних методів у сфері кібербезпеки. Так, модель гібридного виявлення

аномалій у мережевому трафіку поєднує класичні методи аналізу із застосуванням ML, що дозволяє підвищити точність і швидкість ідентифікації відхилень у мережевій активності. Глибинна нейронна мережа для виявлення вторгнень у мережевий трафік демонструє ефективність у роботі з великими обсягами даних та здатність адаптуватися до нових видів атак. Використання трансформерів для аналізу кіберзагроз у текстових даних забезпечує високу точність у розпізнаванні складних текстових шаблонів і виявленні потенційних загроз на основі інформації з відкритих джерел. Модель автоматизованої роботи DI для виявлення кіберзагроз інтегрує алгоритми ML та методи NLP для аналізу даних із відкритих джерел.

Описані моделі вносять вагомий внесок у розвиток сучасних технологій забезпечення кібербезпеки, надаючи нові інструменти для проактивного виявлення та аналізу загроз у складному цифровому середовищі. Їх подальший розвиток і адаптація до специфічних завдань кібербезпеки відкривають нові горизонти для наукових досліджень та практичної реалізації.

# РОЗДІЛ 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

## 4.1 Охорона праці

Охорона праці є важливою складовою будь-якого проєкту, включаючи роботи, пов’язані з автоматизацією виявлення кіберзагроз із застосуванням ШІ. Забезпечення охорони праці у сфері інформаційних технологій та кібербезпеки включає кілька важливих компонентів. Одним із ключових аспектів є створення безпечних умов праці, які дозволять працівникам виконувати свої обов’язки, мінімізуючи ризики для здоров’я. Робоче місце спеціалістів має відповідати наступним вимогам: ергономічності, тобто забезпеченню комфортних умов для роботи, правильній організації робочого простору, зручному розташуванню комп’ютерів та обладнання; достатньому природному і штучному освітленню, яке не викликає втоми очей та не створює тіней на робочій поверхні; зниженню рівня шуму в робочій зоні, використанню звукоізоляційних матеріалів та пристроїв для зменшення акустичного впливу.

Забезпечення безпеки праці має включати регулярне проведення інструктажів з охорони праці та навчання працівників діям у надзвичайних ситуаціях; забезпеченню робочих місць засобами пожежогасіння, встановлення пожежних сигналізацій та навчання працівників правилам пожежної безпеки; дотриманню правил безпечного користування електроприладами, регулярній перевірці електрообладнання та заземлення.

Не менш важливою є психологічна безпека працівників. Важливо створити умови, що сприятимуть підтримці позитивного морального клімату через психологічну підтримку, наприклад, доступ до консультацій психолога, організацію тренінгів з управління стресом тощо; створення комфортного та дружнього робочого середовища, що сприятиме командній роботі та ефективній комунікації.

Забезпечення охорони праці не можливе без постійного моніторингу та оцінки умов праці через регулярні перевірки та зворотній зв’язок. Моніторинг

може бути реалізований через проведення регулярних інспекцій та аудитів умов праці для виявлення потенційних небезпек та негайного їх усунення, а також активне залучення працівників до процесу оцінки умов праці та збору зворотного зв’язку для постійного покращення безпеки на робочому місці.

Впровадження цих заходів дозволить забезпечити високу ефективність роботи спеціалістів з автоматизації виявлення кіберзагроз, знизити ризики для їхнього здоров’я та підвищити загальний рівень безпеки в компанії.

## 4.2 Стратегії захисту систем штучного інтелекту від кібератак для запобігання надзвичайних ситуацій

Атаки на системи ШІ є відносно новою областю для кібербезпеки, але з кожним роком стають дедалі актуальнішими. Уразливість таких систем полягає в їхній залежності від даних, на яких вони навчаються, та від моделей, які обробляють ці дані. Зловмисники можуть спрямовувати атаки безпосередньо на систему ML, впливаючи на її здатність робити точні прогнози та класифікації. Це створює ризики для багатьох важливих напрямків, таких як безпілотний транспорт, розпізнавання облич, фінансовий моніторинг, медична діагностика тощо, де навіть невелике порушення може мати серйозні наслідки. Наприклад, компанія Microsoft повідомляє про зростання атак на моделі розпізнавання облич. У одному з досліджень показано, як незначні фізичні зміни, такі як додавання наклейок на знаки на дорозі, можуть призвести до того, що автомобіль з автоматичним управлінням неправильно розпізнає ці знаки, створюючи небезпеку на дорозі. Ще одним прикладом є камуфляж для розпізнавання облич. Змінивши зовнішній вигляд або додавши певні об’єкти (наприклад, сонцезахисні окуляри або шарфи специфічних кольорів), зловмисники можуть обдурити систему розпізнавання облич, уникаючи ідентифікації.

На відміну від традиційного програмного забезпечення, де криптографічна перевірка файлів та бібліотек вже стала стандартною практикою, системи ML зазвичай не мають настільки розвинених механізмів безпеки. Публічні датасети, які часто використовуються для навчання моделей, можуть містити помилки або

навіть навмисно змінені дані, що ускладнює розпізнавання шкідливих атак. Важливим є також розуміння різних типів атак, таких як цільові та нецільові атаки, а також атак ухилення та крадіжки інтелектуальної власності.

Розробка моделей для виявлення кіберзагроз засобами ШІ неможлива без врахування ймовірних потенційних атак, спрямованих на їх підрив або маніпуляцію результатами. У сучасному цифровому середовищі зловмисники все частіше використовують складні методи, такі як атаки ухилення, «отруєння» даних, інверсні атаки, крадіжка моделей тощо, які спрямовані на ослаблення ефективності ШІ.

У методі атаки ухилення (Evasion Attacks) зловмисник маніпулює вхідними даними таким чином, щоб модель не змогла виявити загрозу. У випадку дослідженої нами моделі гібридного виявлення аномалій у мережевому трафіку, атака ухилення може полягати у фрагментації пакетів даних, що приховує шкідливі дії у, на перший погляд, звичайному трафіку. Як приклад, розглянемо ситуацію, при якій зловмисник використовує зміну заголовків IP-пакетів або додавання непомітних змін до мережевого трафіку, щоб обійти фільтри, які моделюють аномальні шаблони. Це може призвести до зниження точності виявлення, особливо якщо модель недостатньо навчена на таких випадках. Як результат – такі атаки можуть значно зменшити чутливість моделі, що дозволить зловмисникам залишатися непоміченими.

Метод «отруєння» даних (Data Poisoning) передбачає внесення шкідливих даних у тренувальний набір, що змушує модель робити неправильні висновки. Протидія даному методу особливо актуальна для моделі глибинної нейронної мережі, яка залежить від великих обсягів якісних даних. Приклад даного методу, це коли зловмисники генерують фіктивний мережевий трафік, який виглядає як нормальний, але насправді містить шкідливі інструкції. Наприклад, додавання «легітимного» шкідливого трафіку до набору даних для навчання може навчити модель ігнорувати цей тип загроз у майбутньому. Як результат, модель стає менш ефективною і починає пропускати шкідливу активність, що може призвести до серйозних інцидентів кібербезпеки.

Метод інверсних атак полягає у створенні спеціально модифікованих вхідних даних, які вводять модель в оману. Наприклад, у моделі, яка використовує трансформери для аналізу кіберзагроз у текстових даних, зловмисники можуть маніпулювати текстом, щоб змусити модель класифікувати загрозу як безпечну. Як варіант, зловмисник змінює ключові слова у фішингових повідомленнях на синоніми або додає до тексту «безпечні» контексти, щоб обійти модель. Як-от, замість слова «вірус» може використовуватися «програма для тестування». Як наслідок, модель буде помилково класифікувати шкідливі текстові дані як нешкідливі, дозволяючи зловмисникам досягати своїх цілей.

Реалізація атак на модель автоматизованої роботи DI може полягати у маніпулюванні інформацією у відкритих джерелах або генерації фейкових даних. Наприклад, внесення фейкових записів у форуми або бази даних, які DI використовує для аналізу, створення тисяч позитивних коментарів про шкідливе ПЗ, що моделює «безпечність» загрози. Модель починає працювати з хибними даними, що призводить до неправильних оцінок загроз.

Крадіжка моделей є ще одним видом атак, яка може мати серйозні наслідки для систем, що використовують ШІ. У цій атаці зловмисник прагне відтворити структуру або функціонал оригінальної моделі, не маючи доступу до її внутрішньої архітектури чи навчальних даних. Досягається це шляхом численних запитів до моделі, які дозволяють зібрати достатньо інформації про її поведінку. Наприклад, якщо модель є частиною сервісу ML as a Service, зловмисник може використати відкритий API для тестування моделі в умовах різних вхідних даних. Зловмисник поступово аналізує відповіді на ці запити, створюючи «тіньову модель», яка імітує оригінальну. Такі атаки можуть бути спрямовані не лише на копіювання моделі, а й на виявлення її слабких місць, які потім будуть використані для інших типів атак, таких як атаки ухилення чи «отруєння» даних. Практичний приклад такого підходу можна побачити у випадках зловмисного використання моделей класифікації шкідливого програмного забезпечення. Атакуючий може подати різноманітні фрагменти коду на перевірку, отримуючи дані про те, як саме модель поділяє їх на шкідливі чи безпечні. На основі цих даних створюється тіньова модель, яка потім

використовується для розробки обхідних методів, що дозволяють створювати зразки шкідливого ПЗ, які оригінальна модель не розпізнає.

Окрім того, крадіжка моделі також має серйозні економічні наслідки. Зловмисник може використати отриману інформацію для створення аналогічного продукту, позбавляючи розробника оригінальної моделі конкурентної переваги. У випадку автоматизованих систем DI для виявлення кіберзагроз, відтворення тіньової моделі може дозволити зловмисникам прогнозувати, які типи загроз буде ідентифікувати система, а які залишаться непоміченими, що значно підвищує їхні шанси на успішне здійснення атак.

Для мінімізації ризиків впливу атак на моделі ШІ, спрямованих на їхню компрометацію, необхідно впроваджувати комплексні заходи. У рамках цього дослідження ми виділяємо наступні ключові підходи до захисту вище зазначених моделей:

*1. Використання технік захисту від інверсних атак*. Наприклад, одним із дієвих методів є регуляризація моделей, яка зменшує їхню чутливість до змін у вхідних даних, що знижує ймовірність введення в оману. Також ефективною є генерація контрзразків для тренування, що дозволяє підготувати модель до роботи з потенційно шкідливими вхідними даними. Наприклад, включення спеціально модифікованих прикладів до навчального набору підвищує стійкість моделі до атак.

*2. Ретельна перевірка та фільтрація даних перед їх використанням для навчання.* Як приклад, застосування методів валідації даних, таких як перевірка на дублікати, видалення аномалій та виявлення шкідливого вмісту, що є критично важливими для зменшення ризику «отруєння» моделі шкідливими даними.

*3. Інтеграція моделей із системами багаторівневого захисту.* Наприклад, поєднання алгоритмів ML з традиційними засобами кіберзахисту, такими як міжмережеві екрани, IDS/IPS (системи виявлення та запобігання вторгненням) для забезпечення багаторівневої системи безпеки. Даний підхід дозволить компенсувати можливі недоліки моделі через підтримку з боку інших компонентів.

*4. Постійне оновлення та тестування моделей.* Оновлення навчальних наборів даних, регулярне тестування моделей на нових сценаріях атак, а також впровадження процедур перевірки продуктивності дозволяють адаптувати моделі до мінливого середовища кіберзагроз. Наприклад, періодичне розширення набору контрзразків допоможе зберігати актуальність захисту.

*5. Розробка методів безперервного навчання та самонавчання.* Наприклад, через впровадження алгоритмів, які можуть автоматично оновлювати свої знання на основі нових даних, що дозволить моделі динамічно адаптуватися до нових викликів, зменшуючи необхідність повного перевчання. Це також допоможе підтримувати актуальність у середовищах, що швидко змінюються.

*6. Використання розширених та збалансованих наборів даних для тренування моделі.* Різноманітність даних у навчальному наборі допомагає підготувати модель до роботи з реальними сценаріями. Додавання прикладів із менш представлених категорій загроз або аномалій знижує ризик пропуску специфічних типів атак.

*7. Використання захищених джерел даних.* Наприклад, для мінімізації ризиків «отруєння» необхідно використовувати перевірені та захищені джерела даних. Використання захищених джерел даних зменшує ймовірність проникнення шкідливих зразків у тренувальний або тестовий набір.

*8. Розробка процедур для оцінки цілісності моделі.* Постійний моніторинг результатів роботи моделі на предмет аномальних змін дозволяє своєчасно виявляти можливі порушення в її роботі. Використання метрик для оцінки стійкості моделі до атак забезпечує додатковий рівень контролю.

Розуміння потенційних сценаріїв атак, їхніх механізмів та впливу є критично важливим для побудови стійких моделей, здатних ефективно протистояти загрозам. Комплексний підхід до забезпечення безпеки моделей ШІ гарантує їхню надійність навіть у складних і динамічних умовах сучасного кіберпростору.

Більшість методів захисту на сьогодні є експериментальними або знижують ефективність моделі, що стає перешкодою для впровадження ШІ у критичних

сферах. Наприклад, методи формального доведення стійкості до атак лише частково ефективні для невеликих моделей, тоді як для великих моделей вони практично не застосовуються через обчислювальну складність.

Перспективним напрямом є розробка стійких алгоритмів, які адаптуються до нових типів атак, а також створення гібридних систем, що поєднують різні методи захисту. Одним з перспективних підходів є використання «змагальних прикладів» під час навчання моделей, що дозволить їм адаптуватися до можливих атак у реальних умовах.

Отже, атаки на системи ШІ становлять новий виклик для сфери кібербезпеки, оскільки традиційні методи захисту виявляються неефективними. Основними завданнями для забезпечення стійкості є захист моделей від атак ухилення, «отруєння» даних, крадіжки моделей, інверсних атак тощо. Безпека таких систем залежить від застосування спеціалізованих стратегій захисту, що включають обмеження доступу, перевірку цілісності даних і розробку нових моделей, стійких до модифікацій. Прогрес у цій галузі є ключовим для подальшого безпечного впровадження систем ШІ у важливі сфери, такі як медицина, транспорт, оборона та ін.

# ВИСНОВКИ

Стрімке зростання кількості та складності кіберзагроз висуває нові вимоги до засобів їх виявлення та протидії. Результати дослідження, представлені у даній кваліфікаційній роботі, продемонстрували значний потенціал застосування ШІ для розв’язання зазначених завдань.

Аналіз теоретичних засад та сучасних алгоритмів ML показав, що ШІ виступає ключовим інструментом для класифікації, аналізу та прогнозування кіберзагроз. Зокрема, такі алгоритми, як нейронні мережі, методи кластеризації та дерева рішень, ефективно використовуються для аналізу великих обсягів даних і виявлення аномалій, які можуть свідчити про загрозу.

Особливу увагу приділено інтеграції методів DI з технологіями ШІ. Удосконалення таких підходів, як OSINT, SOCMINT та GEOINT засобами ШІ, дозволяє значно підвищити точність і швидкість аналізу даних з відкритих джерел. Це, в свою чергу, сприяє більш ефективному виявленню локалізованих загроз, моніторингу соціальних мереж та ідентифікації потенційно небезпечних сценаріїв.

Практична частина роботи продемонструвала можливості автоматизованого виявлення кіберзагроз із застосуванням ШІ через моделі: гібридного виявлення аномалій у мережевому трафіку, глибинної нейронної мережі для виявлення вторгнень у мережевий трафік, використання трансформерів для аналізу кіберзагроз у текстових даних, автоматизованої роботи DI для виявлення кіберзагроз. Впровадження запропонованих моделей у системи кібербезпеки дозволятиме не лише оперативно виявляти загрози, але й прогнозувати їх розвиток для ефективного реагування на кіберінциденти до того, як вони завдадуть шкоди.

Важливим аспектом дослідження є питання захисту самих систем ШІ від кібератак. Наведені ключові підходи до захисту моделей автоматизованого виявлення кіберзагроз засобами ШІ: використання технік захисту від інверсних атак, ретельна перевірка та фільтрація даних перед їх використанням для навчання, інтеграція моделей із системами багаторівневого захисту, постійне

оновлення та тестування моделей, розробка методів безперервного навчання та самонавчання, використання розширених та збалансованих наборів даних для тренування моделі тощо. Зазначені підходи спрямовані на забезпечення стійкості моделей до зловмисних впливів, зокрема через використання захищених джерел даних, розробку процедур для оцінки цілісності моделі тощо.

Отримані результати підтверджують, що використання ШІ у сфері кібербезпеки відкриває нові горизонти для протидії сучасним викликам у кіберпросторі. Проте необхідне постійне вдосконалення методологічних підходів, оновлення технологічних засобів та активна співпраця між дослідниками й практиками. Представлені у кваліфікаційній роботі розробки та результати можуть слугувати основою для подальших наукових досліджень та практичного впровадження у сфері кібербезпеки.

Апробація та обговорення [14 – 16; 37; 38; 47 – 55, дод. А] окремих результатів дослідження на міжнародних та всеукраїнських конференціях отримала позитивні відгуки. Перспективи подальших досліджень включають удосконалення запропонованих методів, розширення сфери їх застосування та інтеграцію новітніх технологій для підвищення ефективності захисту інформаційних систем.

# СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Zhang, Z., Ning, H., Shi, F., & Others. (2022). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. Artificial Intelligence Review, 55, 1029–1053.

2. Farheen, A. M., Bibhu, D., Pawankumar, S., & Nikhitha, Y. (2022). The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review. International Journal of Advanced Research in Computer and Communication Engineering. Retrieved from https://ssrn.com/abstract=4323317.

3. Tymoshchuk, V., Vorona, M., Dolinskyi, A., Shymanska, V., & Tymoshchuk, D. (2024). SECURITY ONION PLATFORM AS A TOOL FOR DETECTING AND ANALYSING CYBER THREATS. Collection of scientific papers «ΛΌГОΣ», (December 13, 2024; Zurich, Switzerland), 232-237.

4 Tymoshchuk, V., Mykhailovskyi, O., Dolinskyi, A., Orlovska, A., & Tymoshchuk, D. (2024). OPTIMISING IPS RULES FOR EFFECTIVE DETECTION OF MULTI-VECTOR DDOS ATTACKS. Матеріали конференцій МЦНД, (22.11. 2024; Біла Церква, Україна), 295-300.

5. Balantrapu, S.S. (2022). Evaluating AI-Enhanced Cybersecurity Solutions Versus Traditional Methods: A Comparative Study. International Journal of Sustainable Development Through AI, ML, and IoT, 1(1), 1–15.

6. Tymoshchuk, V., Vantsa, V., Karnaukhov, A., Orlovska, A., & Tymoshchuk, D. (2024). COMPARATIVE ANALYSIS OF INTRUSION DETECTION APPROACHES BASED ON SIGNATURES AND ANOMALIES. Матеріали конференцій МЦНД, (29.11. 2024; Житомир, Україна), 328-332.

7. Саєнко, Д.О., & Колісник, Т.П. (2023). Основні напрями застосування технологій ШІ у кібербезпеці. Протидія кіберзлочинності та торгівлі людьми: Збірн. мат. міжн. наук.-практ. конф.ї. Вінниця: ХНУВС, 158 – 160.

8. Болілий, В., Суховірська, Л., & Лунгол, О. (2021). Операційний центр безпеки як послуга на основі SIEM. Наукові записки БДПУ. 177 – 186.

9. Гладка, Ю.А., & Назаренко, Є.О. (2021). Аналіз застосування технологій штучного інтелекту в кібербезпеці. Сучасні тенденції розвитку інформаційних

систем і телекомунікаційних технологій: збірник матеріалів III міжнародної науково-практичної конференції Київ: НУХТ. 64 – 66.

10. Tymoshchuk, D., Yasniy, O., Mytnyk, M., Zagorodna, N., Tymoshchuk, V., (2024). Detection and classification of DDoS flooding attacks by machine learning methods. CEUR Workshop Proceedings, 3842, pp. 184 - 195.

11. Лунгол, О.М., & Габорець, О.А. (2023). OSINT-технології в правоохоронній діяльності: навчальний посібник. Кропивницький: Book Creator.

12. Шевчишен, А.В., Романов, М.Ю., Волобоєв, А.О., Лунгол, О.М., Габорець, О.А., Головкін, С.В., & Вітвіцький, С.С. (2023). Організація розкриття шахрайств, учинених в кіберпросторі: Монографія. Київ: Алерта.

13. Лунгол, О.М. (2024). Кібербезпека в умовах гібридної війни: виклики та шляхи протидії. In Безпекова ситуація в Україні в умовах війни: Матеріали Всеукраїнської науково-практичної конференції (pp. 120–123). Вінниця: ТВОРИ.

14. Лунгол, О.М., & Агішева, А.В. (2023). Технології створення та застосування систем захисту інформаційно-комунікаційних систем. Topical aspects of modern scientific research: Proceedings of the 2nd International Scientific and Practical Conference (pp. 255–260). Tokyo, Japan: CPN Publishing Group.

15. Лунгол, О., & Агішева, А. (2023). Використання технології Deception у боротьбі з кіберзагрозами. Кібербезпека в Україні: правові та організаційні питання: матеріали Міжн. наук.-практ. конф. Одеса: ОДУВС, 75–76.

16. Лунгол, О., & Макаринська, А. (2023). Кіберзлочинність як складова інформаційної війни: аналіз досвіду України. Всеукраїнська науково-практична конференція «Сучасні пріоритети розвитку України: економічна та інформаційна безпека» (10.10.2023 року, м. Дніпро) (с. 35–38). Дніпро: ДДУВС.

17. Звіт про результати роботи Департаменту кіберполіції Національної поліції України у 2023 році. (2023). URL: https://griml.com/MjWuw (Дата звернення: 15.10.2024).

18. Просвіріна, Т.В., Габорець, О.А., & Лунгол, О.М. (2023). Використання геоінформаційних технологій для покращення організаційно-аналітичного забезпечення правоохоронної діяльності. Наукові інновації та передові технології, 5(19), 340–345.

19. Lohmann, G.T. (2023). Human Intelligence Throughout History: An Analysis of the Changes in Human Intelligence Collection Techniques from the Cold War to the Present. URL: https://dspace.cuni.cz/handle/20.500.11956/187416 (Дата звернення: 17.10.2024).

20. Lypa, B., Horyn, I., Zagorodna, N., Tymoshchuk, D., Lechachenko T., (2024). Comparison of feature extraction tools for network traffic data. CEUR Workshop Proceedings, 3896, pp. 1-11.

21. ТИМОЩУК, Д., & ЯЦКІВ, В. (2024). USING HYPERVISORS TO CREATE A CYBER POLYGON. MEASURING AND COMPUTING DEVICES IN TECHNOLOGICAL PROCESSES, (3), 52-56

22. ТИМОЩУК, Д., ЯЦКІВ, В., ТИМОЩУК, В., & ЯЦКІВ, Н. (2024). INTERACTIVE CYBERSECURITY TRAINING SYSTEM BASED ON SIMULATION ENVIRONMENTS. MEASURING AND COMPUTING DEVICES IN TECHNOLOGICAL PROCESSES, (4), 215-220.

23. Tymoshchuk, V., Pakhoda, V., Dolinskyi, A., Karnaukhov, A., & Tymoshchuk, D. (2024). MODELLING CYBER THREATS AND EVALUATING THE PERFORMANCE OF INTRUSION DETECTION SYSTEMS. Grail of Science, (46), 636–641. https://doi.org/10.36074/grail-of-science.29.11.2024.081

24. Frasca, D., Venturi, G., Ustenko, M., & Zanasi, A. (2023). Technologies for IMINT and SIGINT. 2023 IEEE International Workshop on Technologies for Defense and Security (TechDefense), Rome, Italy https://doi.org/10.1109/ TechDefense59795.2023.10380827 (Дата звернення: 18.10.2024).

25. Kotaridis, I., & Benekos, G. (2023). Integrating Earth observation IMINT with OSINT data to create added-value multisource intelligence information: A case study of the Ukraine-Russia war. Security & Defence Quarterly, 43(3), 1–21.

26. Державна служба спеціального зв’язку та захисту інформації України. (2023). Звіт оперативного центру реагування на кіберінциденти ДЦКЗ. URL: https://griml.com/xz5su (Дата звернення: 15.10.2024).

27. Школа OSINT. (2024). Дія. Освіта. Освітній серіал. URL: https://osvita.diia.gov.ua/courses/osint-school (Дата звернення: 17.10.2024).

28. OSINT – розвідка з відкритих джерел та інформаційна безпека. (2024). Prometheus. Освітній курс. URL: https://griml.com/5Py6l (Дата звернення: 22.10.2024).

29. Живило, Є.О., & Дамян, М.Ю. (2024). Інструменти OSINT framework. Стан, досягнення та перспективи інформаційних систем і технологій: Матеріали XXIV Всеукраїнської науково-технічної конференції молодих вчених, аспірантів та студентів, Одеса, 18–19.04.2024 р. (с. 105–106). Одеса: ОНТУ.

30. Сторчак, А.С., & Самойлов, І.В. (2023). Використання OSINT-технології в сучасних сервісах. Комплексне забезпечення якості технологічних процесів та систем – 2023, 300.

31. Горбач, М.М. (2023). Використання технології OSINT для збору, узагальнення та аналізу інформації на основі різних соціальних мереж: кваліф. робота бакалавра за спец. 125 – Кібербезпека. Тернопіль: ТНТУ.

32. Рябцева, Д.О., & Гудзь, Т.І. (2023). OSINT як засіб боротьби в російсько-українській війні. Протидія кіберзлочинності та торгівлі людьми: збірник мат. міжн. науково-практ. конф. (с. 155 – 157). Вінниця: ХНУВС.

33. Цуркан, І.О., Шимко, А.О., Верзілов, М.Р., & Стецик, Р.М. (2024). Використання методу OSINT під час дії правового режиму воєнного стану в Україні. The 6th International scientific and practical conference European congress of scientific achievements (с. 222). Barcelona: Barca Academy Publishing.

34. Комiссарчук, Ю.А., & Черевко, В.В. (2022). OSINT як один із інструментів для збирання інформації про воєнні злочини російської федерації в Україні. The 14th International scientific and practical conference «Modern science: innovations and prospec» (с. 451–454).

35. Віннічук, Б.В., & Стрелков, В.В. (2024). Роль OSINT-розвідки в забезпеченні національної безпеки. Мат. VIII Міжнародної науково-практичної конференції (ДДУВС, 15.03.2024), Частина ІІ. 335–336.

36. Лунгол, О., & Торгало, П. (2023). Аналіз та управління ризиками в сфері інформаційної безпеки. У ІІ Міжнародна науково-практична Інтернет-конференція “Інновації та перспективні шляхи розвитку інформаційних технологій” (ІПШРІТ-2023) Черкаси. 57–61.

37. Лунгол, О., & Макаринська, А. (2023). Сучасні методи виявлення та аналізу соціально-інженерних атак в інформаційних системах. Тези доповідей II Міжнародної науково-практичної конференції «Інновації та перспективні шляхи розвитку інформаційних технологій» (06.12.2023 р., Черкаси). 243–244.

38. Muzh, V., & Lechachenko, T. (2024). Computer technologies as an object and source of forensic knowledge: challenges and prospects of development. Scientific Journal of TNTU, 115(3), 17–22.

39. Запорожченко, М.М. (2023). Місце OSINT в життєвому циклі кібератаки. Телекомунікаційні та інформаційні технології, (1), 53–60.

40. Stanko, A., Wieczorek, W., Mykytyshyn, A., Holotenko, O., & Lechachenko, T. (2024). Realtime air quality management: Integrating IoT and Fog computing for effective urban monitoring. CITI, 2024, 2nd.

41. Riebe, T. (2023). Privacy concerns and acceptance factors of OSINT for cybersecurity: A representative survey. У Technology Assessment of Dual-Use ICTs: How to Assess Diffusion, Governance and Design. Wiesbaden: Springer Fachmedien Wiesbaden. 221–248.

42. Riebe, T., Biselli, T., Kaufhold, M.-A., & Reuter, C. (2023). Privacy concerns and acceptance factors of OSINT for cybersecurity: A representative survey. Proceedings on Privacy Enhancing Technologies (PoPETs). URL: https://doi.org/10.56553/popets-2023-0028

43. Derkach, M., Skarga-Bandurova, I., Matiuk, D., & Zagorodna, N. (2022, December). Autonomous quadrotor flight stabilisation based on a complementary filter and a PID controller. In 2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT) (pp. 1-7). IEEE.

44. Hwang, Y.W., Lee, I.Y., Kim, H., & Kim, D. (2022). Current status and security trend of OSINT. Wireless Communications and Mobile Computing. 21 – 35.

45. Лунгол, О.М., & Ковальова, О.В. (2022). Збірник практичних завдань з навчальної дисципліни «Інформаційне забезпечення професійної діяльності»: інтеракт. навч. посіб. з ел-ми білінгв. підходу. Кропивницький: Book Creator.

46. Компендіум OSINT-інструментів. (2024). URL: https://griml.com/XM7Ij (Дата звернення: 22.10.2024).

47. Габорець, О., & Лунгол, О. (2023). Criminal analysis paradigm in the context of digital security: theoretical foundations and practical challenges. Матеріали Міжн. наук.-практ. конф. «Кібербезпека в Україні: правові та організаційні питання», м. Одеса, ОДУВС, 17.11.2023 р., 156-157.

48. Lunhol, O. (2024). The relevance of teaching OSINT technologies in the training of Law Enforcement specialists in the context of modern security challenges. In Learning & Teaching: In the World after the War: Conference Proceedings of III International Scientific & Practical Conference (p. 228). H.S. Skovoroda Kharkiv National Pedagogical University, Kharkiv, Ukraine, November 8, 2024.

49. Лунгол, О. М., & Макаринська, А. В. (2023). Сучасні методи виявлення та аналізу соціально-інженерних атак в інформаційних системах. Тези доповідей II Міжнар. наук.-практич. конф. «Інновації та перспективні шляхи розвитку інформаційних технологій», 06.12.2023 р., м. Черкаси, 243-244.

50. Лунгол, О.М., & Агішева, А.В. (2023). Технології створення та застосування систем захисту інформаційно-комунікаційних систем. Topical aspects of modern scientific research. Proceedings of the 2nd Int. scient, and practical conference. CPN Publishing Group, Tokyo, Japan, 26-28.10.2023 p., 255-260.

51. Лунгол, О., & Агішева, А. (2023). Використання технології Deception у боротьбі з кіберзагрозами. Мат. міжн. наук.-практ. конф. «Кібербезпека в Україні: правові та організаційні питання», 17.11.2023 р., м. Одеса, 75-76.

52. Лунгол, О. М., & Позігун, Б. В. (2024). Методи захисту цифрового відбитку пристрою в онлайн-середовищі. Міжн. наук.-практ. конф. «Актуальні питання підготовки фахівців для сектору безпеки і оборони в умовах війни», 19.04.2024 р., м. Кропивницький, 79-80.

53. Лунгол, О., & Макаринська, А. (2023). Кіберзлочинність як складова інформаційної війни: аналіз досвіду України. Всеукраїнська науково-практична конференція «Сучасні пріоритети розвитку України: економічна та інформаційна безпека», 10 жовтня 2023 р., м. Дніпро. Дніпро: ДДУВС, 84-86.

54. Лунгол, О., & Шаєц, Є. (2023). Кіберпростір як квінтесенція глобалізованого суспільства. Регіональні особливості злочинності: сучасні тенденції та стратегії протидії: III Всеукр. Наук.-практ. конф., 02.06.2023 р., КННІ ДонДУВС, м. Кривий Ріг, 92-94.

55. Лунгол, О., & Торгало, П. (2024). Ідентифікація небезпек та розвиток стратегій захисту у віртуальному світі. Безпека дітей в Інтернеті: попередження, освіта, взаємодія: збірник мат. III Всеукр. наук.-практ. конф., 05-09.02.2024 р., Кропивницький: КЗ «КОІППО імені Василя Сухомлинського», 63-64.

# ДОДАТКИ

Додаток А Публікації за темою роботи

Лунгол О. Огляд методів та стратегій кібербезпеки засобами штучного інтелекту. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка». № 1(25). 2024. С. 379 – 389.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.25.379389
УДК 004.056.55:004.8

**Лунгол Ольга Миколаївна**
кандидат педагогічних наук, доцент,
доцент кафедри оперативно-розшукової діяльності та інформаційної безпеки
Донецький державний університет внутрішніх справ, Кропивницький, Україна
ORCID ID: 0000-0001-8128-0072
*olyalungol@gmail.com*

# ОГЛЯД МЕТОДІВ ТА СТРАТЕГІЙ КІБЕРБЕЗПЕКИ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

**Анотація.** У сучасному світі інформаційні технології стрімко розвиваються, що призводить до зростання кількості та складності кіберзагроз, зокрема фішингу, шкідливого програмного забезпечення та атак з використанням соціальної інженерії. Зростання кількості та складності кіберзагроз створює нагальну потребу у вдосконаленні методів захисту інформаційних систем. Штучний інтелект (ШІ), особливо технології машинного навчання та глибокого навчання, демонструють значний потенціал у підвищенні рівня кібербезпеки. Ця стаття присвячена огляду сучасних методів та стратегій кібербезпеки, що базуються на застосуванні ШІ, а також оцінці їх ефективності у виявленні та протидії кіберзагрозам. Проаналізовано останні дослідження як вітчизняних, так і закордонних науковців, які акцентують увагу на здатності ШІ аналізувати великі обсяги даних, виявляти приховані закономірності, прогнозувати потенційні загрози та автоматизувати процеси реагування на інциденти. Висвітлено ключові напрямки досліджень, включаючи виявлення аномалій, моделювання загроз, автоматизацію процесів реагування на інциденти та забезпечення розуміння рішень, прийнятих системами ШІ. Особлива увага приділяється інтеграції ШІ в існуючі системи кібербезпеки та його здатності до адаптації у відповідь на нові загрози. Стаття також обговорює основні виклики та перспективи застосування ШІ у кібербезпеці, включаючи етичні та правові аспекти, такі як питання приватності, прозорості рішень та відповідальності за дії, здійснені на основі рішень ШІ-систем. Останні статистичні дані свідчать про стрімке зростання ринку засобів ШІ для забезпечення кібербезпеки, що підкреслює важливість і актуальність цієї теми у сучасних умовах. Результати аналізу підтверджують, що використання ШІ дозволяє автоматизувати процеси моніторингу, виявлення та реагування на загрози, що зменшує час реакції на інциденти та підвищує загальний рівень захисту інформаційних систем. Разом з тим, впровадження ШІ у кібербезпеку стикається з низкою викликів, таких як забезпечення прозорості рішень, прийнятих ШІ, а також захист від потенційних загроз, створених з використанням тих самих технологій. Дослідження цієї теми сприяє стратегічному розвитку та інноваціям у сфері кібербезпеки, надаючи дослідникам та фахівцям нові інструменти та методи для забезпечення безпеки інформаційних систем. Отже, з огляду на швидке зростання та еволюцію кіберзагроз, дослідження ролі ШІ в кібербезпеці є надзвичайно актуальним та важливим. Це дозволяє не тільки підвищити ефективність захисту, але й сприяє розвитку нових стратегій та технологій для протидії загрозам у цифрову епоху.

**Ключові слова:** штучний інтелект; кібербезпека; загрози; інформаційний простір.

## ВСТУП

У сучасному світі інформаційні технології стрімко розвиваються, що призводить до зростання кількості та складності кіберзагроз. Це зумовлює необхідність вдосконалення методів захисту інформації. Останні роки характеризуються значним збільшенням кількості кіберзагроз, таких як фішинг, шкідливе програмне забезпечення та атаки з використанням соціальної інженерії. Кіберзлочинці застосовують дедалі

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

витонченіші методи, що вимагає впровадження нових технологій для ефективного захисту інформації.

Штучний інтелект (ШІ) став потужним інструментом для підвищення рівня кібербезпеки та планування і здійснення кібератак. Технології ШІ, зокрема машинне навчання та глибоке навчання, значно просунулися вперед і знайшли широке застосування у різних сферах, включаючи кібербезпеку. Здатність ШІ до аналізу великих обсягів даних та виявлення прихованих закономірностей робить його незамінним у боротьбі з кіберзагрозами.

**Постановка проблеми.** Збільшення кількості та складності кіберзагроз у сучасному світі створює нагальну потребу у вдосконаленні методів захисту інформації. Сучасні кіберзлочинці застосовують все більш складні та витончені методи, що викликає необхідність впровадження нових технологій для ефективного захисту інформаційних систем. У зв'язку з цим, виникає потреба у дослідженні можливостей використання ШІ для підвищення кібербезпеки.

Дослідження методів та стратегій кібербезпеки з використанням ШІ є актуальним та важливим напрямом у сучасних умовах зростання кількості кіберзагроз. Застосування технологій ШІ дозволяє значно підвищити ефективність захисту інформації, забезпечуючи швидке виявлення та протидію кіберзагрозам. Наукові дослідження у цій галузі є необхідними для розробки нових, більш ефективних методів захисту, що відповідатимуть сучасним викликам у сфері кібербезпеки.

**Аналіз останніх досліджень і публікацій.** На актуальність дослідження місця ШІ у кібербезпеці вказує низка наукових робіт як вітчизняних, так і зарубіжних науковців. Так, Магденко А. Р. [1], Бучацький І. О. [1] та Бондаренко І. О. [1] у своїх дослідженнях зазначають, що завдяки здатності аналізувати великі обсяги даних, виявляти вразливості в системах безпеки та прогнозувати потенційні кіберзагрози, ШІ стає ключовим інструментом у боротьбі із кіберзагрозами та захисті від кібератак. Вони також зазначають, що сучасні хакери та шахраї активно використовують ШІ для вдосконалення своїх атак, збільшуючи їх швидкість, масштаб та складність.

Товстуха Н. [2] вказує, що наразі близько 50% підприємств вже використовують комбінацію ШІ та інструментів машинного навчання, а понад 90% організацій планують запровадити такі інструменти у майбутньому. Устименко В. [3] та Олішевський І. [3] наводять статистичні дані, згідно з якими, за оцінкою MarketsandMarkets, ринок засобів ШІ для забезпечення кібербезпеки буде зростати на 23,3% щороку в період з 2019 до 2026 року, збільшуючись з 8,8 млрд до 38,2 млрд доларів США.

Каун Ю. [4] та Собчук О. [4] наводять результати дослідження вчених Університету Індіани, які виявили понад 200 вкрадених і зламаних відкритих мовних моделей, що пропонуються для хакерської діяльності. Це свідчать про зростаючу тенденцію використання ШІ для зловмисних цілей. Зокрема, створення дипфейків на основі цілком реалістичних та оманливих відео, які можуть мати значний вплив на громадську думку та хід подій. Зневажливі відео, підроблені документи та фейкові акаунти в соціальних мережах можуть бути використані для маніпулювання громадськістю, підриву довіри до інституцій та розпалювання соціальних заворушень.

Старший науковий співробітник Українського науково-дослідного інституту спеціальної техніки та судових експертиз Служби безпеки України Гуржій С. [5] зазначає, що алгоритм роботи системи захисту від кіберзагроз з використанням ШІ повинен працювати автономно та виконувати наступні дії: перевіряти мережі на наявність вразливостей з високою точністю; виявляти нові загрози в режимі реального часу; з високою ефективністю реалізовувати індивідуальні заходи з пом'якшення

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

наслідків, такі як автоматичне виправлення програмного забезпечення або блокування шкідливих IP-адрес.

**Метою статті** є огляд сучасних наукових досліджень вітчизняних й закордонних науковців щодо методів та стратегій кібербезпеки, що базуються на застосуванні ШІ, а також оцінка їх ефективності у виявленні та протидії кіберзагрозам.

## РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Вітчизняні науковці Котенко Д. [6] та Хлапонін Ю. [6] виділяють ключову перевагу використання ШІ в кібербезпеці — це його здатність аналізувати великі обсяги даних для виявлення патернів та аномалій, які можуть свідчити про кіберзагрози. Ця здатність стає все більш важливою, оскільки обсяг даних, які генеруються та збираються, постійно зростає. Люди аналітики просто не в змозі впоратися з таким обсягом даних самостійно, тому ШІ є незамінним інструментом для виявлення кіберзагроз [6]. Здатність до аналізу великих обсягів даних для виявлення патернів та аномалій за допомогою ШІ в кібербезпеці можливо реалізувати за допомогою різноманітних засобів, включаючи: методи машинного навчання, аналіз поведінки користувачів, системи виявлення вторгнень (IDS), аналіз журналів подій (SIEM), технології Big Data. Систематизовану інформацію щодо перелічених методів представлено у табл. 1. [6] – [12].

*Таблиця 1*

**Методи для аналізу великих обсягів даних у кібербезпеці за допомогою ШІ**

| Назва методу | Зміст методу | Характеристика | Зміст характеристики |
|---|---|---|---|
| Методи машинного навчання | Відіграє критичну роль у кібербезпеці, дозволяючи автоматизовано аналізувати великі обсяги даних та виявляти закономірності й аномалії, що можуть свідчити про кібератаки. | «Супервізоване» або «кероване» навчання (англ. Supervised Learning) | Використовується для класифікації даних, де система навчається на маркованих даних (наприклад, нормальні та аномальні дії) і потім застосовує цей досвід для нових даних. Приклади алгоритмів: логістична регресія, дерева рішень, SVM (Support Vector Machines). |
| | | «Ненаглядне» або «некероване» навчання (англ. Unsupervised Learning) | Використовується для виявлення невідомих патернів у даних без попереднього маркування. Приклади алгоритмів: кластеризація (k-means, DBSCAN), алгоритми зниження розмірності (PCA, t-SNE). |
| | | Глибоке навчання (англ. Deep Learning) | Включає використання нейронних мереж для складних завдань, таких як розпізнавання зображень або обробка природної мови. Рекурентні нейронні мережі (RNN) та згорткові нейронні мережі (CNN) є популярними методами для аналізу послідовних даних та зображень відповідно. |

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

| | | | |
|---|---|---|---|
| Аналіз поведінки користувачів (UBA) | User Behavior Analytics (UBA) використовує ШІ та ML (Machine Learning) для моніторингу та аналізу поведінки користувачів у реальному часі з метою виявлення аномалій, які можуть свідчити про внутрішні загрози або компрометацію облікових записів. | **Моніторинг дій** | Збирає дані про дії користувачів, включаючи логін, доступ до файлів, використання мережевих ресурсів та інші операції. |
| | | **Виявлення аномалій** | Використовує моделі ML для порівняння поточної поведінки з історичними даними та виявлення відхилень, які можуть бути ознаками загрози. |
| | | **Адаптивність** | Системи UBA можуть адаптуватися до змін у поведінці користувачів, забезпечуючи постійне оновлення моделей для більш точного виявлення загроз. |
| Системи виявлення вторгнень (IDS) | Intrusion Detection Systems (IDS) використовують ШІ для аналізу мережевого трафіку та виявлення підозрілих дій, які можуть свідчити про вторгнення. | **Мережеві IDS (NIDS)** | Моніторять мережевий трафік і аналізують пакети даних для виявлення аномалій. Застосовуються методи сигнатурного та евристичного аналізу. |
| | | **Хостові IDS (HIDS)** | Моніторять активність на окремих пристроях, такі як журнали подій, цілісність файлів, системні виклики, для виявлення вторгнень. |
| | | **Гібридні системи** | Комбінують NIDS та HIDS для більш комплексного підходу до виявлення загроз. |
| Аналіз журналів подій (SIEM) | Security Information and Event Management (SIEM) системи об'єднують та аналізують журнали подій з різних джерел для виявлення загроз і забезпечення відповідності вимогам безпеки. | Централізоване збирання даних | Збирають дані з мережевих пристроїв, серверів, додатків та інших джерел. |
| | | Кореляція подій | Використовують алгоритми для кореляції подій з різних джерел, що дозволяє виявити складні атаки, які можуть бути непомітні при аналізі окремих подій. |
| | | Робота в режимі реального часу | Забезпечують моніторинг у реальному часі та автоматичне реагування на виявлені загрози. |
| Технології Big Data | Технології Big Data дозволяють обробляти та аналізувати великі обсяги даних з високою швидкістю, що є критичним для виявлення та реагування на кіберзагрози. | Розподілена обробка | Використання платформ, таких як Hadoop та Apache Spark, для обробки великих обсягів даних у розподіленому середовищі. |
| | | Аналіз у реальному часі | Використання технологій потокової обробки, таких як Apache Kafka, для аналізу даних у реальному часі. |
| | | Інтеграція даних | Здатність інтегрувати дані з різних джерел, включаючи мережевий трафік, журнали подій, соціальні мережі та інші джерела. |

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

Кожен із описаних методів має свої переваги та особливості, що робить їх важливими інструментами в арсеналі засобів кібербезпеки. Використання ШІ та машинного навчання дозволяє суттєво покращити виявлення загроз і реагування на них, забезпечуючи більш ефективний захист інформаційних систем у сучасному світі.

Серед закордонних досліджень за темою статті виділяємо роботи Саркер І. [13], Фурхад М. [13], Новрози Р. [13], Камачо Н. [14], Кузлу М. [15], Фэйр Ч. [15], Гюлер О. [15], Гупта М. [16], Акири К. [16], Арьял К. [16], Паркер Е. [16], Прахарадж Л. [16], Ясін А. [17], Мохамед Н. [18], Капуано Н. [19], Фенза Г. [19], Лоя В. [19], Станзіоне К. [19], Адевусі А. [20], Околі У. [20], Олорунсого Т. [20], Адага Е. [20], Дараоджимба Д. [20], Обі О. [20] та ін.

Перелічені роботи підкреслюють значний потенціал ШІ у підвищенні ефективності кібербезпеки. Основними напрямками досліджень є виявлення аномалій, моделювання загроз, автоматизація процесів реагування на інциденти та забезпечення розуміння рішень, прийнятих системами ШІ. Науковці акцентують увагу на важливості інтеграції ШІ в існуючі системи кібербезпеки для адаптації до наявних та нових загроз. Так, дослідження Саркера І. [13], Фурхада М. [13] та Новрози Р. [13] містять огляд застосування ШІ у кібербезпеці, акцентуючи увагу на моделюванні безпеки та можливих напрямках досліджень. Автори досліджують різні підходи до використання ШІ для підвищення ефективності кібербезпеки, аналізують поточні технології. Основні акценти зроблено на моделюванні загроз, виявленні аномалій та реагуванні на інциденти безпеки.

Камачо Н. [14] досліджує роль ШІ в сучасній кібербезпеці, зосереджуючи увагу на реагуванні на загрози у цифрову епоху. Автор детально аналізує, як технології ШІ можуть допомогти виявляти та протидіяти новим кіберзагрозам, обговорює переваги та виклики впровадження ШІ в системи кібербезпеки. Особлива увага приділяється інтеграції ШІ в існуючі системи та його здатності до адаптації у відповідь на нові загрози.

Кузлу М. [15], Фэйр Ч. [15] та Гюлер О. [15] демонструють результати дослідження ролі ШІ у забезпеченні кібербезпеки в контексті Інтернету речей (IoT). Вони розглядають специфічні виклики, пов'язані з безпекою IoT, і те, як ШІ може бути використаний для їх вирішення. Основні акценти зроблено на методах виявлення аномалій, прогнозуванні загроз та забезпеченні цілісності даних у мережах IoT.

Досить популярним за рівнем цитування у науковій спільності є колективне дослідження Гупти М., Акири К., Арьял К., Паркери Е. та Прахарадж Л. [16], які аналізують вплив генеративного ШІ, такого як ChatGPT, на кібербезпеку та приватність. Автори досліджують, як генеративні моделі можуть бути використані як для захисту, так і для здійснення кібератак. Вони обговорюють потенційні загрози та можливості, які надають ці технології, а також розглядають етичні аспекти та питання приватності.

Коло наукових інтересів Ясін А. [17] також пов'язане із парадигмальною зміною у кібербезпеці, спричиненою впровадженням ШІ для виявлення загроз та реагування на них. Науковець акцентує увагу на нових методах виявлення загроз, які базуються на машинному навчанні, та їхній здатності швидко адаптуватися до нових загроз. Автор роботи також розглядає питання автоматизації процесів реагування на інциденти.

Найновіші методи та підходи, які використовуються у сфері ШІ та ML для кібербезпеки, відображені у дослідженнях Мохамед Н. [18]. Науковець оцінює їхню ефективність і обмеження, а також робить акцент на міждисциплінарному підході та інтеграції різних технологій для покращення систем кібербезпеки.

Адевусі А., Околі У., Олорунсого Т., Адага Е., Дараоджимба Д. та Обі О. [20] досліджують специфічні виклики, пов'язані з захистом критичної інфраструктури, такі як енергетичні системи, транспорт та комунікації, і те, як ШІ може допомогти у

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

вирішенні цих питань. Вони розглядають конкретні приклади успішного впровадження ШІ та пропонують рекомендації для подальшого розвитку цієї галузі.

Отже, дослідження за темою статті є надзвичайно важливим з кількох ключових причин:

1. Зростання складності та обсягу кіберзагроз, оскільки з кожним роком кіберзагрози стають дедалі складнішими і частішими. Традиційні методи захисту вже не можуть ефективно протистояти новим, складним атакам. Штучний інтелект та машинне навчання здатні аналізувати великі обсяги даних, виявляти складні закономірності та адаптуватися до нових загроз швидше, ніж будь-які інші методи;

2. Застосування ШІ дозволяє автоматизувати багато аспектів кібербезпеки, від виявлення загроз до реагування на інциденти. Це значно підвищує ефективність та швидкість реагування, зменшуючи навантаження на людські ресурси та знижуючи ризик людських помилок;

3. Покращення точності та швидкості виявлення за рахунок аналізу аномальної поведінки в режимі реального часу. Це дозволяє зменшити кількість хибних спрацьовувань та підвищити загальний рівень безпеки;

4. ШІ-системи можуть постійно навчатися та адаптуватися до нових загроз, що дозволяє їм залишатися актуальними та ефективними навіть у швидкозмінному середовищі кібербезпеки;

5. Інтеграція ШІ з іншими сучасними технологіями, такими як IoT та Big Data, створює нові можливості для покращення кібербезпеки. Інтеграція дозволяє здійснювати більш детальний аналіз даних та розробляти комплексні захисні стратегії.

Дослідження за темою «Огляд методів та стратегій кібербезпеки засобами штучного інтелекту» також охоплює важливі етичні та правові аспекти, пов'язані з використанням ШІ в кібербезпеці: питання приватності, прозорості рішень та відповідальності за дії, здійснені на основі рішень ШІ-систем. Також, дослідження цієї теми сприятиме стратегічному розвитку та інноваціям у сфері кібербезпеки, надаючи дослідникам та фахівцям нові інструменти та методи для забезпечення безпеки інформаційних систем.

Отже, з огляду на швидке зростання та еволюцію кіберзагроз, дослідження ролі ШІ в кібербезпеці є надзвичайно актуальним та важливим. Це дозволяє не тільки підвищити ефективність захисту, але й сприяє розвитку нових стратегій та технологій для протидії загрозам у цифрову епоху. Останні статистичні дані свідчать про стрімке зростання ринку засобів ШІ для забезпечення кібербезпеки. Згідно з оцінками MarketsandMarkets [21], обсяг цього ринку з 2019 до 2026 року зростає на 23,3% щороку, збільшуючись з 8,8 млрд до 38,2 млрд доларів США. Цей ріст відображає збільшення інтересу до ШІ як інструменту для підвищення ефективності захисту інформаційних систем.

У звіті IBM X-Force Threat Intelligence Index [22] за 2023 рік зазначено, що 30% всіх кіберінцидентів пов'язані з використанням технологій штучного інтелекту та машинного навчання. IBM також повідомляє, що компанії, які використовують ШІ для забезпечення кібербезпеки, змогли скоротити час виявлення та реагування на інциденти на 27%.

За даними Gartner [23], до 2025 року 60% організацій будуть активно використовувати технології ШІ для виявлення та реагування на кіберзагрози. Gartner також зазначає, що компанії, які впроваджують ШІ у свої стратегії кібербезпеки, демонструють значне підвищення ефективності та точності виявлення загроз.

Statista [24] повідомляє, що у 2022 році обсяг інвестицій у технології ШІ для кібербезпеки становив близько 12,5 млрд доларів США. Очікується, що до 2025 року ці

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

інвестиції зростуть до 25 млрд доларів США. Цей ріст свідчить про важливість та актуальність технологій ШІ у сфері кібербезпеки.

Національний інститут стандартів і технологій США (NIST) [25] активно досліджує застосування ШІ у кібербезпеці. Вони публікують рекомендації та стандарти, які допомагають організаціям впроваджувати ефективні стратегії захисту з використанням технологій ШІ. NIST підкреслює важливість інтеграції ШІ у системи виявлення та реагування на загрози для підвищення загального рівня безпеки.

Отже, аналіз статистичних даних з різних офіційних джерел свідчить про значний потенціал та необхідність впровадження технологій ШІ у сферу кібербезпеки. ШІ дозволяє автоматизувати процеси моніторингу, виявлення та реагування на загрози, що зменшує час реакції на інциденти та підвищує загальний рівень захисту інформаційних систем. Інвестиції у технології ШІ для кібербезпеки продовжують зростати, що підкреслює важливість цих технологій у боротьбі з кіберзагрозами. Аналіз показує, що сучасні методи ШІ, такі як машинне навчання, глибоке навчання та аналіз великих даних, активно впроваджуються у системи кібербезпеки для автоматизації процесів моніторингу та реагування на загрози. Це дозволяє значно скоротити час реакції на інциденти та підвищити загальний рівень безпеки. Однак, незважаючи на значні досягнення, використання ШІ у кібербезпеці стикається з низкою викликів. Серед них — необхідність забезпечення прозорості рішень, прийнятих ШІ, а також захист від потенційних загроз, які можуть бути створені з використанням тих самих технологій ШІ. Наприклад, генеративні моделі, такі як GPT-3, можуть бути використані для створення більш складних фішингових атак або дезінформаційних кампаній.

## ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Дослідження методів та стратегій кібербезпеки із застосуванням ШІ підтверджує значний потенціал цієї технології у боротьбі з кіберзагрозами. ШІ дозволяє ефективно аналізувати великі обсяги даних, виявляти складні закономірності та аномалії, що дає змогу швидше та точніше реагувати на наявні та можливі загрози. Завдяки здатності аналізувати великі обсяги даних у режимі реального часу, ШІ дозволяє значно підвищити точність і швидкість виявлення кіберзагроз, зменшуючи кількість хибних спрацьовувань та підвищуючи загальний рівень безпеки. Використання ШІ дозволяє автоматизувати багато аспектів кібербезпеки, що зменшує навантаження на людські ресурси та знижує ризик людських помилок. Даний процес особливо важливий у контексті швидкозмінного інформаційного середовища, де швидкість реакції на інциденти є критичною. Поєднання ШІ з IoT та Big Data створює нові можливості для покращення кібербезпеки. Інтеграція цих технологій дозволяє здійснювати більш детальний аналіз даних та розробляти комплексні захисні стратегії.

Незважаючи на значні досягнення, використання ШІ у кібербезпеці стикається з низкою викликів, таких як забезпечення прозорості рішень, прийнятих ШІ, а також захист від потенційних загроз, створених з використанням тих самих технологій. Використання ШІ у кібербезпеці порушує важливі питання приватності, прозорості рішень та відповідальності за дії, здійснені на основі рішень ШІ-систем. Важливо розробити етичні та правові стандарти для регулювання використання ШІ у цій сфері.

Подальші наукові дослідження у галузі використання ШІ для підвищення кібербезпеки є необхідними для розробки нових, ще більш ефективних методів захисту, що відповідатимуть сучасним викликам. Інвестиції у технології ШІ для кібербезпеки

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

продовжують зростати, що підкреслює важливість цих технологій у боротьбі з кіберзагрозами.

Отже, дослідження показало, що ШІ є потужним інструментом для підвищення рівня кібербезпеки. Використання ШІ дозволяє не тільки підвищити ефективність захисту, але й сприяє розвитку нових стратегій та технологій для протидії загрозам у цифрову епоху. Інтеграція ШІ у системи кібербезпеки є необхідною для адаптації до нових викликів та забезпечення безпеки інформаційних систем у майбутньому.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Магденко, А. Р., Бучацький, І. О., & Бондаренко, І. О. (2024). *Штучний інтелект: нова зброя у руках кіберзлочинців та шахраїв.* https://ir.lib.vntu.edu.ua//handle/123456789/42455
2. Товстуха, Н. А. (2023). Використання штучного інтелекту для покращення кібербезпеки: за та проти. *Комп'ютерні системи та мережні технології: XIII Міжнар. науково-практ. конф.,* 153–154.
3. Устименко, В. О., & Олішевський І. Г. (2022). Перспективи застосування технологій штучного інтелекту в сфері кібербезпеки. *Молодь: наука та інновації: мат. X Міжнар. науково-тех. Конф. студентів, аспірантів та молодих вчених,* 375–377.
4. Каун, Ю., & Собчук, О. (2024). Штучний інтелект як інструмент кібератак і кібербезпеки. *Міжнародна науково-практична конференція «Проблеми комп'ютерних наук, програмного моделювання та безпеки цифрових систем»,* 40–41.
5. Гуржій, С. В. (2024). Потенціал штучного інтелекту у сфері кібербезпеки. *Матеріали XLII-ої Міжнар. науково-практ. конф. «Сучасні аспекти модернізації науки: стан, проблеми, тенденції розвитку»,* 260–261.
6. Котенко, Д., & Хлапонін, Ю. (2024). Штучний інтелект у системах виявлення і запобігання кібератакам: перспективи та виклики. *Підводні технології: промислова та цивільна інженерія, 1(14),* 48–55. https://doi.org/10.32347/uwt.2024.14.1203
7. Субач, І. Ю., & Власенко, О. В. (2023). Архітектура інтелектуальної SIEM-системи для виявлення кіберінцидентів у базах даних інформаційно-комунікаційних систем військового призначення. *Системи і технології зв'язку, інформатизації та кібербезпеки, 4,* 82–92.
8. Савицька, Л., Коробейнікова, Т., Волос, О., & Тарновський, М. (2023). Метод та засіб моніторингу безпеки в комп'ютерній мережі засобами SIEM. *Інформаційні технології та комп'ютерна інженерія, 58(3),* 22–32.
9. Коробейнікова, Т., & Цар, О. (2023). Аналіз сучасних відкритих систем виявлення та запобігання вторгнень. *Grail of Science, 27,* 317–325.
10. Bondarenko, A., & Statsenko, V. (2024). Using artificial intelligence methods and models to improve expert intrusion detection systems. *Technical sciences, 333(2),* 99–106.
11. Цеба, К. Я. (2024). Огляд сучасних інструментів та технологій виявлення кіберзагроз. *Матеріали XV-ої Міжнародної науково-практичної конференції «Free and Open Source Software».*
12. Бабкін, А. А., & Кудін, О. В. (2020). Огляд нейромережевих моделей систем виявлення вторгнень. *Вчені записки Таврійського національного університету ім. В.І. Вернадського. Серія: Технічні науки, 31(70),* 77–82.
13. Sarker, I. H., Furhad, M. H., Nowrozy, R. (2021). AI-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science, 2(3).*
14. Camacho, N. G. (2024). The Role of AI in Cybersecurity: Addressing Threats in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS), 3(1),* 143–154.
15. Kuzlu, M., Fair, C., Guler, O. (2021). Role of artificial intelligence in the Internet of Things (IoT) cybersecurity. *Discover Internet of things, 1(1),* 7–17.
16. Gupta, M., Akiri, C., Aryal, K., Parker, E., Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access, 11,* 80218–80245. https://doi.org/10.1109/ACCESS.2023.3300381
17. Yaseen, A. (2023). AI-driven threat detection and response: A paradigm shift in cybersecurity. *International Journal of Information and Cybersecurity, 7(12),* 25–43.
18. Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering, 10(2).* https://doi.org/10.1080/23311916.2023.2272358

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

19. Capuano, N., Fenza, G., Loia, V., Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access, 10,* 93575–93600.
20. Adewusi, A. O., Okoli, U. I., Olorunsogo, T., Adaga. E., Daraojimba, D. O., & Obi, O. C. (2024). Artificial intelligence in cybersecurity: Protecting national infrastructure: A USA. *World Journal of Advanced Research and Reviews, 21(1),* 2263–2275.
21. *MarketsandMarkets.* (n. d.). https://www.marketsandmarkets.com
22. *IBM X-Force Threat Intelligence Index.* (n. d.). https://www.ibm.com/reports/threat-intelligence
23. *Gartner.* (n. d.). https://www.gartner.com/en
24. *Statista.* (n. d.). https://www.statista.com
25. *National Institute of Standards and Technology (NIST).* (n. d.). https://www.nist.gov

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

**Olha Lunhol**
PhD in Pedagogical Sciences, Associate Professor,
Associate Professor of the Department of
Operational-search Activities and Information Security
Donetsk State University of Internal Affairs, Kropyvnytskyi, Ukraine
ORCID ID: 0000-0001-8128-0072
olyalungol@gmail.com

# OVERVIEW OF CYBERSECURITY METHODS AND STRATEGIES USING ARTIFICIAL INTELLIGENCE

**Abstract.** In today's world, information technology is rapidly evolving, leading to an increase in both the number and complexity of cyber threats, including phishing, malware, and social engineering attacks. The growth in the quantity and sophistication of cyber threats creates an urgent need to improve methods for protecting information systems. Artificial Intelligence (AI), particularly machine learning and deep learning technologies, shows significant potential in enhancing cybersecurity. This article is dedicated to reviewing contemporary AI-based cybersecurity methods and strategies, as well as evaluating their effectiveness in detecting and countering cyber threats. The paper analyzes recent research by both domestic and international scientists, emphasizing AI's ability to analyze large volumes of data, uncover hidden patterns, predict potential threats, and automate incident response processes. It highlights key research directions, including anomaly detection, threat modeling, incident response automation, and ensuring the interpretability of decisions made by AI systems. Special attention is given to the integration of AI into existing cybersecurity systems and its capacity to adapt to new threats. The article also discusses the main challenges and prospects of applying AI in cybersecurity, including ethical and legal aspects such as privacy issues, decision transparency, and accountability for actions taken based on AI system decisions. Recent statistical data indicate a rapid growth in the market for AI-based cybersecurity tools, underscoring the importance and relevance of this topic in contemporary conditions. The analysis results confirm that using AI allows for automating monitoring, threat detection, and response processes, reducing incident response time and enhancing the overall protection level of information systems. At the same time, implementing AI in cybersecurity faces several challenges, such as ensuring the transparency of AI decisions and protecting against potential threats created using the same technologies. Research in this field promotes strategic development and innovation in cybersecurity, providing researchers and professionals with new tools and methods for ensuring information system security. Thus, given the rapid growth and evolution of cyber threats, studying the role of AI in cybersecurity is extremely relevant and important. It not only enhances protection efficiency but also fosters the development of new strategies and technologies to counter threats in the digital age.

**Keywords:** Artificial Intelligence; cyber security; threats; information space.

## REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Magdenko, A. R., Buchatskyi, I. O., & Bondarenko, I. O. (2024). *Artificial intelligence: a new weapon in the hands of cybercriminals and fraudsters.* https://ir.lib.vntu.edu.ua//handle/123456789/42455.
2. Tovstukha, N. A. (2023). Using artificial intelligence to improve cybersecurity: pros and cons. *Computer systems and network technologies: XIII International Scientific and Practical Conference,* 153–154.
3. Ustymenko, V. O., & Olishevskyi, I. G. (2022). Prospects for the application of artificial intelligence technologies in the field of cybersecurity. *Youth: science and innovations: mat. X Internat. scientific and technical conf. of students, graduate students and young scientists,* 375–377.
4. Kaun, Y., & Sobchuk, O. (2024). Artificial intelligence as a tool for cyberattacks and cybersecurity. *International scientific and practical conference "Problems of computer science, software modeling and security of digital systems",* 40–41.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 1 (25), 2024
ISSN 2663 - 4023

5. Gurzhiy, S. V. (2024). The potential of artificial intelligence in the field of cybersecurity. *Proceedings of the XLIIth International Scientific and Practical Conference "Modern Aspects of Modernization of Science: State, Problems, Development Trends"*, 260–261.
6. Kotenko, D., & Khlaponin, Y. (2024). Artificial Intelligence in Cyber Attack Detection and Prevention Systems: Prospects and Challenges. *Underwater technologies: industrial and civil engineering, 1(14),* 48–55. https://doi.org/10.32347/uwt.2024.14.1203
7. Subach, I. Y., & Vlasenko, O. V. (2023). Architecture of an intelligent SIEM system for detecting cyber incidents in databases of information and communication systems for military purposes. *Systems and technologies of communication, informatization and cybersecurity, 4,* 82–92.
8. Savytska, L., Korobeynikova, T., Volos, O., & Tarnovskyi, M. (2023). Method and means of monitoring security in a computer network by means of SIEM. *Information technology and computer engineering, 58(3),* 22–32.
9. Korobeynikova, T., & Tsar, O. (2023). Analysis of modern open systems for intrusion detection and prevention. *Grail of Science, 27,* 317–325.
10. Bondarenko, A., & Statsenko, V. (2024). Using artificial intelligence methods and models to improve expert intrusion detection systems. *Technical sciences, 333(2),* 99–106.
11. Tseba, K. Y. (2024). An overview of modern tools and technologies for detecting cyber threats. *Proceedings of the XVth International Scientific and Practical Conference "Free and Open Source Software".*
12. Babkin, A. A., & Kudin, O. V. (2020). An overview of neural network models of intrusion detection systems. *Scientific Notes of the V.I. Vernadsky Taurida National University. Series: Technical Sciences, 31(70),* 77–82.
13. Sarker, I. H., Furhad, M. H., Nowrozy, R. (2021). AI-driven cybersecurity: an overview, security intelligence modeling and research directions. *SN Computer Science, 2(3).*
14. Camacho, N. G. (2024). The Role of AI in Cybersecurity: Addressing Threats in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS), 3(1),* 143–154.
15. Kuzlu, M., Fair, C., Guler, O. (2021). Role of artificial intelligence in the Internet of Things (IoT) cybersecurity. *Discover Internet of things, 1(1),* 7–17.
16. Gupta, M., Akiri, C., Aryal, K., Parker, E., Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access, 11,* 80218–80245. https://doi.org/10.1109/ACCESS.2023.3300381
17. Yaseen, A. (2023). AI-driven threat detection and response: A paradigm shift in cybersecurity. *International Journal of Information and Cybersecurity, 7(12),* 25–43.
18. Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. *Cogent Engineering, 10(2).* https://doi.org/10.1080/23311916.2023.2272358
19. Capuano, N., Fenza, G., Loia, V., Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access, 10,* 93575–93600.
20. Adewusi, A. O., Okoli, U. I., Olorunsogo, T., Adaga. E., Daraojimba, D. O., & Obi, O. C. (2024). Artificial intelligence in cybersecurity: Protecting national infrastructure: A USA. *World Journal of Advanced Research and Reviews, 21(1),* 2263–2275.
21. *MarketsandMarkets.* (n. d.). https://www.marketsandmarkets.com
22. *IBM X-Force Threat Intelligence Index.* (n. d.). https://www.ibm.com/reports/threat-intelligence
23. *Gartner.* (n. d.). https://www.gartner.com/en
24. *Statista.* (n. d.). https://www.statista.com
25. *National Institute of Standards and Technology (NIST).* (n. d.). https://www.nist.gov

Додаток Б Методи цифрової розвідки

| ***Метод*** | ***Призначення*** | ***Використання*** | ***Переваги*** | ***Недоліки*** |
|---|---|---|---|---|
| **CYBINT** | Збір та аналіз інформації про кіберзагрози та кіберактивність | Виявлення атак, аналіз кіберінцидентів, захист мереж | Швидка ідентифікація кіберзагроз, автоматизація процесів | Великий обсяг даних для аналізу, залежність від технологій та доступу до інфраструктури |
| **HUMINT** | Збір інформації через людські контакти | Шпигунство, інтерв'ю, вербування | Доступ до недоступної і закритої інформації, людський фактор | Ризик для агентів, обмежена достовірність через людський фактор |
| **TECHINT** | Збір технічної інформації про обладнання та технології | Аналіз обладнання, технологій, об'єктів | Знання про нові технології та потенційні загрози | Дороговартісне обладнання, спеціалізовані знання |
| **SIGINT** | Перехоплення та аналіз електронних сигналів та повідомлень | Перехоплення радіо- і телефонних сигналів, мережевий аналіз | Доступ до зашифрованих або прихованих комунікацій | Потреба у спеціалізованому обладнанні, законодавчі обмеження |
| **MASINT** | Збір і аналіз вимірювань та підписів об’єктів, фізичних характеристик | Виявлення ядерної активності, аналіз хімічних і біологічних загроз | Висока точність при вимірюваннях, можливість автоматизації процесів | Висока вартість обладнання, обмежена можливість застосування в цивільних умовах |
| **FININT** | Збір і аналіз фінансових даних для виявлення підозрілих фінансових операцій | Виявлення тероризму, відмивання коштів, фінансових злочинів | Допомога в боротьбі з фінансовими злочинами, високий рівень виявлення підозрілих транзакцій | Обмежений доступ до фінансових даних, законодавчі обмеження |
| **IMINT** | Збір і аналіз зображень (фото, відео) з різних джерел | Супутниковий моніторинг, аерофотозйомка, дрони | Здатність відстежувати активність на великих територіях, виявлення військових загроз | Залежність від погодних умов, вартість супутникових зйомок та іншого обладнання |

# Додаток В Методологія цифрової розвідки

| ***Етап*** | ***Опис*** | ***Результат*** |
|---|---|---|
| 1. Постановка мети та завдань | Визначення цілей, формулювання гіпотез та завдань дослідження | Чітке розуміння цілей розвідки |
| 2. Збір даних (Data Collection) | Збір інформації з різних джерел (OSINT, SOCMINT, SIGINT, FININT тощо) | Зібрані дані з різних джерел |
| 3. Попередня обробка даних | Очищення, нормалізація та підготовка даних для подальшого аналізу | Якісні, готові для аналізу дані |
| 4. Аналіз даних (Data Analysis) | Використання методів обробки даних, статистики та машинного навчання | Ідентифіковані аномалії, тренди, закономірності |
| 5. Візуалізація та інтерпретація | Відображення результатів у вигляді графіків, карт або діаграм для полегшення інтерпретації | Легкий для розуміння звіт або діаграми |
| 6. Прогнозування та рекомендації | Побудова прогнозів щодо загроз і формування рекомендацій щодо дій | Прогнозування ризиків і рекомендації для дій |
| 7. Оцінка ефективності та оптимізація | Оцінка результатів, виявлення слабких місць і впровадження покращень | Підвищення ефективності методології |

# Додаток Д Інструменти для реалізації SOCMINT

## TweetDeck та Hootsuite

- інструменти для керування соціальними медіа, які широко використовуються для спрощення роботи з акаунтами, моніторингу активності та взаємодії з підписниками. Кожен із них має свої унікальні функції, але їхнє призначення полягає в оптимізації управління соціальними мережами, особливо для професіоналів у сфері маркетингу, аналітики, цифрової розвідки та комунікацій.

## CrowdTangle

- аналітичний інструмент від компанії Meta (Facebook), розроблений для відстеження та аналізу контенту в соціальних медіа. Його основна мета — допомагати медіа, журналістам, аналітикам, дослідникам і маркетологам виявляти та моніторити популярний контент на таких платформах, як Facebook, Instagram, Twitter (X) і Reddit. CrowdTangle широко застосовується для дослідження інформаційних тенденцій, виявлення вірусного контенту та аналізу соціальних настроїв.

## Talkwalker та Brandwatch

- аналітичні платформи для моніторингу та аналізу соціальних медіа, які допомагають бізнесу, маркетологам, аналітикам та дослідникам розуміти ставлення та реакції аудиторії, контролювати репутацію бренду і розробляти ефективні стратегії на основі даних.

## Meltwater

- надає інструменти для моніторингу соціальних медіа, що дозволяють аналізувати згадки, настрої та тенденції у реальному часі. Ця платформа підходить для відстеження інфлюенсерів, аналізу конкурентів і виявлення загальних настроїв аудиторії.

## BuzzSumo

- фокусується на аналізі найпопулярнішого контенту та трендів, що допомагає визначати теми, які найбільше цікавлять аудиторію. Він відображає аналітику за платформами, охоплення та залученість, що корисно для SOCMINT для моніторингу вірусного контенту.

## Crimson Hexagon

- використовує штучний інтелект для аналізу великих обсягів даних у соціальних медіа. Вона відслідковує настрої, уподобання та реакції користувачів у динаміці, що дозволяє глибше розуміти поведінку аудиторії та можливі ризики.

## Sprinklr

- інтегрує дані з різних соціальних мереж і надає інструменти для моніторингу, аналітики та управління комунікацією з аудиторією. Це рішення підходить для комплексного моніторингу великих обсягів контенту в реальному часі.

## NetBase Quid

- аналітичний інструмент, що дозволяє відслідковувати розмови, настрої та тенденції в соціальних медіа. Він також корисний для аналізу кризових ситуацій, репутаційного менеджменту та детального вивчення поведінки аудиторії.

## Keyhole

- відомий своєю здатністю відслідковувати хештеги, ключові слова і кампанії в реальному часі. Він надає візуалізації для оцінки залученості та дозволяє відстежувати інфлюенсерів, що допомагає отримати більш комплексне уявлення про впливовий контент у соціальних мережах.

## Mention

- дозволяє моніторити згадки в режимі реального часу і створювати звіти про настрої аудиторії, що робить його корисним для швидкого виявлення актуальних тем та можливих загроз.

# Додаток Е Інструменти реалізації GEOINT

### MaxMind GeoIP

- один із найпопулярніших інструментів для визначення географічного місця розташування IP-адрес, що дозволяє ідентифікувати локації підозрілих запитів та зловмисників.

### ArcGIS та QGIS

- геоінформаційні системи, що дозволяють візуалізувати дані, пов'язані з кіберзагрозами, та знаходити просторові закономірності.

### Google Earth Engine

- платформа для аналізу та візуалізації геопросторових даних, що дозволяє відстежувати розповсюдження кіберзагроз у реальному часі.

### GeoServer

- платформа з відкритим кодом для обробки геопросторових даних, що дозволяє публікувати і переглядати просторові дані у форматах Web Map Service (WMS), Web Feature Service (WFS), і Web Coverage Service (WCS). GeoServer добре підходить для побудови інтерактивних карт і візуалізації кіберзагроз у реальному часі.

### Sentinel Hub

- платформа для доступу до супутникових зображень з космічних апаратів Sentinel (від Європейського космічного агентства). Sentinel Hub надає API для роботи з зображеннями і дозволяє відстежувати активність та зміни на поверхні землі, що корисно для моніторингу можливих локацій з кіберзагрозами.

### CartoDB

- інструмент для аналізу та візуалізації геопросторових даних. CartoDB дозволяє працювати з великими наборами даних, виконувати аналіз за місцем розташування та виявляти географічні тенденції..

### OpenStreetMap (OSM)

- забезпечує доступ до картографічних даних. Він може бути використаний у поєднанні з іншими інструментами для аналізу навколишнього середовища і моніторингу кіберзагроз.

### Mapbox

- надає SDK та API для інтеграції кастомних карт і геолокаційних даних. Він може бути корисний для візуалізації та створення кастомних карт у додатках, пов’язаних з кібербезпекою.

### Skybox Imaging (тепер Planet Labs)

- сервіс надає високоякісні супутникові зображення, які можна використовувати для моніторингу локацій у реальному часі. Skybox Imaging застосовується для відстеження підозрілих активностей або аномалій, що можуть вказувати на джерела кіберзагроз.

### IP2Location

- онлайн-інструмент для геолокації IP-адрес, який дозволяє швидко визначати країну, регіон, місто та інші параметри. IP2Location особливо корисний для моніторингу підозрілої активності за IP-адресами.

Додаток Ж Програмний код реалізації моделі гібридного виявлення аномалій у мережевому трафіку, використовуючи LSTM та кластеризацію K-Means

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout

# 1. Завантаження і підготовка даних
# Симуляція мережевого трафіку
def generate_network_data(samples=10000, features=10):
    normal_data = np.random.normal(loc=0.5, scale=0.1, size=(samples, features))
    attack_data = np.random.normal(loc=0.8, scale=0.1, size=(int(samples * 0.1), features))
    labels = np.hstack([np.zeros(samples), np.ones(int(samples * 0.1))])
    data = np.vstack([normal_data, attack_data])
    return data, labels

data, labels = generate_network_data()
scaler = MinMaxScaler()
data = scaler.fit_transform(data)

# Перетворення на послідовності для LSTM
sequence_length = 10
X = []
y = []

for i in range(len(data) - sequence_length):
    X.append(data[i:i + sequence_length])
    y.append(labels[i + sequence_length - 1])

X = np.array(X)
y = np.array(y)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 2. Побудова LSTM-моделі
model = Sequential([
    LSTM(64, input_shape=(sequence_length, X.shape[2]), return_sequences=True),
    Dropout(0.2),
    LSTM(32, return_sequences=False),
    Dropout(0.2),
    Dense(1, activation='sigmoid')
])

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
model.summary()
```

```
# 3. Навчання моделі
history = model.fit(X_train, y_train, validation_data=(X_test, y_test), epochs=10, batch_size=64)

# 4. Генерація ембеддінгів для кластеризації
embedding_model = Sequential(model.layers[:-1])  # Використовуємо всі шари, крім останнього
X_embeddings = embedding_model.predict(X)

# 5. Кластеризація K-Means
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(X_embeddings)

# Позначення аномалій
clusters = kmeans.predict(X_embeddings)
anomalies = clusters != np.bincount(clusters).argmax()  # Найбільший кластер вважаємо "нормальним"

# 6. Оцінка результатів
from sklearn.metrics import classification_report

print("LSTM Classification Report:")
print(classification_report(y_test, model.predict(X_test).round()))

print("K-Means Classification Report:")
print(classification_report(y, anomalies.astype(int)))

# 7. Збереження моделі
model.save("hybrid_anomaly_model.h5")
```

## Додаток И Реалізація моделі глибинної нейронної мережі для виявлення вторгнень у мережевий трафік

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout

# Створення моделі
model = Sequential()
model.add(Dense(128, input_dim=30, activation='relu'))
# Вхідний шар + перший прихований
model.add(Dropout(0.3))
# Dropout для регуляризації
model.add(Dense(64, activation='relu'))
# Другий прихований шар
model.add(Dropout(0.3))
model.add(Dense(1, activation='sigmoid'))
# Вихідний шар для бінарної класифікації

# Компіляція моделі
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Навчання
model.fit(X_train, y_train, validation_data=(X_val, y_val), epochs=20, batch_size=32)
```

# Додаток К Реалізація моделі використання трансформерів (BERT) для аналізу текстових даних, пов'язаних із кіберзагрозами

## 1. Установка необхідних бібліотек

pip install transformers datasets torch scikit-learn

## 2. Код для реалізації

```
import torch
from transformers import BertTokenizer, BertForSequenceClassification, Trainer, TrainingArguments
from datasets import Dataset, load_metric
from sklearn.model_selection import train_test_split
import pandas as pd

# 1. Завантаження та підготовка даних
data = {
    "text": [
        "Unauthorized access detected on the system",
        "Potential malware found in the attachment",
        "Unusual login location for user admin",
        "Daily backup completed successfully",
        "No threats detected in the system logs"
    ],
    "label": [1, 1, 1, 0, 0]  # 1 - загроза, 0 - нормальний текст
}
df = pd.DataFrame(data)
# Розділення на тренувальний і тестовий набори
train_texts, test_texts, train_labels, test_labels = train_test_split(
    df["text"], df["label"], test_size=0.2, random_state=42
)
# Конвертація до формату Dataset
train_dataset = Dataset.from_dict({"text": train_texts, "label": train_labels})
test_dataset = Dataset.from_dict({"text": test_texts, "label": test_labels})
# 2. Завантаження токенізатора та моделі
tokenizer = BertTokenizer.from_pretrained("bert-base-uncased")
def tokenize_function(examples):
    return tokenizer(examples["text"], padding="max_length", truncation=True, max_length=128)
tokenized_train = train_dataset.map(tokenize_function, batched=True)
tokenized_test = test_dataset.map(tokenize_function, batched=True)
# 3. Підготовка моделі
model = BertForSequenceClassification.from_pretrained("bert-base-uncased", num_labels=2)
# 4. Налаштування метрики
metric = load_metric("accuracy")
def compute_metrics(eval_pred):
    logits, labels = eval_pred
    predictions = torch.argmax(torch.tensor(logits), dim=1)
```

```
    return metric.compute(predictions=predictions, references=labels)
# 5. Налаштування тренера
training_args = TrainingArguments(
    output_dir="./results",
    evaluation_strategy="epoch",
    save_strategy="epoch",
    logging_dir="./logs",
    num_train_epochs=3,
    per_device_train_batch_size=8,
    per_device_eval_batch_size=8,
    learning_rate=5e-5,
    logging_steps=10,
    load_best_model_at_end=True
)
trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_train,
    eval_dataset=tokenized_test,
    tokenizer=tokenizer,
    compute_metrics=compute_metrics
)
# 6. Тренування моделі
trainer.train()
# 7. Оцінка моделі
eval_results = trainer.evaluate()
print(f"Evaluation results: {eval_results}")
# 8. Застосування моделі для аналізу нових даних
def predict_new_texts(texts):
    tokenized_texts = tokenizer(texts, padding="max_length", truncation=True, max_length=128,
return_tensors="pt")
    outputs = model(**tokenized_texts)
    predictions = torch.argmax(outputs.logits, dim=1)
    return predictions.numpy()
# Приклад застосування
new_texts = [
    "Critical vulnerability discovered in the network",
    "System maintenance scheduled for tonight"
]
predictions = predict_new_texts(new_texts)
print(f"Predictions for new texts: {predictions}")
```