Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ДОПУСТИТИ ДО ЗАХИСТУ:
ВО завідувач кафедри
кібербезпеки
та захисту інформації
_____ Іван ПАРХОМЕНКО
“19” травня 2025 р.

ПОЯСНЮВАЛЬНА ЗАПИСКА

кваліфікаційної роботи

магістра
(назва освітнього ступеня)

галузь знань 12 Інформаційні технології
(шифр і назва галузі знань)

спеціальність 125 «Кібербезпека»
(код і назва спеціальності)

освітній ступень бакалавр

освітня програма Кібербезпека
(назва освітньої програми)

на тему: «Метод обробки інцидентів інформаційної безпеки з використанням штучного інтелекту»

Виконавець: студент 2 курсу, групи КБм-21

_____ Віталій КОСТЮЧЕНКО
(підпис) (ім’я прізвище)

	Прізвище, ініціали	**Підпис**
Керівник роботи	Іван ПАРХОМЕНКО	

Нормоконтроль	Лариса МИРУТЕНКО	

Київ 2025

Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ЗАТВЕРДЖЕНО:
ВО завідувач кафедри
кібербезпеки
та захисту інформації
_____ Іван ПАРХОМЕНКО
«25» жовтня 2024 р.

ЗАВДАННЯ
на виконання кваліфікаційної роботи

спеціальності 125 Кібербезпека та захист інформації
(код і назва спеціальності)

освітній ступень магістр

здобувача **КБм-21** (група) **Костюченку Віталію Сергійовичу** (прізвище ім'я по батькові)

Тема кваліфікаційної роботи Метод обробки інцидентів інформаційної безпеки з використанням штучного інтелекту

1. ПІДСТАВИ ДЛЯ ПРОВЕДЕННЯ РОБОТИ

Рішення засідання кафедри кібербезпеки та захисту інформації факультету інформаційних технологій протокол № 4 від 24.10.2024 р

2. МЕТА ТА ВИХІДНІ ДАНІ ДЛЯ ПРОВЕДЕННЯ РОБІТ

Об'єкт дослідження процес аналізу та реалізації механізмів обробки інцидентів інформаційної безпеки з використанням штучного інтелекту.

Предмет дослідження набір механізмів застосування штучного інтелекту, зокрема технологій обробки природної мови для автоматизації аналізу та обробки кіберінцидентів

Мета розробка методу обробки інцидентів інформаційної безпеки з використанням штучного інтелекту

Вихідні дані для проведення роботи Методи обробки інцидентів інформаційної безпеки з

використанням штучного інтелекту

3. ОЧІКУВАНІ НАУКОВІ РЕЗУЛЬТАТИ

Наукова новизна полягає у розробці комбінованого методу з використання ШІ у процесі обробки інцидентів іформаційної безпеки, не лише автоматичне формування підсумків інцидентів, що включає але й інтеграцію системи швидкого пошуку технічної інформації

Практична цінність можливість застосування розробленого методу у системах безпеки для часткової автоматизації та покращення часу вирішення інцидентів.

5. ДАТА ВИДАЧІ ЗАВДАННЯ

Дата видачі завдання: 18 листопада 2024 року

Завдання видав _____ (підпис) Іван ПАРХОМЕНКО (ім'я, прізвище)

Завдання прийняв до виконання _____ (підпис) Віталій Костюченко (ім'я, прізвище)

КАЛЕНДАРНИЙ ПЛАН

Найменування етапів робіт	Строки виконання робіт (початок-кінець)
Уточнення постановки задачі	15.10.2024 – 05.11.2024
Аналіз літератури	20.01.2025 – 13.02.2025
Розгляд нормативно-правової та законодавчої бази, що регулює питання обробки інцидентів інформаційної безпеки	15.02.2025 – 17.02.2025
Дослідження можливостей використання штучного інтелекту у сфері кібербезпеки	18.02.2025 – 23.02.2025
Аналіз моделей обробки природньої мови	25.02.2025 – 28.02.2025
Аналіз сучасних рішень обробки інцидентів на основі штучного інтелекту від вендорів у сфері кібербезпеки	08.03.2025 – 11.03.2025
Розробка методу обробки інцидентів інформаційної безпеки з використанням штучного інтелекту	15.03.2025 – 16.04.2025
Опис розробленого методу та практична реалізація	22.04.2025 – 25.04.2025

Найменування етапів робіт	**Строки виконання робіт (початок-кінець)**
Виконання аналізу використання розробленого методу та перевірка ефективності	27.04.2025 – 04.05.2025
Оформлення пояснювальної записки	05.05.2025 – 14.05.2025
Подача пакету документів на розгляд ЕК	19.05.2025

Завдання видав _____ Іван ПАРХОМЕНКО
(підпис) (ім’я, прізвище)

Завдання прийняв до виконання _____ Віталій Костюченко
(підпис) (ім’я, прізвище)

Дата видачі завдання: 25.10.2024 р.

Термін подання кваліфікаційної роботи до ЕК 19.05.2025 р.

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи «Метод обробки інцидентів інформаційної безпеки з використанням штучного інтелекту»: 99 сторінок, 26 рисунків та 3 таблиці. 50 літературних джерел.

Метою роботи є розробка методу обробки інцидентів інформаційної безпеки з використанням штучного інтелекту.

Об'єктом дослідження є процес аналізу та реалізації механізмів обробки інцидентів інформаційної безпеки з використанням штучного інтелекту.

Предметом дослідження є набір механізмів застосування штучного інтелекту, зокрема технологій обробки природної мови, для автоматизації аналізу та обробки кіберінцидентів.

Методи дослідження: теоретичний аналіз літератури та існуючих підходів до обробки інцидентів кібербезпеки; аналіз існуючих методів та технік обробки інцидентів інформаційної безпеки; аналіз механізмів використання штучного інтелекту під час обробки інцидентів; практична реалізація та експериментальне тестування запропонованих рішень.

Практичною цінністю є можливість застосування розробленого методу у системах безпеки для часткової автоматизації та покращення часу вирішення інцидентів.

Наукова новизна: полягає у розробці комбінованого методу з використання ШІ у процесі обробки інцидентів іформаційної безпеки, що включає не лише автоматичне формування підсумків інцидентів, але й інтеграцію системи швидкого пошуку технічної інформації.

Ключові слова **:** Інцидент, штучний інтелект, програмний моудуль.

Апробація роботи: Костюченко В., Табаченко Д., Білоконь І. Зменшення втоми від тривог в SOC: Покращення реагування на інциденти за допомогою автоматизації

та штучного інтелекту. VIII Міжнародна науково-практична конференція «Проблеми кібербезпеки інформаційно-комунікаційних систем» (PCSICS). 2025. С. 109-110

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕННЬ

ШІ – Штучний інтелект

AI – Artificial intelligence

ПЗ – програмне забезпечення

LMM – Large Language Model

NIST - National Institute of Standards and Technology

ISO - International Organization for Standardization

NLP - Natural Language Processing

CVE - Common Vulnerabilities and Exposures

IEC - International Electrotechnical Commission

IDS - Intrusion Detection System

IPS – Intrusion Prevention System

SP – Special Publication

SANS - SysAdmin, Audit, Network and Security

ТЗІ – Технічний захист інформації

НД- Нормативний документ

SIEM - Security information and event management

XDR - Extended Detection and Response

ВСТУП

Актуальність теми

У сучасному цифровому світі інформаційна безпека стала одним із ключових аспектів діяльності як державних установ, так і приватного сектора. Збільшення кількості та складності кіберінцидентів вимагає ефективних методів їх виявлення, аналізу та реагування. Традиційні підходи до обробки інцидентів, що базуються на ручному аналізі та експертних оцінках, часто виявляються недостатньо швидкими та ефективними в умовах сучасних загроз [1]. Це зумовлює необхідність впровадження новітніх технологій, серед яких штучний інтелект (ШІ) займає провідне місце.

Застосування ШІ в кібербезпеці відкриває нові можливості для аналізу великих обсягів даних, автоматизації виявлення аномалій та прогнозування потенційних загроз. Зокрема, технології обробки природної мови (Natural Language Processing, NLP) та машинного навчання можуть значно підвищити швидкість та точність аналізу інцидентів [2]. Важливо розуміти, що ШІ не може повністю замінити людину у процесі прийняття рішень, але він є незамінним інструментом для автоматизації рутинних завдань, таких як формування Incident Summary та пошук релевантної технічної інформації.

Мета та завдання дослідження

Метою роботи є розробка методу обробки інцидентів інформаційної безпеки з використанням штучного інтелекту та практична реалізація застосунку для покарщення процесу обробки інцидентів за допомогою вирішення наступних задач :

1. Формування узагальненої інформації на основі вхідних логів та технічних даних за допомогою NLP-моделей.
2. Автоматизований пошук технічної інформації про виявлений інцидент (наприклад, пошук релевантних статей, CVE-записів, документації) за допомогою Штучного інтелекту.

3. Підбиття стандартизованих підсумків розслідування інциденту відповідно до ключових методик життєвого циклу інциденту згідно рекомендацій найкращих практик.

Для досягнення цієї мети необхідно виконати наступні **завдання**:

- Дослідити ключові стандартів обробки Інцидентів Інформаційної безпеки
- Проаналізувати рішення обробки інцидентів на основі штучного інтелекту від
- вендорів у сфері кібербезпеки
- Проаналізувати наявні LMM-моделі штучного інтелекту
- Розробити метод та реалізувати його у програмному модулі
- Проаналізувати результатів використання розробленого методу

Об’єктом дослідження є процес аналізу та реалізації методів обробки інцидентів інформаційної безпеки з використанням штучного інтелекту.

Предметом дослідження є набір механізмів застосування штучного інтелекту, зокрема технологій обробки природної мови, для автоматизації аналізу та обробки кіберінцидентів.

У процесі роботи використовуватимуться наступні **методи дослідження** :

Теоретичний аналіз літератури та існуючих підходів до обробки інцидентів кібербезпеки;

Аналіз існуючих методів та технік обробки інцидентів інформаційної безпеки ;

Аналіз механізмів використання штучного інтелекту під час обробки інцидентів;

Практична реалізація та експериментальне тестування запропонованих рішень.

Наукова новизна роботи полягає в у розробці комбінованого методу з використання ШІ у процесі обробки інцидентів іформаційної безпеки, що включає не лише автоматичне формування підсумків інцидентів, але й інтеграцію системи швидкого пошуку технічної інформації.

Практична цінність можливість застосування розробленого методу у системах безпеки для часткової автоматизації та покращення часу вирішення інцидентів.

Ключові слова : Інцидент, штучний інтелект, програмний моудуль.

Апробація роботи: Костюченко В., Табаченко Д., Білоконь І. Зменшення втоми від тривог в SOC: Покращення реагування на інциденти за допомогою автоматизації та штучного інтелекту. VIII Міжнародна науково-практична конференція «Проблеми кібербезпеки інформаційно-комунікаційних систем» (PCSICS). 2025. С. 109-110

РОЗДІЛ 1
МЕТОДИ ОБРОБКИ ІНЦИДЕНТІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ

1.1 Основні підходи до обробки інцидентів

Традиційні моделі реагування на кіберінциденти протягом багатьох років базуються на одних і тих же фундаментальних принципах: виявлення інциденту, його зупинка або обмеження впливу, а потім—відновлення нормального функціонування. Подальший аналіз інциденту застосовується для визначення потенційних покращень у безпеці та процесах організації. В Україні відсутній єдиний, чітко визначений стандарт реагування на інциденти інформаційної безпеки, закріплений на законодавчому рівні. Закони України "Про основні засади кібербезпеки України" та "Про інформацію" встановлюють загальні принципи та вимоги щодо захисту інформації та кібербезпеки, проте не містять деталізованого, поетапного плану дій у випадку виникнення інцидентів. У цьому контексті, одним з найближчих нормативних документів, який можна розглядати як орієнтир, є НД ТЗІ 2.5-004-99 "НОРМАТИВНИЙ ДОКУМЕНТ СИСТЕМИ ТЕХНІЧНОГО ЗАХИСТУ ІНФОРМАЦІЇ Критерії оцінки захищеності інформації в комп’ютерних системах від несанкціонованого доступу". Хоча цей документ насамперед фокусується на критеріях оцінки захищеності інформації від несанкціонованого доступу, він також зачіпає аспекти виявлення та реагування на події безпеки, які потенційно можуть перерости в інциденти. У НД ТЗІ 2.5-004-99 не йдеться безпосередньо про реагування на інциденти в сучасному розумінні цього процесу, однак він закладає основи підходу до виявлення і нейтралізації несанкціонованого доступу, який є типовим прикладом інциденту інформаційної безпеки. Цей нормативний документ фактично слугує базовим інструментом для оцінки рівня захищеності інформаційних систем та встановлення загроз, які можуть бути пов’язані з інцидентами. Саме тому він, попри застарілість (датований 1999 роком), продовжує

згадуватись у сучасних методичних рекомендаціях Держспецзв’язку щодо створення та сертифікації комплексної системи захисту інформації (КСЗІ).

Ці рекомендації, серед іншого, акцентують на необхідності створення механізмів виявлення, реєстрації та реагування на спроби порушення безпеки, що є функціональними складовими процесу управління інцидентами. Вони передбачають наявність системного журналювання, моніторингу подій, а також організаційних заходів для реагування на виявлені аномалії чи порушення. Водночас чітко визначеного формату сценаріїв реагування або стандартного процесу, аналогічного ISO/IEC 27035, в українських нормативних документах немає.

У зв’язку з цим українські організації, особливо ті, що підпадають під категорію об’єктів критичної інформаційної інфраструктури, змушені орієнтуватися на міжнародні стандарти та практики. Найбільш релевантними у цьому контексті є стандарти серії ISO/IEC 27000, зокрема ISO/IEC 27035, який детально описує структуру процесу реагування на інциденти, включаючи підготовку, виявлення, аналіз, реагування, відновлення та постінцидентний розбір.

Деякі підприємства також беруть за основу керівництва від ENISA (Агентства ЄС з кібербезпеки) або методології NIST, які мають більш деталізовану структуру обробки інцидентів, що дозволяє досягти вищого рівня узгодженості у діях персоналу та кращої керованості ризиками.

Посилаючись на новітню інформацію публікації слів Олександра Потія [4] – «Державна служба спеціального зв’язку та захисту інформації України готова відійти від комплексної системи захисту інформації (КСЗІ) і перейти на декларування відповідності систем захисту інформації в інформаційних, електронних комунікаційних та інформаційно-комунікаційних системах… У 2025 році Держспецзв’язку прагне перейти на повноцінне запровадження системи декларативного принципу під час побудови захисту інформаційних систем на основі базових профілів безпеки, які базуються на провідних світових стандартах, зокрема Risk Management Framework американського Національного інституту стандартів і технології NIST», – сказав Олександр Потій. Після чого Верховна Рада України

ухвалила [3] у другому читанні Закон України «Про внесення змін до деяких законів України щодо захисту інформації та кіберзахисту державних інформаційних ресурсів, об’єктів критичної інформаційної інфраструктури», у якому ухвалила відмову від КСЗІ.

Очевидно, що Україна намагається наслідувати найкращі практики західних країн, дійсно, такі підходи ефективно ілюструються класичними моделями, запропонованими Інститутом SANS у 2003 році та Національним інститутом стандартів і технологій США (NIST) у 2012 році, які фокусуються на циклах швидкого реагування та вдосконалення захисних механізмів.

Однак подібні традиційні підходи не враховують важливість отримання розвідувальної інформації щодо самих атакуючих та їх мотивів, хоча така інформація є критично важливою для довгострокового покращення кібербезпеки. Наприклад, з точки зору військової чи бізнес-розвідки, метою є не лише нейтралізація атаки, а й розуміння атакуючої сторони: хто саме атакує, з якими мотивами та цілями, які ресурси та методи використовує. Відсутність такої інформації створює дефіцит знань («інформаційний голод»), що послаблює організацію в майбутньому.

Штучний інтелект здатний значно допомогти подолати ці недоліки традиційних підходів. Зокрема, технології NLP можуть автоматично опрацьовувати великі об'єми неструктурованих даних (наприклад, звіти про інциденти, форуми, соціальні мережі), виокремлюючи важливі деталі щодо мотивів і цілей атакуючих. Машинне навчання, своєю чергою, здатне ідентифікувати закономірності в поведінці зловмисників, прогнозувати їхні подальші кроки, що покращує ситуаційну обізнаність організації. Також, експертні системи можуть узагальнювати накопичені знання і створювати базу для оперативного реагування, знижуючи ризик помилок і підвищуючи ефективність прийняття рішень.

Обробка інцидентів інформаційної безпеки є одним із найважливіших процесів у забезпеченні стійкості організації до кіберзагроз. Мета цього процесу полягає у своєчасному виявленні, аналізі, нейтралізації та подальшому запобіганні

повторних інцидентів. Однак, в умовах постійно зростаючої кількості, складності та варіативності кіберінцидентів, традиційні ручні підходи все частіше втрачають ефективність, потребуючи доповнення інноваційними технологіями, зокрема, штучним інтелектом (ШІ).

Життєвий цикл обробки інцидентів, що пропонується Національним інститутом стандартів і технологій США (NIST), включає чотири основні етапи [5]:

Підготовка (Preparation) – на цьому етапі відбувається розробка політик інформаційної безпеки, підготовка процедур реагування, впровадження систем моніторингу та виявлення загроз, а також регулярне навчання персоналу. Особлива увага приділяється профілактиці інцидентів шляхом управління ризиками та усунення вразливостей.

Виявлення та аналіз (Detection & Analysis) – цей етап включає постійний моніторинг стану інформаційних систем, збір та аналіз логів, ідентифікацію підозрілої активності та підтвердження факту інциденту. Аналіз передбачає визначення масштабу та типу інциденту, класифікацію рівня загрози, оцінку потенційних наслідків та початкову реакцію.

Реагування (Containment, Eradication & Recovery) – на цьому етапі здійснюються заходи з локалізації та нейтралізації інциденту, включаючи тимчасове ізолювання заражених систем або сегментів мережі. Після локалізації загрози проводиться ретельний пошук та усунення джерела інциденту, відновлення інформаційних систем до нормального стану та аналіз виявлених вразливостей для запобігання подальшим атакам.

Навчання на досвіді (Post-Incident Activity) – після завершення реагування здійснюється аналіз отриманого досвіду, проводиться детальний огляд причин виникнення інциденту, ефективності реагування та необхідності вдосконалення існуючих процедур і систем безпеки. Результатом цього етапу є рекомендації щодо змін у політиках безпеки та процесах реагування.

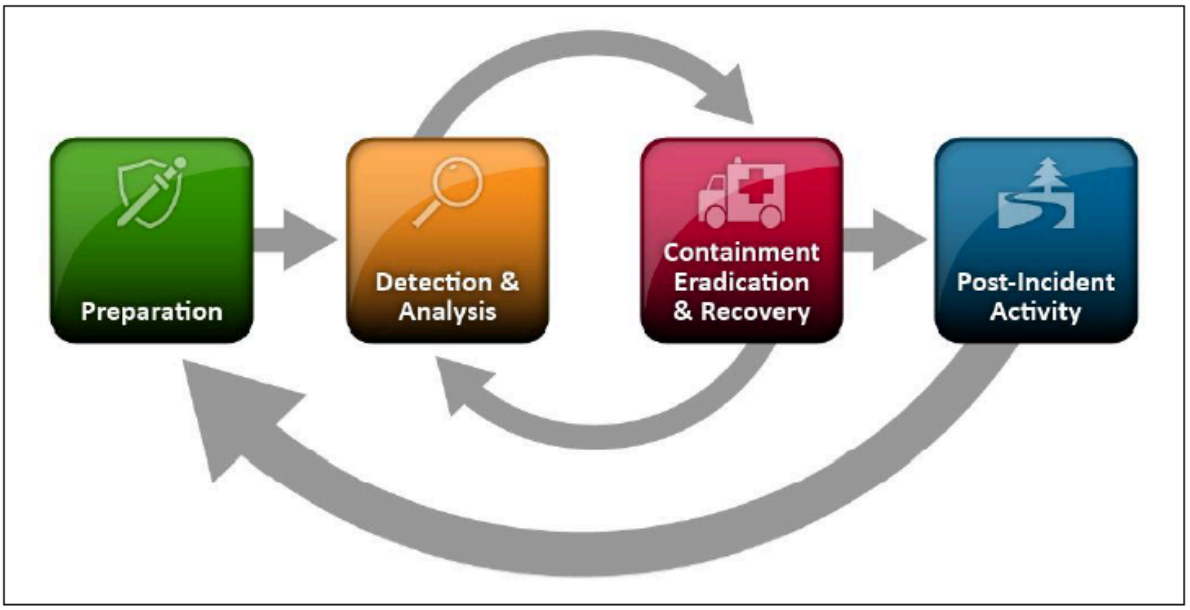


Рисунок 1.1 – Процес обробки інциденту NIST SPECIAL PUBLICATION 800-61

Дотримання зазначеного циклу забезпечує системність та ефективність роботи організації з обробки кіберінцидентів, допомагаючи знизити потенційні ризики та втрати.

Традиційні методи включають наступні основні підходи:

Ручний аналіз – передбачає детальний перегляд та інтерпретацію логів, журналів подій, мережевих трафіків, інформації системного моніторингу експертами з кібербезпеки. Цей процес передбачає значні затрати часу і ресурсів, вимагаючи високої кваліфікації та досвіду спеціалістів. Перевагою методу є можливість глибокого аналізу та точного розуміння інцидентів, однак через людський фактор є ризик пропуску деталей та помилок.

Підхід на основі сигнатур (Signature-based approach) – базується на використанні заздалегідь створених підписів (сигнатур) відомих загроз і атак. Такі сигнатури описують характерні ознаки відомого шкідливого програмного забезпечення (malware), типових кібератак (наприклад, SQL-ін'єкції або Cross-Site Scripting). Системи виявлення вторгнень (IDS) та системи запобігання вторгненням (IPS) використовують цей підхід для швидкого реагування на вже відомі загрози.

Головний недолік – неможливість ефективного виявлення нових або модифікованих атак, які не відповідають наявним сигнатурам.

Евристичний аналіз – використовує методики визначення аномальної поведінки інформаційних систем на основі попередньо встановлених правил та шаблонів. Наприклад, евристичний аналіз може визначати нехарактерні об'єми мережевого трафіку, незвичайні запити до баз даних або несанкціоновані спроби доступу до системи. Перевагою такого підходу є можливість виявлення нових та раніше невідомих загроз. Проте недоліком є велика кількість хибних спрацьовувань (false positives), які створюють додаткове навантаження на команди реагування, потребуючи додаткового аналізу. Ці традиційні методи потребують доповнення сучасними інструментами для покращення оперативності та точності реагування.

Згідно з NIST SP 800-61 Revision 2 "Computer Security Incident Handling Guide", перший етап процесу обробки інцидентів інформаційної безпеки називається Preparation (Підготовка). Цей етап є критично важливим, оскільки закладає основу для ефективного виявлення, аналізу, стримування, ліквідації та відновлення після інцидентів. Ефективна підготовка значно підвищує здатність організації швидко та результативно реагувати на загрози інформаційній безпеці, мінімізуючи потенційні збитки та час простою.

Етап Preparation включає в себе розробку та впровадження організаційних, технічних та процедурних заходів, необхідних для ефективного управління інцидентами. Одним з ключових аспектів є розробка та документування політик та процедур обробки інцидентів. Політика обробки інцидентів є документом високого рівня, який визначає загальний підхід організації до управління інцидентами інформаційної безпеки. Вона має включати визначення інциденту інформаційної безпеки, цілі та завдання процесу обробки інцидентів, ролі та відповідальності різних підрозділів та осіб (наприклад, групи реагування на інциденти, керівництва, юридичного відділу, відділу комунікацій), порядок взаємодії з зовнішніми сторонами (правоохоронними органами, постачальниками послуг, клієнтами), критерії ескалації інцидентів, вимоги до документування інцидентів та порядок проведення аналізу

після інциденту та внесення покращень. Процедури обробки інцидентів, у свою чергу, є детальними покроковими інструкціями щодо виконання конкретних дій на різних етапах обробки інциденту. Вони можуть включати процедури виявлення та повідомлення про підозрілі події, процедури первинної оцінки та класифікації інцидентів за рівнем серйозності та потенційним впливом, процедури стримування інцидентів (наприклад, ізоляція уражених систем, блокування мережевого трафіку), процедури ліквідації інцидентів (наприклад, видалення шкідливого програмного забезпечення, відновлення систем з резервних копій), процедури відновлення після інцидентів (наприклад, відновлення даних, відновлення працездатності систем), процедури документування кожного етапу обробки інциденту та процедури комунікації щодо інцидентів (внутрішньої та зовнішньої).

Важливим елементом етапу Preparation є створення та підготовка групи реагування на інциденти (Incident Response Team - IRT). До складу IRT повинні входити представники різних підрозділів організації, які володіють необхідними знаннями та навичками (наприклад, фахівці з інформаційної безпеки, системні адміністратори, мережеві інженери, представники юридичного відділу, відділу комунікацій), з чітким визначенням ролей та відповідальностей кожного члена команди (наприклад, керівник команди, аналітик, комунікатор). Забезпечення необхідного навчання та підготовки, включаючи регулярні тренінги та симуляції інцидентів (наприклад, "червона команда" проти "синьої команди"), є критично важливим для ефективної роботи команди. Також необхідно забезпечити команду надійними та захищеними каналами зв'язку для координації дій під час інцидентів (наприклад, виділені канали зв'язку, системи миттєвих повідомлень, конференц-зв'язок).

На етапі підготовки здійснюється закупівля та налаштування необхідних інструментів та ресурсів. До них належать системи виявлення вторгнень (Intrusion Detection Systems - IDS) та системи запобігання вторгненням (Intrusion Prevention Systems - IPS) для автоматизованого виявлення та блокування підозрілої активності, системи управління інформацією та подіями безпеки (Security Information and Event

Management - SIEM) для централізованого збору, аналізу та кореляції журналів подій з різних джерел для виявлення інцидентів, інструменти для аналізу шкідливого програмного забезпечення (Malware Analysis Tools) для дослідження виявлених зразків шкідливого коду, інструменти для криміналістичного аналізу (Forensic Tools) для збору та аналізу цифрових доказів, інструменти для відновлення даних (Data Recovery Tools) для відновлення інформації з резервних копій, захищені сховища для зберігання доказів та інформації про інциденти, а також контактна інформація зовнішніх експертів та організацій (наприклад, CERT, постачальники послуг безпеки, юридичні консультанти). Важливим є визначення та документування критичних активів організації та їхньої цінності, що допомагає приоритизувати зусилля під час обробки інцидентів та визначити критичність інциденту. Створення та підтримка планів резервного копіювання та відновлення, включаючи регулярне створення резервних копій критичних даних та конфігурацій систем, а також розробку та тестування планів відновлення, є невід'ємною частиною підготовки, забезпечуючи можливість швидкого відновлення працездатності систем після інцидентів.

Проведення оцінки ризиків та визначення потенційних сценаріїв інцидентів, включаючи ідентифікацію потенційних загроз та вразливостей, які можуть призвести до інцидентів інформаційної безпеки, та розробку можливих сценаріїв інцидентів (наприклад, фішингова атака, атака типу "відмова в обслуговуванні", компрометація облікових даних), допомагає організації підготуватися до найбільш вірогідних типів інцидентів та розробити відповідні плани реагування. Встановлення каналів комунікації та процедур обміну інформацією, включаючи визначення відповідальних осіб за комунікацію під час інцидентів (як внутрішню, так і зовнішню) та розробку шаблонів повідомлень та процедур інформування зацікавлених сторін, є важливим для своєчасного інформування. Нарешті, проведення періодичних навчань та симуляцій для співробітників організації з питань розпізнавання та повідомлення про підозрілу активність та для членів IRT для перевірки ефективності розроблених процедур та їхньої підготовленості, є

важливим елементом етапу підготовки. Етап Preparation є безперервним процесом, який потребує регулярного перегляду та оновлення у відповідь на зміни в ландшафті загроз та інфраструктурі організації. Інвестиції в якісну підготовку є ключем до ефективного та успішного управління інцидентами інформаційної безпеки.

Наступним етапом після Preparation є Detection & Analysis (Виявлення та Аналіз). Цей етап є критично важливим для своєчасного виявлення інцидентів інформаційної безпеки та розуміння їхньої природи, масштабу та потенційного впливу. Ефективне виявлення та ретельний аналіз дозволяють організації вжити адекватних заходів для стримування, ліквідації та відновлення після інциденту, мінімізуючи завдані збитки.

Етап Detection & Analysis включає в себе процеси моніторингу та виявлення підозрілої активності, а також подальший аналіз виявлених подій для визначення, чи є вони фактичним інцидентом інформаційної безпеки, і якщо так, то якого типу, масштабу та потенційного впливу. Процес виявлення може ініціюватися різними джерелами, включаючи автоматизовані системи моніторингу, повідомлення користувачів, внутрішні аудити та зовнішні повідомлення.

Одним з ключових аспектів етапу виявлення є моніторинг та аналіз подій безпеки. Це включає в себе постійний моніторинг журналів подій з різних систем (операційних систем, мережевого обладнання, застосунків, систем безпеки), аналіз мережевого трафіку, виявлення аномальної поведінки користувачів та систем. Системи управління інформацією та подіями безпеки (SIEM) відіграють важливу роль у цьому процесі, забезпечуючи централізований збір, кореляцію та аналіз великих обсягів даних для виявлення потенційних інцидентів. Також важливим є використання систем виявлення вторгнень (IDS) та систем запобігання вторгненням (IPS) для автоматизованого виявлення та блокування відомих загроз.

Іншим важливим джерелом виявлення інцидентів є повідомлення користувачів. Співробітники організації повинні бути навчені розпізнавати підозрілу активність (наприклад, фішингові листи, незвичайні запити, несанкціонований доступ) та повідомляти про неї відповідальним особам або до групи реагування на

інциденти. Наявність чітких та простих процедур повідомлення є критично важливим для своєчасного виявлення інцидентів.

Після виявлення потенційної події безпеки розпочинається етап аналізу. Метою аналізу є визначення, чи є ця подія фактичним інцидентом інформаційної безпеки. Цей процес включає в себе збір додаткової інформації про подію, її контекст, залучені системи та дані. Група реагування на інциденти проводить первинну оцінку для визначення серйозності події та її потенційного впливу на бізнес.

На етапі аналізу здійснюється класифікація інциденту. Визначення типу інциденту (наприклад, атака шкідливого програмного забезпечення, несанкціонований доступ, атака типу "відмова в обслуговуванні", витік даних) є важливим для вибору відповідних процедур реагування. Також визначається рівень серйозності інциденту на основі його потенційного впливу на конфіденційність, цілісність та доступність інформаційних активів організації. Критерії класифікації повинні бути чітко визначені на етапі підготовки.

Важливим аспектом аналізу є збір та аналіз доказів. Зібрані дані можуть включати журнали подій, мережевий трафік, образи дисків уражених систем, інформацію про облікові записи користувачів та інші релевантні дані. Аналіз цих доказів допомагає зрозуміти причини інциденту, методи атаки, залучені системи та обсяг завданої шкоди. Для цього можуть використовуватися спеціалізовані інструменти криміналістичного аналізу.

Протягом етапу аналізу важливо документувати всі дії та знахідки. Детальне документування процесу аналізу, зібраних доказів, зроблених висновків та прийнятих рішень є критично важливим для подальших етапів обробки інциденту, а також для проведення аналізу після інциденту та внесення покращень до системи безпеки.

Ефективний етап Detection & Analysis вимагає наявності добре підготовленої групи реагування на інциденти, належних інструментів та технологій моніторингу та

аналізу, а також чітких процедур виявлення, повідомлення та первинної оцінки інцидентів.

Після етапу Detection & Analysis настає етап, який часто розглядається як три взаємопов'язані фази: Containment (Стримування), Eradication (Ліквідація) та Recovery (Відновлення). Ці етапи спрямовані на обмеження шкоди від інциденту, усунення загрози та відновлення нормального функціонування систем та бізнес-процесів.

Containment (Стримування) має на меті запобігти подальшому поширенню інциденту та мінімізувати його вплив на організацію. Стратегії стримування можуть варіюватися залежно від типу та масштабу інциденту. Загальні підходи включають ізоляцію уражених систем від мережі, сегментацію мережі для запобігання горизонтальному переміщенню зловмисників, блокування певного мережевого трафіку або портів, тимчасове виведення з експлуатації скомпрометованих сервісів або систем. Прийняття рішень щодо стратегії стримування повинно враховувати потенні збитки для бізнесу від застосування тих чи інших заходів, необхідність збереження доказів для подальшого аналізу та можливість повного припинення функціонування критично важливих систем. Важливо ретельно документувати всі вжиті заходи зі стримування.

Eradication (Ліквідація) передбачає усунення причини інциденту та видалення всіх компонентів загрози з уражених систем. Це може включати видалення шкідливого програмного забезпечення, виправлення вразливостей, скидання або зміну скомпрометованих облікових даних, відновлення систем до відомого безпечного стану (наприклад, шляхом перевстановлення операційної системи або відновлення з резервної копії). Перед початком ліквідації важливо переконатися, що процес стримування був успішним і загроза не поширюється далі. Також необхідно ретельно спланувати процес ліквідації, щоб уникнути втрати важливих даних або порушення стабільності систем. Після завершення ліквідації слід провести перевірку для підтвердження повного усунення загрози.

Recovery (Відновлення) фокусується на відновленні нормального функціонування уражених систем та бізнес-процесів. Цей етап може включати відновлення даних з резервних копій, відновлення працездатності сервісів, повернення систем в онлайн-режим. Процес відновлення повинен бути пріоритезованим на основі критичності систем та бізнес-потреб. Важливо провести тестування відновлених систем для перевірки їхньої працездатності та безпеки. Після відновлення слід здійснювати посилений моніторинг відновлених систем для виявлення будь-яких ознак повторної компрометації або залишкової шкідливої активності. Процес відновлення також повинен бути ретельно задокументований, включаючи всі виконані кроки та виявлені проблеми.

Важливо зазначити, що ці три етапи часто є ітеративними та можуть відбуватися паралельно. Наприклад, стримування може бути застосоване на початковому етапі, потім може відбуватися аналіз для кращого розуміння загрози, що може призвести до коригування стратегії стримування. Ліквідація може розпочатися на одних уражених системах, поки на інших ще триває аналіз. Відновлення може починатися після ліквідації на певних системах, тоді як на інших ще тривають заходи з ліквідації. Комунікація між членами групи реагування на інциденти та іншими зацікавленими сторонами є критично важливою протягом усіх цих етапів для забезпечення скоординованих та ефективних дій.

Останнім основним етапом процесу обробки інцидентів інформаційної безпеки є Post-Incident Activity (Діяльність після інциденту). Цей етап є надзвичайно важливим для вилучення уроків з інциденту, вдосконалення процесів безпеки та реагування, а також для запобігання подібним інцидентам у майбутньому. Ключовим елементом цього етапу є підбиття підсумків інциденту, яке часто називають "lessons learned" або "post-mortem". Метою підбиття підсумків інциденту є детальний аналіз інциденту після його успішного стримування, ліквідації та відновлення. Завдання полягає не в пошуку винних, а в об'єктивному вивченні подій, виявленні слабких місць у системах безпеки та процесах реагування, а також у визначенні конкретних кроків для покращення. Результати підбиття підсумків інциденту мають бути

задокументовані у вигляді звіту, який розповсюджується серед відповідних зацікавлених сторін. Згідно з NIST SP 800-61 Revision 2 "Computer Security Incident Handling Guide", останнім основним етапом процесу обробки інцидентів інформаційної безпеки є Post-Incident Activity (Діяльність після інциденту). Цей етап є надзвичайно важливим для вилучення уроків з інциденту, вдосконалення процесів безпеки та реагування, а також для запобігання подібним інцидентам у майбутньому. Ключовим елементом цього етапу є підбиття підсумків інциденту, яке часто називають "lessons learned". Метою підбиття підсумків інциденту є детальний аналіз інциденту після його успішного стримування, ліквідації та відновлення. Завдання полягає не в пошуку винних, а в об'єктивному вивченні подій, виявленні слабких місць у системах безпеки та процесах реагування, а також у визначенні конкретних кроків для покращення. Результати підбиття підсумків інциденту мають бути задокументовані у вигляді звіту, який розповсюджується серед відповідних зацікавлених сторін.

Першим кроком є створення звіту про підсумки інциденту (follow-up report). Цей документ має бути всеосяжним та детально описувати всі аспекти інциденту та процесу його обробки. Він слугує офіційним записом події та основою для подальшого аналізу та впровадження покращень, які необхідно виконати. Звіт повинен містити докладний опис інциденту, включаючи час виявлення та тривалість, тип атаки, уражені системи та дані, а також вектор атаки. Важливо провести глибокий аналіз причин інциденту, виявивши використані вразливості, недоліки в конфігурації, політиках безпеки або людському факторі. Звіт також повинен містити оцінку впливу інциденту на бізнес, включаючи фінансові втрати, репутаційні ризики та порушення операційної діяльності. Окремий розділ звіту має бути присвячений оцінці ефективності дій команди реагування на кожному етапі, висвітлюючи як успішні моменти, так і області, що потребують покращення. Кульмінацією звіту є формулювання вилучених уроків та конкретних рекомендацій щодо усунення виявлених слабких місць та вдосконалення існуючих процесів. NIST надає

інформацію, яка наведена у Таблиці 1.1, щодо контрольного списку необхідної інформації у підсумку інциденту.

Таблиця 1.1

Контрольний список обробки інцидентів згідно NIST SP 800-61R2

	Detection and Analysis	
1	Determine whether an incident has occurred	
1.1	Analyze the precursors and indicators	
1.2	Look for correlating information	
1.3	Perform research (e.g., search engines, knowledge base)	
1.4	As soon as the handler believes an incident has occurred, begin documenting the investigation and gathering evidence	
2	Prioritize handling the incident based on the relevant factors (functional impact, information impact, recoverability effort, etc.)	
3	Report the incident to the appropriate internal personnel and external organizations	
	Containment, Eradication, and Recovery	
4	Acquire, preserve, secure, and document evidence	
5	Contain the incident	
6	Eradicate the incident	
6.1	Identify and mitigate all vulnerabilities that were exploited	
6.2	Remove malware, inappropriate materials, and other components	
6.3	If more affected hosts are discovered (e.g., new malware infections), repeat the Detection and Analysis steps (1.1, 1.2) to identify all	

	other affected hosts, then contain (5) and eradicate (6) the incident for them	
7	Recover from the incident	
7.1	Return affected systems to an operationally ready state	
7.2	Confirm that the affected systems are functioning normally	

продовження таблиці 1.1

7.3	If necessary, implement additional monitoring to look for future related activity	
	Post-Incident Activity	
8	Create a follow-up report	
9	Hold a lessons learned meeting (mandatory for major incidents, optional otherwise)	

Другим важливим кроком є проведення зустрічі для обговорення вилучених уроків (lessons learned meeting). Згідно з наданою інформацією, ця зустріч є обов'язковою для серйозних інцидентів і залишається на розсуд для менш значних. Метою цієї зустрічі є відкрите обговорення інциденту за участю всіх ключових членів команди реагування та інших зацікавлених сторін. Під час зустрічі обговорюються питання про те, що було зроблено правильно, які дії виявилися неефективними, які помилки були допущені та які конкретні кроки необхідно вжити для запобігання подібним інцидентам у майбутньому та оптимізації процесів реагування. Важливо створити атмосферу, в якій кожен учасник може вільно висловлювати свою думку та ділитися спостереженнями без страху бути звинуваченим. Результати цієї зустрічі, включаючи узгоджені кроки для покращення, повинні бути задокументовані та включені до плану дій щодо впровадження рекомендацій, зазначених у звіті про підсумки інциденту. Ефективне виконання етапу Post-Incident Activity є запорукою постійного вдосконалення системи

інформаційної безпеки організації та підвищення її стійкості до майбутніх кіберзагроз.

Теоретично, штучний інтелект та моделі LLM мають можливість допомогти на кожному з етапів життєвого циклу інциденту. На етапі Підготовки Штучний інтелект міг би допомогти з процесом пошуку загроз використовуючи пошукові здібності таких моделей, що персоналізує навчання та розвиток аналітиків та зменшує час на пошук необхідних новин. На етапі Виявлення та Аналізу ШІ без значних недоліків може збагачувати контекст інцидентів при коректному доступі до даних. Етап Containment, Eradication, and Recovery (Стримування, Ліквідація та Відновлення) на мою думку, наразі недоступний для штучного інтелекту, оскільки це процес який має бути контрольований та погоджений людиною аналітиком, використання та довіра до штучного інтелекту на даному етапі може нанести значну шкоду підприємству та бізнес операціям.

На етапі Діяльності після інциденту (Post-Incident Activity) ШІ відіграє ключову роль у вилученні максимальної користі з пережитого досвіду. Він здатен автоматизовано аналізувати великі обсяги даних про інцидент, виявляючи глибинні причини, слабкі місця в системі безпеки та недоліки в процесах реагування з більшою об'єктивністю та швидкістю, ніж ручний аналіз. ШІ може генерувати детальні звіти про підсумки інциденту, автоматично узагальнюючи ключову інформацію, хронологію подій, технічні деталі, вжиті заходи та оцінку їхньої ефективності. Це звільняє час аналітиків для більш стратегічних завдань. Аналізуючи вилучені уроки з минулих інцидентів, ШІ може допомогти в оцінці ефективності впроваджених після інциденту рекомендацій. Застосування ШІ на цьому етапі дозволяє організації не просто реагувати на інциденти, а й активно навчатися на власному досвіді, постійно вдосконалюючи свою систему кібербезпеки та підвищуючи загальну стійкість до майбутніх атак.

1.2 Альтернативні підходи до обробки інцидентів

Огляд літератури здійснювався з двох перспектив. Першою була практична перспектива, яка передбачала аналіз документів щодо найкращих практик у сфері кібербезпеки та суміжних галузях. Другою стала академічна перспектива, де дослідження вже активно виявляли прогалини та недоліки в існуючих підходах. Обидва погляди були об'єднані для створення консолідованого списку факторів, що впливають на реагування на кіберінциденти, а також виявлення відсутніх факторів, які можуть бути враховані в майбутніх моделях. Ці фактори та перспективи будуть детальніше описані нижче.

Традиційний підхід до реагування на кіберінциденти, що використовується з початку широкого поширення комп'ютерних систем, полягав у виявленні інциденту, його локалізації або пом'якшенні, з подальшим відновленням нормальної роботи системи. Наступний пост-інцидентний аналіз використовувався для визначення можливих удосконалень в інфраструктурі та процесах. Це добре ілюструє модель Інституту SANS, що була деталізована у 2003 році, а також більш пізні рекомендації NIST, які розглядають процес реагування на інцидент як замкнутий цикл із застосуванням досвіду для покращення наступних заходів.

Міжнародні стандарти, такі як ISO27001, пропонують аналогічний підхід, який базується на циклі Демінга (Plan-Do-Check-Act). При цьому основна увага традиційно приділяється збереженню конфіденційності, цілісності та доступності (Confidentiality, Integrity, Availability). Однак слід зазначити, що такий традиційний підхід може вступати в конфлікт із завданнями інших зацікавлених сторін, зокрема з точки зору військової чи бізнес-розвідки, де пріоритетним є отримання більш глибоких розвідувальних даних щодо нападників, їхніх цілей, мотивації та використовуваних ресурсів.

Саме тут може бути ефективним використання штучного інтелекту. Традиційні моделі мають тенденцію зупиняти аналіз інциденту одразу після його локалізації. Штучний інтелект, особливо з його здатністю швидко аналізувати великі масиви даних та автоматично виокремлювати значущі деталі, може значно розширити межі такого аналізу. Наприклад, NLP-технології дозволяють автоматично структурувати

великі обсяги текстової інформації, виявляючи зв'язки між атаками та їх виконавцями, допомагаючи зберігати та накопичувати цінну інформацію про минулі інциденти для використання в майбутньому. Також методи машинного навчання здатні не лише ефективно визначати патерни атак, але й прогнозувати майбутні події на основі історичних даних, тим самим посилюючи ситуаційну обізнаність аналітиків та дозволяючи більш ефективно передбачати потенційні загрози.

Крім того, штучний інтелект допомагає вирішити проблему недостатності інформації («інформаційного голоду»), характерної для традиційних підходів. Завдяки автоматизованому збору та систематизації даних з різних джерел ШІ може оперативно забезпечувати аналітиків якісною інформацією, необхідною для стратегічного аналізу інцидентів та формування ефективних довгострокових стратегій реагування.

Штучний інтелект створює перспективні можливості для автоматизації окремих етапів процесу обробки інцидентів інформаційної безпеки. Важливо підкреслити, що, попри широкі можливості автоматизації, повністю довіряти процес прийняття рішень (Decision Making) штучному інтелекту на даний момент неможливо через ризики помилок та хибних висновків. Проте, ШІ ефективно застосовується у допоміжних задачах, що значно прискорюють аналіз інцидентів та мінімізують навантаження на експертів.

Конкретні приклади застосування ШІ включають такі технології:

Обробка природної мови (NLP) – ця технологія використовується для автоматичного створення зрозумілих Incident Summary на основі технічних логів або звітів. Наприклад, NLP може бути використано для автоматичної генерації зрозумілих звітів про фішингові атаки на основі аналізу листів і текстів повідомлень, що дозволяє швидко передати інформацію командам реагування.

Машинне та глибоке навчання – дозволяє виявляти та аналізувати аномалії у поведінці систем. Наприклад, моделі на основі машинного навчання ефективно ідентифікують атаки типу DDoS (Distributed Denial of Service), аналізуючи незвичайний трафік та шаблони мережевої активності. Прикладом успішного

виявлення (True Positive) є своєчасна детекція та блокування масштабної DDoS-атаки, яка могла б призвести до тривалого простою сервісів. Однак, такі системи можуть також помилково ідентифікувати як загрозу пікове навантаження на сайт у день розпродажів (False Positive), що може викликати зайве втручання команди реагування.

Експертні системи – використовують базу знань та встановлені правила для надання рекомендацій щодо подальших дій на основі типових інцидентів. Наприклад, експертна система може автоматично пропонувати дії при виявленні типових ознак компрометації акаунта (наприклад, багаторазові невдалі спроби входу). Це прискорює процес початкового реагування, хоча рішення про блокування акаунта завжди має приймати спеціаліст.

Розвиваючи цю ідею прогнозування далі, ще на початку 2000-х років, дослідники розглядали важливість інтеграції елементів інтелектуального аналізу у стандартні процеси реагування на кіберінциденти, які до того часу розглядалися переважно як «ремонт після атаки, нейтралізація та стримування» (ARNC). Ще до появи процесу «SEI State of the Practice, визначили стандартну модель реагування на інциденти як недостатньо ефективну через її суто реактивну природу.

Натомість було запропоновано розширену модель аналізу та прогнозування ходу кібератак, відому як Attack Repair, Neutralization and Containment (ARNC), доповнену підходом «Cyber Intelligence Preparation of the Battlefield» (C-IPB). Цей підхід включав можливість формування оцінок про ймовірні та можливі дії атакуючої сторони, що дозволяло значно покращити точність і актуальність заходів реагування. Процес C-IPB охоплював чотири основні кроки: визначення операційного середовища (визначення зони відповідальності), аналіз дій та можливостей атакуючої сторони, оцінку вразливостей, а також прогнозування наступних можливих кроків атакуючих. Кіберрозвідка на той час вже поділялась на аналіз виконаних дій зловмисників, їх можливостей, персональних характеристик та намірів. Відтоді, хоча фундаментальні принципи залишаються незмінними, значно розширилися типи інформації, які необхідно враховувати при аналізі кіберзагроз. Це,

зокрема, знайшло відображення в сучасних стандартах, таких як Structured Threat Information eXpression (STIX), і використанні механізмів обміну інформацією про загрози (наприклад, Structured Threat Information eXpression, а також Threat Intelligence Exchange), що підтримуються завдяки сучасним технологіям, зокрема штучному інтелекту. STIX забезпечує уніфіковану архітектуру, яка об'єднує різноманітні типи інформації про кіберзагрози, зокрема:

Спостереження – наприклад, створення ключа реєстру, трафік мережі до певних IP-адрес, отримання електронних листів із конкретної адреси тощо.

Індикатори (Indicators) – потенційні спостереження, що мають визначений контекст та значення.

Інциденти (Incidents) – конкретні випадки дій зловмисника.

Тактики, техніки та процедури зловмисників (Tactics, Techniques, and Procedures – TTP) – шаблони атак, шкідливе програмне забезпечення, експлойти, послідовності дій (kill chains), інструменти, інфраструктура, цілі атак тощо.

Цілі експлуатації (Exploit Targets) – наприклад, уразливості, слабкі місця або конфігурації.

Плани дій (Courses of Action) – реагування на інциденти або усунення вразливостей.

Кампанії кіберзагроз (Cyber Attack Campaigns) – набори пов’язаних інцидентів або TTP, об'єднаних спільною метою.

Суб'єкти кіберзагроз (Cyber Threat Actors) – ідентифікація та/або характеристика зловмисника.

Для того щоб зробити цю агреговану архітектуру максимально практичною для конкретних випадків використання, можуть залучатися існуючі структуровані мови, такі як Cyber Observable Expression (CybOX™), Malware Attribute Enumeration and Characterization (MAEC™), Common Attack Pattern Enumeration and Classification (CAPEC™) та інші. Крім того, мова STIX передбачає гнучкість у своєму використанні: практично всі її елементи є необов'язковими, що дозволяє

використовувати лише ті частини STIX, які є релевантними для конкретного сценарію – від одного окремого поля до всього спектру можливостей мови.

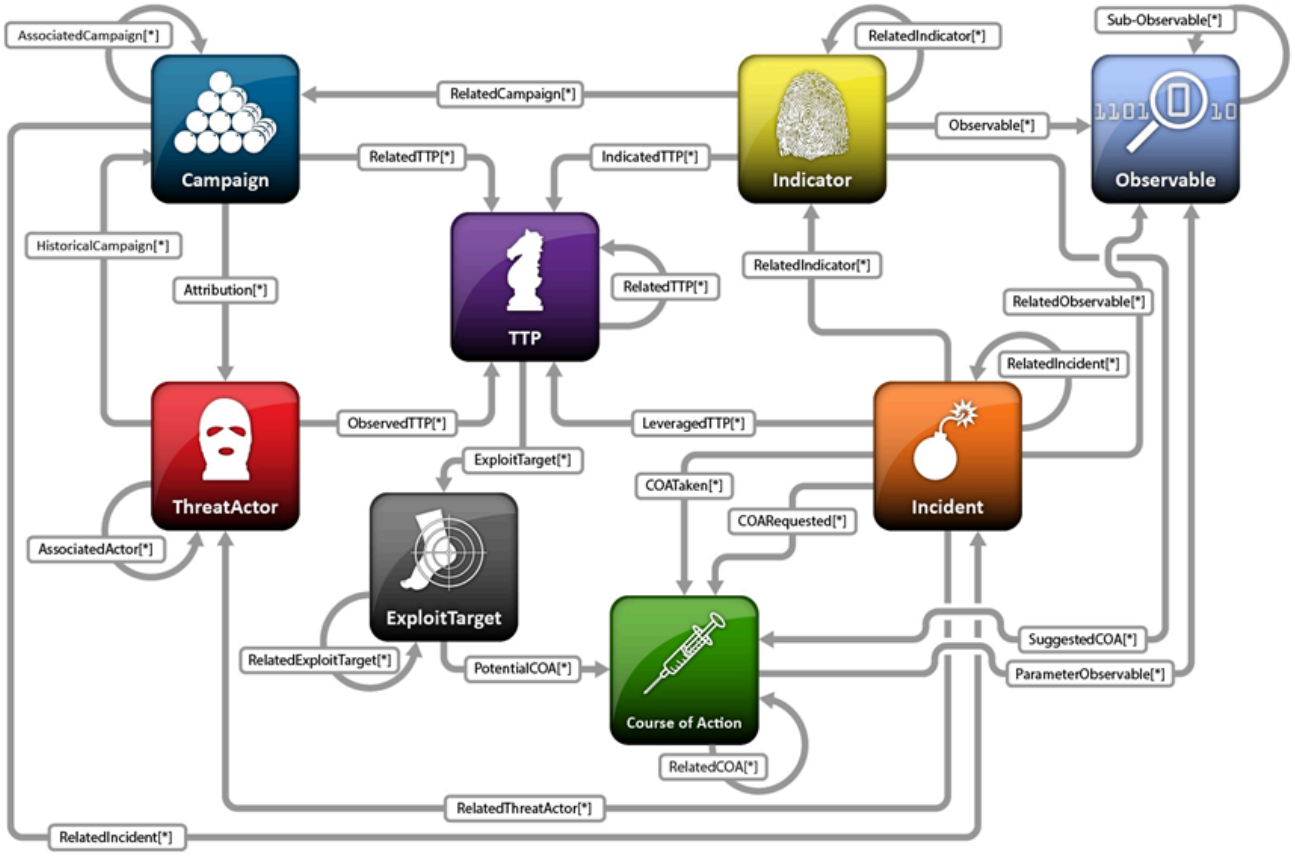


Рисунок 1.2 – Цикли обробки інцидентів згідно MITRE CORPORATION STRUCTURED THREAT INFORMATION EXPRESSION (STIX)

STANAG Threat Information eXpression (STIX) призначений для підтримки обміну інформацією щодо кіберзагроз між різними організаціями. Ця модель дозволяє стандартизовано представляти інформацію про загрози, підтримуючи такі задачі: виявлення кіберінцидентів, реагування на них, обмін інформацією про кіберзагрози, а також аналітичну роботу з прогнозування майбутніх атак.

Окрім питань забезпечення належної роботи сенсорів і наявності підготовленого персоналу для своєчасного виявлення інцидентів, у документі приділяється увага питанням управління ризиками, відповідальності за них, а також довірі до апаратного забезпечення, персоналу та партнерів. Цікаво, що в документі також

коротко порушується дилема вибору між негайною зупинкою виявлених атак та їх подальшим моніторингом для отримання додаткової розвідувальної інформації.

Документ також наголошує на важливості міжнародного обміну інформацією про кіберінциденти, зокрема в рамках ініціатив НАТО, таких як Multi-National Experiment 7 – Access to the Global Commons (MNE7), а також Multinational Capability Development Campaign (MCDC) Cyber Implications for Combined Operational Access (CICOA), що активно реалізовувалась у 2013-2014 роках.

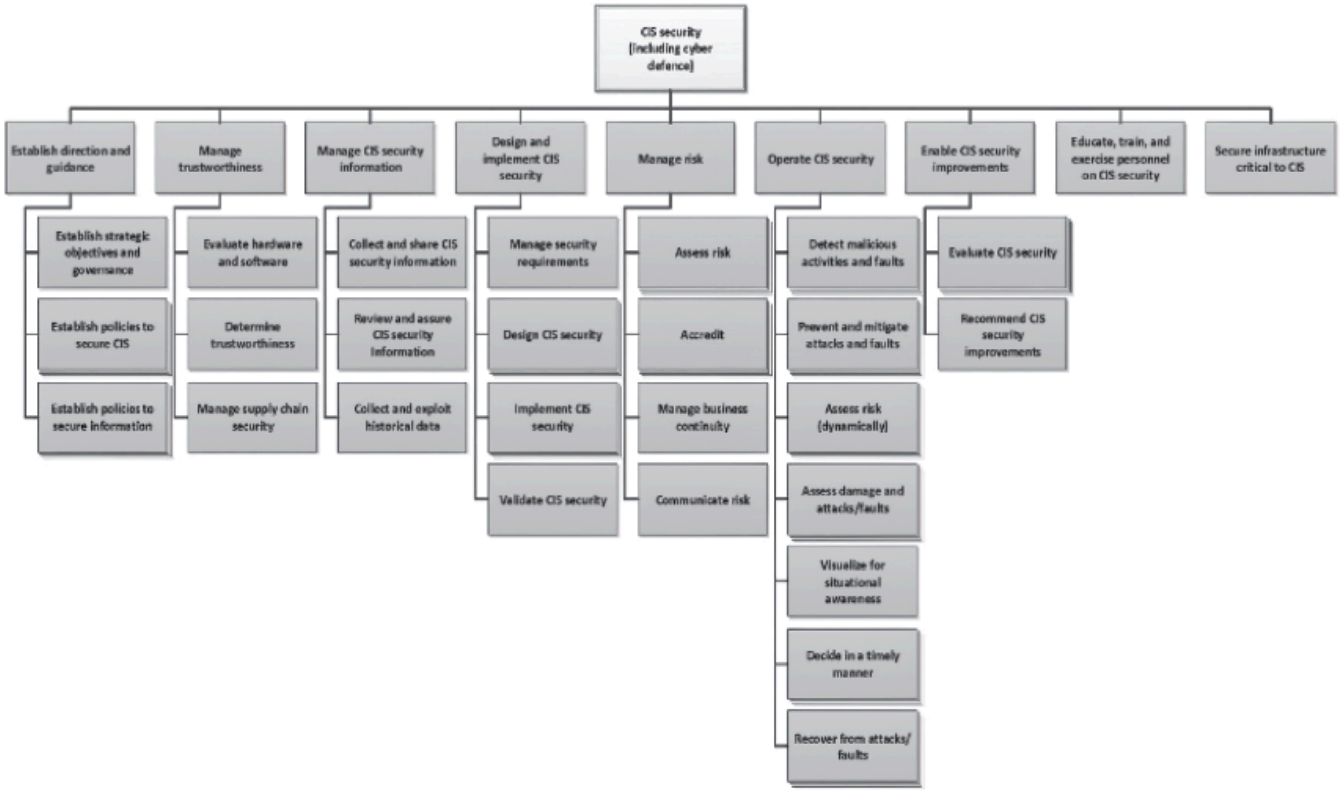


Рисунок 1.3 – Цикли обробки інцидентів згідно NC3A CIS SECURITY FRAMEWORK

Зрештою, будь-яка відповідь на кіберінцидент, яку обирає уповноважений фахівець, повинна бути здійснена достатньо швидко, щоб вплинути на кінцевий результат. Це найкраще описується моделлю, запропонованою полковником ВПС США Джоном Бойдом (Colonel John Boyd USAF), відомою як цикл OODA (Orr, 1983). Ця модель включає такі етапи: спостереження (Observe) — моніторинг дій

противника, орієнтація (Orient) — оцінка можливих дій та їх наслідків на основі спостережень за противником і знань про власні можливості, прийняття рішення (Decide) — вибір відповідного способу реагування, та дія (Act) — реалізація прийнятого рішення.

Цикл OODA був створений для того, щоб описати процес отримання переваги у повітряних боях. Згідно з цією моделлю, той, хто завершує цикл OODA швидше за свого противника, не дозволяє йому адекватно відреагувати, що гарантує перевагу.

Застосування штучного інтелекту може допомогти суттєво скоротити цикл OODA під час кіберінцидентів. Зокрема, машинне навчання та системи штучного інтелекту можуть автоматично здійснювати швидкий збір та аналіз великих обсягів інформації, що дозволяє скоротити етапи спостереження та орієнтації. Крім того, автоматичний пошук релевантної інформації та формування Incident Summary за допомогою технологій обробки природної мови дає фахівцям змогу оперативніше приймати обґрунтовані рішення. Завдяки цьому, організації можуть набагато швидше виконати весь цикл OODA, ефективніше реагуючи на дії зловмисників та підвищуючи свою перевагу в динамічних умовах кіберконфліктів. Зрештою, будь-яка відповідь на кіберінцидент, яку обирає уповноважений фахівець, повинна бути здійснена достатньо швидко, щоб вплинути на кінцевий результат.

Застосування штучного інтелекту може допомогти суттєво скоротити цикл OODA під час кіберінцидентів. Зокрема, машинне навчання та системи штучного інтелекту можуть автоматично здійснювати швидкий збір та аналіз великих обсягів інформації, що дозволяє скоротити етапи спостереження та орієнтації. Крім того, автоматичний пошук релевантної інформації та формування Incident Summary за допомогою технологій обробки природної мови дає фахівцям змогу оперативніше приймати обґрунтовані рішення. Завдяки цьому, організації можуть набагато швидше виконати весь цикл OODA, ефективніше реагуючи на дії зловмисників та підвищуючи свою перевагу в динамічних умовах кіберконфліктів.

1.3 Важливість написання стандартизоваої звітності та після-інцидентих операцій

Досі мало уваги приділялося тому, як організації оцінюють ефективність свого навчання на вже опрацьованих інцидентах. Незрозуміло, як організаціям слід аналізувати власну здатність до навчання та постійно її вдосконалювати. У галузі науки управління було проведено багато досліджень організацій, Однак нам поки що невідомо про жодні дослідження, які б оцінювали, чи можна застосувати ці підходи до того, як команди з кібербезпеки можуть покращити своє навчання на прикладі інцидентів кібербезпеки [5]. Щоб повністю зрозуміти це, в майбутніх дослідженнях необхідно оцінити, чи дійсно заходи, впроваджені на основі уроків інцидентів, зменшують поширеність майбутніх інцидентів.

У відібраних дослідженнях не повідомлялося про спроби виміряти, як навчання на основі інцидентів покращило безпеку, або про те, що досліджувані організації оцінювали власну здатність до навчання. Як уже згадувалося, важко достовірно виміряти інциденти, але більшість проведених досліджень ґрунтується на неявному припущенні, що навчання на основі інцидентів є позитивним, без пошуку явних доказів на підтвердження цієї точки зору. Дослідники виявили, що аналіз після інцидентів, як правило, проводиться лише після великих інцидентів, і не обов'язково тих, які мають найбільші можливості для навчання або час, вкладений у вивчення системних причин [6]. Лише в небагатьох дослідженнях визнаються дилеми, з якими стикаються організації, коли виділяють час і ресурси для інвестування в навчання. Ахмад та ін. [6] припустили, що необхідні подальші дослідження конкуруючих пріоритетів управління безпекою та команд реагування на інциденти, щоб зрозуміти, як їх можна збалансувати, щоб максимізувати безпеку технологій організації. Хоча дослідники закликають організації не лише зосереджуватися на інцидентах з високим рівнем впливу, а й ширше залучати співробітників, необхідні додаткові дослідження, щоб зрозуміти «золоту середину», де інвестиції в навчальні заходи переважають над вигодами. Це можна вивчити,

порівнюючи організації або підрозділи всередині організацій, які інвестували різну кількість часу в навчальну діяльність, та оцінюючи вплив на поширеність інцидентів. Однак для проведення порівняння необхідно, щоб кожна організація зіткнулася з однаковою кількістю інцидентів в якості базової лінії.

Однак використання даних для вимірювання ефективності навчання з точки зору зменшення кількості майбутніх інцидентів є надзвичайно складним через відсутність незалежних або надійних джерел даних про інциденти, доступних для академічних дослідників.

Організаціям необхідно збалансувати зусилля між навчанням на основі інцидентів та реагуванням на них. Однак усунення першопричин, якщо вони були правильно визначені, має зменшити кількість інцидентів. Спираючись на системне мислення, відобразили [7] цю петлю зворотного зв'язку на своїх системних діаграмах команд реагування на інциденти, але жодне з досліджень не вивчало цей компроміс в організаціях, щоб зрозуміти оптимальний баланс. Дослідники [8] могли б застосувати підхід науки про дизайн, щоб перевірити, як вирішити практичні проблеми балансування між реагуванням на інциденти та часом, необхідним для розслідування та вивчення інцидентів. Дослідження показують, що ширша організація рідко залучається до аналізу післяінцидентних ситуацій, а звіти часто не поширюються за межами групи реагування на інциденти та їхнього безпосереднього керівництва. У випадку з інцидентами з безпекою виникло протиріччя: люди хотіли бути залученими до навчальної діяльності, але були стурбовані додатковими зусиллями. Це ще раз підкреслює важливість того, щоб дослідники враховували комерційну реальність інвестування часу і ресурсів у навчання. В організації, яку досліджували Тату та ін. (2018) [9], вони виявили, що обізнаність користувачів покращилася після інциденту з програмою-вимагачем, однак потрібні додаткові дослідження того, як уроки з інцидентів кібербезпеки можуть бути засвоєні більш широкими організаціями. Потрібні додаткові дослідження, щоб з'ясувати цінність різних учасників у розслідуванні причин інцидентів, в тому числі наслідки

залучення людей з інших підрозділів IT, юридичних, кадрових, операційних працівників і сторонніх експертів, а також ідеальне поєднання старшинства.

Дослідники вказують [10] на недостатність залучення постачальників до навчання після інцидентів. Організації все більше залежать від еко-системи постачальників. Необхідно провести додаткові дослідження, щоб прояснити відносини, які організації мають з постачальниками щодо навчання на основі інцидентів, включаючи контрактні умови, участь в аналізі після інцидентів, обмін інформацією та впровадження отриманих уроків. Подальші тематичні дослідження та спостереження за аналізом післяінцидентних ситуацій допоможуть краще зрозуміти виклики, з якими стикаються організації, коли вчаться на інцидентах.

У дослідженнях, присвячених процесу управління інцидентами, часто застосовуються стандарти ISO/IEC 27035 або NIST SP-800-61. Це добре, оскільки ці стандарти передбачають врахування отриманих уроків як важливий крок, що розширює коло досліджень, які включають навчання в свої дослідження. Хоча стандарти передбачають навчання, вони не пояснюють, як ефективно витягувати уроки і, що особливо важливо, використовувати їх для зменшення ймовірності та впливу інцидентів у майбутньому (Ahmad et al., 2020) [11]. Здається, що стандарти припускають, що організації здатні ідентифікувати цінні уроки без вказівок щодо оптимального підходу до їх вилучення.

Для вивчення того, як визначаються уроки, дослідники запозичили моделі з психології, науки управління або науки про безпеку. Деякі беруть уроки з наявних звітів про інциденти, щоб представити висновки в новому форматі Для покращення безпеки необхідна подальша робота, щоб зрозуміти, що заважає кращому навчанню. У багатьох дослідженнях визнається потенційна цінність навчання на основі інцидентів для вирішення основних проблем, які можуть покращити безпеку організації в цілому. Однак, оскільки процес навчання найбільше висвітлювався на етапі причинно-наслідкового аналізу, дослідники [11] стверджують, що «потенційна роль реагування на інциденти як інструменту навчання і зворотного зв'язку для

досягнення ширших організаційних цілей, зокрема, управління безпекою, була мало визнана».

Дослідники знайшли мало доказів того, що організаційне навчання виходить за рамки вдосконалення процесів команди з управління інцидентами. Галузеві стандарти управління інцидентами навмисно надають організаціям гнучкість у визначенні того, як їх впроваджувати, і вони зосереджені на управлінні інцидентами, а не на тому, як отримані уроки можуть покращити загальний стан безпеки в організації.

Висновок до першого розділу

У першому розділі дипломної роботи було здійснено ґрунтовне дослідження ключових стандартів та підходів до обробки інцидентів інформаційної безпеки. Проаналізовано як традиційні, так і сучасні методи реагування на кіберінциденти, зокрема моделі, запропоновані NIST, ISO/IEC 27035, ENISA, а також підходи, що базуються на використанні штучного інтелекту.

Встановлено, що в Україні відсутній єдиний уніфікований стандарт реагування на інциденти, що ускладнює впровадження системного підходу до управління кіберзагрозами. У зв’язку з цим українські організації орієнтуються на міжнародні стандарти, серед яких найбільш релевантними є ISO/IEC 27035 та NIST SP 800-61. Ці документи визначають життєвий цикл обробки інцидентів, що включає етапи підготовки, виявлення та аналізу, стримування, ліквідації, відновлення та постінцидентної діяльності.

Особливу увагу приділено ролі інноваційних технологій, зокрема штучного інтелекту, у підвищенні ефективності реагування на інциденти. Встановлено, що ШІ здатен значно покращити процеси виявлення, аналізу та постінцидентного навчання, знижуючи навантаження на аналітиків та підвищуючи точність оцінки загроз. Водночас, повна автоматизація критичних етапів реагування (стримування, ліквідація) наразі є недоцільною через високі ризики.

Також розглянуто альтернативні підходи до реагування, зокрема моделі C-IPB, STIX та цикл OODA, які акцентують увагу на розвідувальному компоненті та швидкості прийняття рішень. Визначено, що ефективне управління інцидентами потребує не лише технічних засобів, а й організаційної зрілості, стандартизованої звітності та здатності до навчання на основі попередніх інцидентів.

Таким чином, дослідження підтвердило необхідність інтеграції міжнародних стандартів, адаптації сучасних технологій та формування культури безперервного вдосконалення процесів кіберзахисту в українських організаціях.

РОЗДІЛ 2

МОЖЛИВОСТІ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ПРИ ОБРОБЦІ ІНЦИДЕНТІВ

2.1 Штучний інтелект як допоміжний інструмент у обробці інцидентів

Зростаюча кількість та складність кібератак підкреслили нагальну потребу в інноваційних рішеннях для посилення безпеки цифрової інфраструктури. Серед таких рішень штучний інтелект (ШІ) постає як перспективна технологія, що має потенціал значно підвищити рівень кібербезпеки та ефективність реагування на інциденти. ШІ здатний покращити швидкість і точність виявлення загроз, реагування та їхнього усунення, а також зменшити навантаження на фахівців із безпеки [10].

Розглянемо роль ШІ в ключових аспектах кібербезпеки та реагування на інциденти, зокрема оцінці вразливостей, виявленні та запобіганні вторгненням, а також цифровій криміналістиці. ШІ, завдяки своїм вбудованим можливостям, може стати вирішальним фактором, що дає організаціям змогу ефективніше виявляти, реагувати та нейтралізувати загрози.

ШІ не є панацеєю для кібербезпеки. Як і будь-яка технологія, він має свої обмеження та потенційні вразливості [11]. Тому в даній роботі також розглядається необхідність постійного розвитку у сфері ШІ, щоб подолати ці обмеження та виклики. Усвідомлюючи потребу у безперервному вдосконаленні, наукова робота підкреслює важливість подальших досліджень і розробок, спрямованих на максимізацію потенційних переваг ШІ у сфері кібербезпеки.

Початковий етап реагування

На цьому етапі продовжується збір даних про інцидент, що розпочався на попередньому етапі. Мета – зібрати достатньо інформації для формування ефективної стратегії реагування. Як правило, на цьому етапі отримуються дані з інтерв’ю з особами, які повідомили про підозрілий інцидент, а також доступні

журнали мережевого моніторингу або звіти IDS, які можуть підтвердити факт інциденту.

Формулювання стратегії реагування передбачає врахування всіх обставин інциденту, таких як критичність уражених систем або даних, тип атакувальника та можливі наслідки. Політика реагування організації, яка визначає її підхід до кіберінцидентів, також може суттєво вплинути на вибір стратегії реагування [12].

Розслідування інциденту

Під час цього етапу збираються різні типи доказів, що мають відношення до інциденту, наприклад, дані з хостів або мережеві логи, щоб реконструювати події, що призвели до кіберінциденту. Реконструкція повинна [13] дати відповіді на питання: що сталося, коли, як і чому, а також хто за це відповідає.

Розслідування зазвичай поділяється на два етапи:

Збір даних (Data Collection)

Аналіз даних (Data Analysis)

Завершальний етап

Мета цього етапу – вжити необхідних заходів для стримування інциденту, усунути першопричини проблеми та запобігти повторенню подібних інцидентів у майбутньому. Всі важливі кроки повинні бути задокументовані, а їх виконання проконтрольоване для перевірки ефективності.

Будь-які зміни у вражених системах слід проводити лише після збору всіх можливих доказів, інакше вони можуть бути втрачені. Після вирішення інциденту може виникнути необхідність оновлення політик безпеки або процедур реагування на інциденти [14], якщо під час аналізу було виявлено слабкі місця в поточних практиках.

Штучний інтелект у сфері кібербезпеки є корисним, оскільки покращує аналіз, дослідження та розуміння кіберзлочинності фахівцями з безпеки. Він підсилює технології кіберзахисту, які використовують компанії для боротьби з кіберзлочинцями, допомагаючи захищати як організації, так і клієнтів [15].

З іншого боку, штучний інтелект може вимагати значних ресурсів і не завжди є доцільним у всіх застосуваннях. Більш того, він може стати новою зброєю у руках кіберзлочинців, які використовують цю технологію для вдосконалення своїх атак.

Штучний інтелект у кібербезпеці є перспективною галуззю, що привертає все більшу увагу та інвестиції. Сучасні компанії зазвичай мають багаторівневий захист, який включає заходи безпеки на рівні периметра, мережі, кінцевих пристроїв, програмного забезпечення та даних [16]. Наприклад, це можуть бути апаратні чи програмні міжмережеві екрани (файрволи) та системи захисту мережі, що аналізують і контролюють дозволені підключення. Якщо хакерам вдасться подолати ці рубежі, вони зіткнуться з антивірусними програмами, а далі – із системами виявлення та запобігання вторгненням (IDS/IPS) тощо.

Існує небагато наукових джерел, які описують спроби застосування методів штучного інтелекту для обробки інцидентів [17]. Однак, на основі досвіду впровадження ШІ в тактичну та, зокрема, операційну кіберрозвідку, було зроблено висновок, що основною функцією штучного інтелекту в обробці інцидентів є вирішення задачі класифікації [18]. Тобто, однозначне віднесення поточного інциденту до одного з елементів класифікаційної схеми, для кожного з яких розроблено відповідні методики та робочі процеси.

Протягом тривалого часу процес реагування на інциденти виконувався виключно людьми. Проте автоматизація кібератак значно прискорила темп їх виконання, що ускладнило роботу аналітиків із кібербезпеки. Через величезну кількість інцидентів, які виникають унаслідок автоматизованих атак, фахівці з безпеки стикаються з проблемою перевантаження сигналами тривоги (alert fatigue) [19].

Штучний інтелект виступає як засіб для вирішення цієї проблеми, і вже широко використовується у сфері кібербезпеки – як у наукових дослідженнях, так і у комерційних продуктах. Однак, ШІ також може використовуватися як інструмент для проведення кібератак, що робить необхідним його застосування для захисту, щоб ефективно протистояти швидкості та масштабам таких атак.

У сучасному взаємопов’язаному та цифровому світі ландшафт кібербезпеки стає все більш складним і витонченим. Організації стикаються з численними загрозами з боку кіберзлочинців, які використовують вразливості їхніх систем і мереж для несанкціонованого доступу до конфіденційної інформації, порушення операційної діяльності або завдання фінансових збитків [20].

Щоб протистояти цим ризикам і мінімізувати їхній вплив, організації застосовують комплексні заходи безпеки, серед яких оцінка вразливостей є ключовим елементом стратегії кібербезпеки та реагування на інциденти [21]. Безперечно, оцінка вразливостей відіграє першочергову роль в управлінні кібербезпекою. Вона передбачає ретельну ідентифікацію вразливих місць у програмному забезпеченні та системах, що є проактивним процесом сканування та аналізу потенційних загроз з метою запобігання зловмисним атакам.

Сфера оцінки вразливостей досягла значного рівня зрілості. Однак підтримання контролю над великою кількістю цифрових пристроїв і обчислювальних систем, які потребують аналізу, залишається значним викликом [22]. Ця практика базується на методичному підході до виявлення та оцінки вразливостей в IT-інфраструктурі, додатках та системах організації. Вона включає проактивне сканування, тестування та аналіз потенційних слабких місць, які можуть бути використані зловмисниками.

Традиційні підходи до оцінки вразливостей здебільшого спиралися на ручні методи та статичні системи, що базуються на правилах. Однак вони часто не встигають за швидко змінюваним ландшафтом загроз і стрімким зростанням кількості та складності вразливостей.

Поява штучного інтелекту (ШІ) спричинила революційні зміни у сфері кібербезпеки, включаючи оцінку вразливостей та реагування на інциденти. ШІ пропонує нові можливості та ефективніші підходи, які значно підвищують результативність і швидкість виконання цих важливих процесів.

Використовуючи алгоритми машинного навчання, обробку природної мови (NLP) та глибоке навчання, автоматизовані системи на основі ШІ дозволяють

організаціям виявляти, аналізувати та усувати вразливості набагато швидше, точніше і проактивно [23].

За даними Cybersecurity Ventures, щорічно у світі створюється приголомшливі 111 мільярдів рядків нового програмного коду [24]. Використання автоматизованих механізмів для виявлення вразливостей ще до розгортання систем дозволяє командам розробників приділяти більше уваги розширенню функціональності продуктів і покращенню їхньої продуктивності.

Стрімке зростання кількості пристроїв та додатків, що розгортаються сьогодні, не лише підвищує ризики, пов'язані з мережевими системами, а й забезпечує великий обсяг навчальних даних, які можуть бути використані в поєднанні з методами штучного інтелекту [25].

Роль ШІ в оцінці вразливостей є багатогранною, що робить його одним із ключових інструментів сучасної кібербезпеки. Штучний інтелект (ШІ) має здатність автоматизувати та оптимізувати весь процес оцінки вразливостей, зменшуючи потребу в ручній роботі та дозволяючи командам безпеки зосередитися на більш критичних завданнях.

a) Автоматизація та адаптація

Використовуючи алгоритми машинного навчання, ШІ може аналізувати великі масиви даних, включаючи журнали систем, мережевий трафік і історичну інформацію про вразливості. Такий аналіз допомагає виявляти закономірності та аномалії, які можуть свідчити про потенційні загрози. Крім того, ШІ здатний постійно навчатися та адаптуватися до нових атак і тактик зловмисників, що підвищує стійкість процесу оцінки вразливостей [26].

b) Покращене виявлення та аналіз вразливостей

ШІ виступає каталізатором для більш просунутого та точного аналізу вразливостей. Використовуючи методи глибокого навчання, такі як згорткові нейронні мережі (CNN), рекурентні нейронні мережі (RNN) і генеративно-змагальні мережі (GAN), ШІ може отримувати цінну інформацію з великих і складних наборів даних, включаючи неструктуровані джерела, такі як звіти про безпеку, блоги та

наукові публікації. Це дає можливість організаціям виявляти раніше невідомі вразливості та ефективно ідентифікувати нові загрози.

c) Прискорене реагування на інциденти

Оцінка вразливостей на основі ШІ значно сприяє реагуванню на інциденти, оскільки прискорює ідентифікацію вразливостей з високою точністю. Це дозволяє командам безпеки ефективніше розподіляти ресурси та визначати пріоритети. Скорочення часу між виявленням вразливостей і їх усуненням допомагає мінімізувати ризик атак і зменшити наслідки кіберінцидентів [27].

Життєвий цикл управління вразливостями

Нижче наведено діаграму, що ілюструє життєвий цикл управління вразливостями, а також оптимальні етапи оцінки вразливостей у системі [27]:

1. **Виявлення та ідентифікація забутих пристроїв і активів**

Провести глибокий аналіз мережі для виявлення пристроїв або активів, які могли бути проігноровані або забуті.

2. **Пріоритезація активів**

Оцінити важливість та цінність кожного активу для компанії та розставити пріоритети відповідно до їхнього рівня критичності. Впровадження таких підходів на основі ШІ дозволяє організаціям не тільки ефективніше управляти вразливостями, а й значно підвищувати рівень безпеки своєї IT-інфраструктури.

3. **Комплексне сканування**

Навіть після встановлення пріоритетів не залишати жодного компонента без уваги. Провести ретельне сканування всіх складових системи, щоб виявити приховані вразливості.

4. **Ефективна звітність**

Впровадити ефективний механізм звітності, що дозволить оперативно передавати виявлені невизначеності або проблеми вищому керівництву або відповідальним особам.

5. **Оцінка вразливостей та призначення завдань**

Проаналізувати виявлені вразливості та розподіліть завдання відповідно до рівня прийнятного ризику та терміновості усунення.

6. **Перевірка рішень та усунення вразливостей**

Переконайтися в ефективності застосованих заходів та впевніться, що вони успішно нейтралізують виявлені вразливості.

2.2 Використання штучного інтелекту у оцінці вразливостей

Уразливості в сфері безпеки охоплюють різні недоліки та слабкі місця в інформаційних технологіях і пов’язаних із ними продуктах, що поширюються на різні рівні та компоненти інформаційних систем. Такі дефекти безпосередньо впливають на стабільне функціонування всієї інформаційної інфраструктури. Якщо зловмисники скористаються цими вразливостями, це може призвести до серйозних порушень цілісності, конфіденційності та доступності системи. Саме тому дослідження уразливостей є одним із ключових напрямків у сфері інформаційної безпеки [28].

Дотримуватися ітеративного підходу, регулярно повторюючи цикл оцінки та вдосконалюючи процес управління вразливостями для підвищення ефективності.

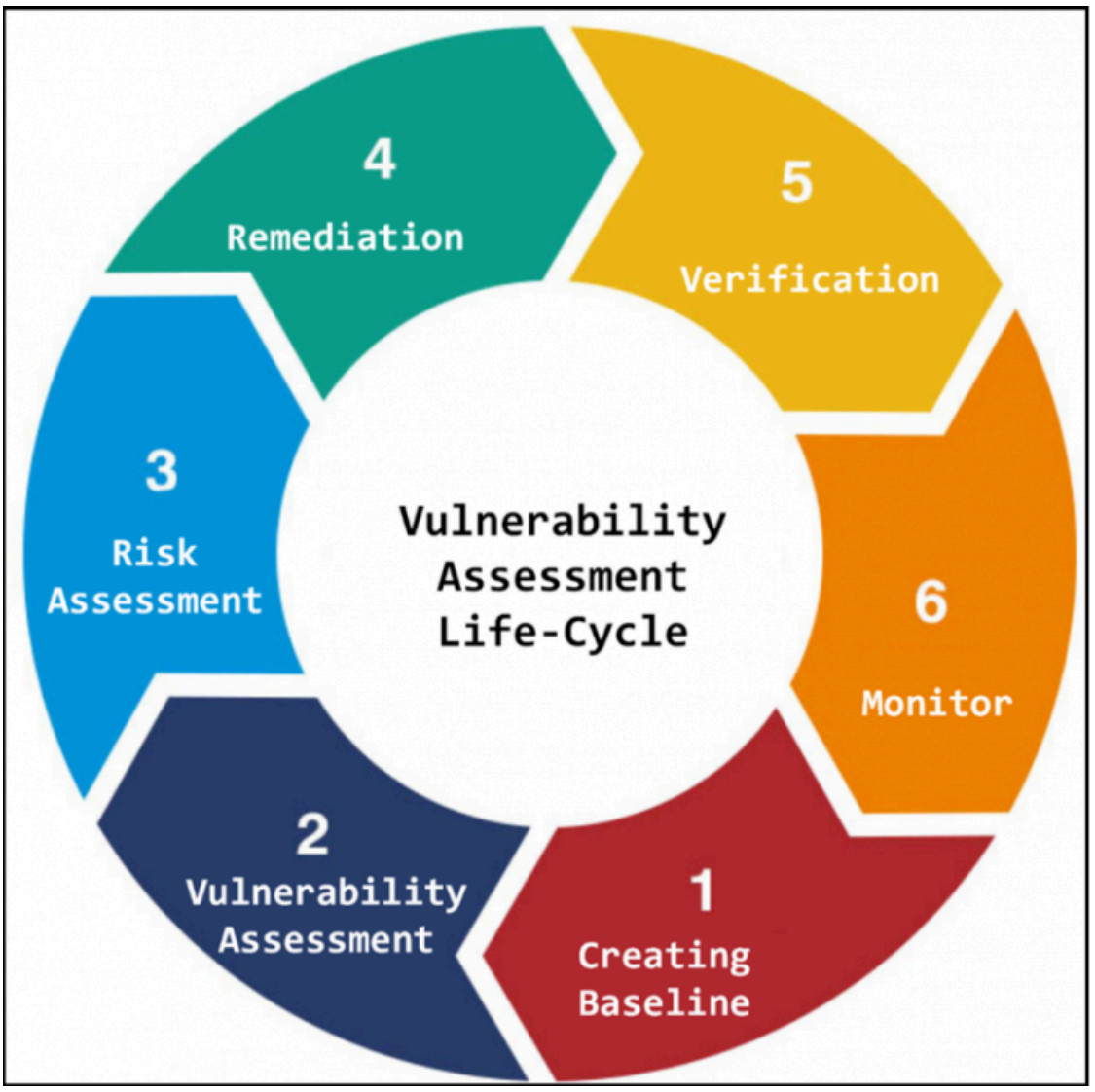


Рисунок 2.1 – Цикл оцінки вразливості

Зі зростанням складності кіберзагроз традиційні методи безпеки вже не можуть забезпечити належний захист від нових, постійно змінюваних ризиків. У зв'язку з цим компанії активно впроваджують штучний інтелект (ШІ) у свої стратегії кібербезпеки [29].

ШІ надає розширені можливості для виявлення та реагування на загрози, підсилює процес управління вразливостями, а також сприяє дотриманню нормативних вимог та покращенню політик управління безпекою. Використовуючи технології ШІ, такі як **машинне навчання, обробка природної мови (NLP), поведінковий аналіз та**

глибоке навчання, організації можуть значно підвищити рівень кіберзахисту [30] та мінімізувати ризики, пов’язані з різними кіберзагрозами, зокрема:

- **Шкідливе програмне забезпечення**
- **Фішингові атаки**
- **Внутрішні загрози**

ШІ вже має безліч застосувань у кібербезпеці, і його роль продовжує зростати, допомагаючи компаніям швидко адаптуватися до нових загроз і проактивно захищати свої інформаційні активи [31].

Виявлення та реагування на загрози за допомогою ШІ

Штучний інтелект відіграє ключову роль у кібербезпеці, забезпечуючи ефективне виявлення загроз і реагування на них. Використовуючи методи машинного навчання та обробки природної мови, організації можуть аналізувати великі обсяги даних для виявлення закономірностей та аномалій, що можуть свідчити про кіберзагрози. Системи виявлення вторгнень на основі ШІ (IDS) здійснюють моніторинг мережевого трафіку для виявлення аномалій або підозрілої активності, а також використовують алгоритми для ідентифікації нетипових патернів, які можуть свідчити про злом.

Проактивне виявлення загроз (Threat Hunting) базується на використанні алгоритмів ШІ для ідентифікації загроз у реальному часі, зокрема складних і довготривалих атак (APT – Advanced Persistent Threats). Прогнозний аналіз (Predictive Analytics) дозволяє оцінювати історичні дані та тренди для передбачення потенційних атак і вразливостей ще до їхнього використання зловмисниками.

Управління вразливостями та пріоритезація загроз

Штучний інтелект значно покращує управління вразливостями, допомагаючи компаніям ефективно виявляти, оцінювати та усувати проблеми безпеки. Автоматизоване сканування вразливостей дозволяє здійснювати пошук слабких місць у системах, мережах та додатках, а також порівнювати знайдені вразливості із базами даних відомих загроз. Для ефективного усунення загроз використовується механізм пріоритезації виправлення уразливостей, що допомагає визначити рівень

ризику кожної вразливості та першочергові проблеми для усунення з урахуванням потенційного впливу на систему.

Автоматизоване тестування на проникнення (Penetration Testing) симулює атаки для перевірки ефективності заходів безпеки. ШІ використовується для аналізу реакції системи на спроби зламу, що дозволяє виявляти потенційні слабкі місця ще до того, як вони будуть експлуатовані зловмисниками.

Забезпечення відповідності та управління безпекою

ШІ допомагає організаціям забезпечувати відповідність нормативним вимогам, таким як GDPR та HIPAA, та вдосконалювати заходи безпеки. Автоматизований моніторинг політик безпеки виявляє порушення політик та автоматично реагує на них, здійснюючи аналіз ризиків і формування рекомендацій для покращення політик безпеки. Аналіз загроз на основі великих даних дозволяє обробляти величезні обсяги інформації, включаючи логи системи та звіти про загрози, що дає змогу швидко виявляти можливі атаки та адаптувати заходи безпеки.

Методи автоматизованої ідентифікації вразливостей

Статичний аналіз коду дозволяє оцінювати вихідний код без його виконання для пошуку небезпечних шаблонів або потенційних вразливостей у коді. Динамічний аналіз коду включає запуск програм у тестовому середовищі для виявлення потенційних проблем у процесі виконання. Мережевий аналіз трафіку використовує алгоритми машинного навчання для моніторингу трафіку та виявлення підозрілих дій, таких як сканування портів або аномальні запити.

2.1.5 Сканування вразливостей за допомогою ШІ

Автоматизовані сканери вразливостей використовують ШІ [28] для перевірки систем, мереж і додатків на наявність відомих проблем безпеки. Вони зіставляють знайдені вразливості з базами даних, порівнюючи їх із відомими експлойтами та методами атак. Також застосовується аналіз поведінки, що дозволяє виявляти аномальні дії користувачів або процесів, які можуть свідчити про потенційну загрозу. Використання обробки природної мови (NLP) [32] у кібербезпеці дозволяє

аналізувати звіти про вразливості, технічні блоги та наукові публікації, що допомагає швидко виявляти нові загрози.

Глибоке навчання у кібербезпеці

ШІ використовує глибокі нейронні мережі для виявлення та аналізу вразливостей. Згорткові нейронні мережі (CNN) застосовуються для аналізу зображень інтерфейсів програм або мережевих схем, тоді як рекурентні нейронні мережі (RNN) аналізують послідовності подій та логи для виявлення аномалій. Генеративно-змагальні мережі (GANs) використовуються для створення нових загроз та перевірки стійкості системи до потенційних атак.

Виявлення та запобігання вторгненням (IDS/IPS)

Мережеве виявлення вторгнень (NIDP) передбачає аналіз мережевого трафіку на рівні пакетів для виявлення шкідливих дій. Використовуються методи поведінкового аналізу для визначення аномалій, що можуть свідчити про потенційну атаку. Хостове виявлення вторгнень (HIDP) здійснює моніторинг дій на рівні окремого пристрою або сервера, аналізуючи журнали подій, системні виклики та зміни у файлах. Інтегровані системи виявлення та запобігання вторгненням (IDPS) поєднують різні методи детекції, включаючи сигнатурний, поведінковий та аномалійний аналіз, що забезпечує більш комплексний підхід до виявлення та запобігання загрозам.

Системи штучного інтелекту ефективно об'єднують дані з різних джерел, таких як бази даних вразливостей, потоки безпеки та системні журнали, створюючи цілісну картину потенційних загроз. Кореляція інформації з цих різнорідних джерел підвищує точність і надійність ідентифікації вразливостей, що дозволяє зміцнити заходи кібербезпеки.

Виявлення та запобігання вторгненням

Виявлення та запобігання вторгненням передбачає постійний моніторинг системних журналів і мережевого трафіку з метою ідентифікації потенційних загроз безпеці. Автоматизовані інструменти безпеки відіграють ключову роль у цьому

процесі, оскільки вони збирають і аналізують великі обсяги даних у реальному часі. Для виявлення підозрілих активностей та аномальних патернів ці системи використовують методи на основі сигнатурного аналізу, виявлення аномалій та поведінкового аналізу.

Попри високу автоматизацію процесу виявлення загроз, роль аналітиків залишається незамінною. Людський фактор є вирішальним у перевірці результатів аналізу, оцінці серйозності загроз та розробці оптимальної стратегії реагування. Аналітики безпеки аналізують отримані дані, перевіряють їхню достовірність та ухвалюють обґрунтовані рішення щодо усунення виявлених загроз. Хоча автоматизація значно покращує процес виявлення загроз, вона не може повністю замінити експертне розуміння контексту, що забезпечує аналітик. Його участь гарантує, що реакція на загрозу буде відповідати конкретним обставинам, політикам організації та не спричинить зайвих перешкод у роботі легітимного мережевого трафіку. Крім того, аналітики відіграють важливу роль у постійному вдосконаленні систем виявлення та запобігання вторгненням, адаптуючи їх до нових типів атак і змінюючи конфігурацію відповідно до сучасних викликів кібербезпеки [33].

Мережеве виявлення та запобігання вторгненням (NIDP)

Системи мережевого виявлення та запобігання вторгненням (NIDP) виконують моніторинг мережевого трафіку для ідентифікації та реагування на потенційні вторгнення. Вони застосовують різні методи аналізу даних на рівні мережевих пакетів для виявлення аномальної або шкідливої активності. Одним з основних підходів у NIDP є аналіз пакетів, що передбачає детальне дослідження заголовків та вмісту мережевого трафіку для виявлення загроз або нетипових моделей поведінки. Використання методів аналізу мережевого трафіку та поведінкових характеристик дозволяє ефективно виявляти та блокувати підозрілу активність, забезпечуючи додатковий рівень захисту мережевої інфраструктури [34].

Аналіз заголовків і вмісту пакетів мережевого трафіку дозволяє виявляти шаблони або аномалії, які можуть вказувати на потенційні вторгнення. Серед

основних методів аналізу пакетів використовуються глибока перевірка пакетів (DPI) та аналіз протоколів.

Виявлення аномалій є ще одним важливим аспектом мережевого виявлення вторгнень (NIDP), що передбачає встановлення базової поведінки для порівняння з поточною мережею з метою ідентифікації відхилень. Для виявлення аномалій широко застосовуються статистичні методи, алгоритми машинного навчання та поведінковий аналіз. Порівнюючи поточні мережеві патерни з історичними даними або попередньо визначеними пороговими значеннями, системи NIDP можуть генерувати сповіщення або застосовувати превентивні заходи.

Сигнатурний аналіз є перевіреним методом у NIDP, який передбачає порівняння мережевого трафіку з базою відомих сигнатур атак. Якщо система знаходить збіг, вона видає попередження. Однак такий метод ефективний лише для відомих атак і може бути недостатнім для виявлення нових або невідомих атак. Щоб подолати це обмеження, системи виявлення та запобігання вторгнень часто поєднують сигнатурний аналіз із методами виявлення аномалій для забезпечення більш надійного рівня безпеки.

Моніторинг мережевого трафіку є невід’ємною частиною NIDP, оскільки включає збір та аналіз мережевих потоків, зокрема інформацію про IP-адреси джерела та призначення, порти, протоколи та тривалість сеансів. Завдяки аналізу мережевих потоків адміністратори безпеки можуть ідентифікувати підозрілі шаблони, такі як аномальні обсяги переданих даних або незвичні комунікаційні схеми. Дані про мережеві потоки також можна використовувати для візуалізації мережевої активності та виявлення прихованих закономірностей, які неможливо розпізнати іншими методами аналізу.

Система виявлення та запобігання вторгненням на рівні хоста (HIDP) зосереджується на моніторингу активності та подій на окремих пристроях чи кінцевих точках з метою захисту від внутрішніх атак або загроз у мережі. Такі методи забезпечують детальну видимість процесів на рівні хоста та відіграють важливу роль у захисті систем. Аналіз логів є ключовим компонентом HIDP,

оскільки системні журнали містять важливу інформацію про дії користувачів, зокрема спроби входу в систему, доступ до файлів, системні виклики та зміни конфігурації. Аналіз логів допомагає аналітикам безпеки виявляти підозрілі або несанкціоновані дії. Автоматизовані інструменти аналізу логів сприяють швидкому виявленню шаблонів загроз або небезпечних подій, що підвищує ефективність виявлення вторгнень.

Моніторинг системних викликів є ще однією важливою технікою HIDP, яка передбачає перехоплення та аналіз викликів операційної системи, здійснюваних програмами або процесами. Це дозволяє виявляти шкідливу або аномальну поведінку, таку як спроби несанкціонованого доступу, підвищення привілеїв або маніпуляції з файлами. Виявлені аномалії можуть активувати попереджувальні сигнали або превентивні заходи для нейтралізації потенційних ризиків.

Перевірка цілісності файлів є ще одним механізмом, який використовується для підтримки безпеки, зокрема через збереження хеш-значень або контрольних сум для кожного файлу та періодичну перевірку їхньої цілісності шляхом повторного обчислення хешу і порівняння з початковим значенням. Будь-які невідповідності можуть свідчити про можливі зміни у файлах або їхню компрометацію, що може вказувати на порушення безпеки.

Методи поведінкового аналізу в HIDP зосереджуються на постійному моніторингу та аналізі поведінки процесів і додатків, що працюють на хостах. Такий підхід дозволяє виявляти відхилення від очікуваних шаблонів поведінки, що може свідчити про аномальну або потенційно шкідливу активність.

Системи виявлення та запобігання вторгненням (IDPS) відіграють ключову роль у виявленні та реагуванні на вторгнення в комп’ютерні мережі та системи. Вони призначені для безперервного моніторингу мережевого трафіку, активності хостів і системних журналів, забезпечуючи можливість виявлення та запобігання загрозам у режимі реального часу. IDPS можуть працювати в різних режимах, включаючи мережевий, хостовий або їх комбінацію, що дозволяє забезпечити комплексне покриття кібербезпеки.

Ці системи базуються на поєднанні технологій, методологій та алгоритмів, які дозволяють виявляти та нейтралізувати загрози. Для ідентифікації зловмисних дій та потенційних вразливостей вони використовують сучасні методи, такі як сигнатурний аналіз, виявлення аномалій і поведінковий аналіз.

Сигнатурний аналіз передбачає порівняння мережевого трафіку, даних хостів або системних журналів із відомими сигнатурами атак. Ці сигнатури базуються на раніше зафіксованих та задокументованих випадках зловмисної активності. У разі збігу система генерує сповіщення, що дозволяє фахівцям із безпеки вжити відповідних заходів. Цей метод ефективний для виявлення вже відомих атак, але може мати труднощі з виявленням нових або невідомих загроз, які не мають попередньо визначених сигнатур.

Виявлення аномалій є ще одним важливим компонентом IDPS. Ця методика базується на створенні еталонної моделі нормальної поведінки мережі або хостів і порівнянні поточної активності з цією моделлю. Виявлені відхилення або аномалії можуть свідчити про потенційне вторгнення. Алгоритми виявлення аномалій використовують статистичні методи, машинне навчання та поведінковий аналіз для виявлення нетипових шаблонів, сплесків мережевого трафіку або нестандартної поведінки системи. Завдяки генеруванню сповіщень про аномальні дії IDPS здатні виявляти раніше невідомі або змінювані загрози.

Поведінковий аналіз є проактивним підходом до виявлення зловмисної активності на основі аналізу поведінки мережевого трафіку, додатків або системних процесів. Вивчаючи послідовність дій, моделі доступу до ресурсів або комунікаційну поведінку, система може виявляти відхилення від очікуваної поведінки та генерувати відповідні сповіщення. Цей метод є особливо ефективним для виявлення складних атак, які можуть залишатися непоміченими при сигнатурному аналізі [36].

Окрім виявлення загроз, IDPS також акцентує увагу на превентивних заходах та реагуванні на інциденти. При виявленні потенційного вторгнення або підозрілої активності система може здійснювати різні дії для запобігання подальшому

пошкодженню або мінімізації впливу загрози. Серед таких заходів може бути блокування мережевого трафіку, ізоляція скомпрометованих хостів, скидання сеансів користувачів або сповіщення персоналу безпеки для подальшого розслідування. Крім того, IDPS можуть інтегруватися з іншими засобами безпеки, такими як міжмережеві екрани, для автоматичного застосування політик контролю доступу або оновлення правил безпеки з метою підвищення загального рівня захисту.

Автоматизоване реагування в кібербезпеці означає використання інструментів і алгоритмів штучного інтелекту для автоматичного виконання певних дій у відповідь на виявлені загрози або інциденти безпеки. Такі автоматизовані заходи допомагають запобігти поширенню кібератак і зменшити їхній вплив.

Розглянемо приклад для кращого розуміння роботи автоматизованого реагування. Уявімо велику організацію, яка використовує розвинену систему виявлення вторгнень на основі ШІ. Ця система безперервно моніторить мережу на предмет підозрілої активності або потенційних загроз. Одного дня вона виявляє серію мережевих пакетів із характерними ознаками DDoS-атаки. Помітивши потенційну загрозу, система безпеки на основі ШІ негайно активує відповідні дії. Вона аналізує вхідний мережевий трафік, ідентифікує шкідливі пакети та визначає оптимальний спосіб нейтралізації атаки. У цьому випадку система вирішує заблокувати IP-адреси, з яких надходять атакуючі пакети.

Завдяки своїм автоматизованим можливостям, ШІ відправляє команди на інфраструктуру мережі, зокрема на міжмережеві екрани або маршрутизатори. Внаслідок цього заблоковані IP-адреси негайно припиняють доступ до ресурсів організації, що ефективно зупиняє потік шкідливого трафіку. Одночасно система ініціює заходи для ізоляції заражених пристроїв у мережі. Вона визначає скомпрометовані комп’ютери або сервери, які можуть бути частиною атаки, і відокремлює їх від решти мережі. Завдяки цьому вдається запобігти подальшому поширенню атаки та зменшити потенційні збитки.

У цьому випадку можливості автоматизованого реагування, які надають інструменти кібербезпеки на основі ШІ, відіграють критично важливу роль у

стримуванні та мінімізації наслідків DDoS-атаки. Автоматичне блокування підозрілого трафіку та ізоляція заражених систем допомагають уникнути перебоїв у роботі мережевих сервісів та значного простою. Крім того, автоматизація таких рутинних завдань зменшує навантаження на аналітиків безпеки. Замість того, щоб вручну ідентифікувати та блокувати шкідливий трафік, вони можуть зосередитися на складніших стратегічних завданнях, таких як аналіз першопричини атаки, виявлення потенційних вразливостей або вдосконалення алгоритмів реагування ШІ.

ШІ також можна використовувати для прогнозного аналізу, що передбачає застосування історичних даних для ідентифікації можливих загроз у майбутньому. Аналізуючи закономірності та тенденції в мережевій активності, алгоритми штучного інтелекту можуть виявляти потенційні вразливості та передбачати загрози ще до їхньої реалізації. Це дозволяє організаціям проактивно усувати ризики, перш ніж вони зможуть завдати шкоди.

Однак важливо пам’ятати, що ШІ не є універсальним рішенням для всіх проблем кібербезпеки і повинен використовуватися разом з іншими інструментами та методами. Наприклад, алгоритми ШІ можуть не виявляти складні загрози, такі як APT-атаки (Advanced Persistent Threats) або уразливості нульового дня, які потребують людської експертизи та інтуїції. Крім того, алгоритми ШІ можуть бути схильні до помилкових спрацьовувань або пропускати реальні загрози, що може призвести до зайвих попереджень або недооцінки небезпеки.

2.3 Аналіз рішень обробки інцидентів на основі штучного інтелекту від вендорів у сфері кібербезпеки

Для розуміння реального стану систем які використовують штучний інтелект як основу виявлення кіберзагроз було проведено дослідження декількох доступних рішень у реальному корпоративному середовищі.

2.3.1 Аналіз ефективності систем штучного інтелекту для виявлення кіберзагроз: дослідження Checkpoint XDR/XPR

Одним досліджених рішень є Checkpoint XDR/XPR, що позиціонується як передова система виявлення та реагування, яка використовує ШІ для аналізу кіберзагроз.

Згідно із заявами виробника, система має можливості хмарних операцій безпеки на основі штучного інтелекту, що дозволяє обробляти великі обсяги даних з різних джерел. Вона забезпечує комплексну профілактику, охоплюючи кінцеві точки, мережу, мобільні пристрої, електронну пошту та хмарні сервіси. Крім того, реалізована автоматизована співпраця у сфері безпеки, що дозволяє швидко реагувати на загрози за рахунок обміну даними між різними модулями системи. XDR/XPR миттєво блокує кіберзагрози та запобігає їх поширенню між елементами корпоративного середовища без необхідності додаткового налаштування. Вбудовані механізми кореляції журналів дозволяють об’єднувати події з різних джерел у єдині інциденти безпеки, що спрощує аналіз. Важливими аспектами є автоматичне виявлення загроз та реагування на них без втручання адміністратора, а також механізм усунення хибних спрацювань завдяки використанню алгоритмів машинного навчання.

У рамках дослідження система була розгорнута в корпоративному середовищі та працювала протягом 52 днів. За цей період система зафіксувала 1382 події безпеки, серед яких 9 виявилися дійсними загрозами (True Positive), а 1373 – хибними спрацюваннями (False Positive). Аналіз хибних спрацювань показав, що система часто маркувала як загрозу очікуваний мережевий трафік, доступ до корпоративної мережі з зареєстрованих та контрольованих пристроїв, а також типове використання офісних пристроїв, наприклад, Pass-the-Ticket атаку через DHCP, яка насправді не була загрозою.

Незважаючи на велику кількість хибних спрацювань, система успішно ідентифікувала 9 реальних атак. Зокрема, були зафіксовані спроби зовнішніх

зловмисників використати експлойт після оновлення вразливої системи, а також виявлення аномальної активності, яка могла свідчити про компрометацію облікових записів.

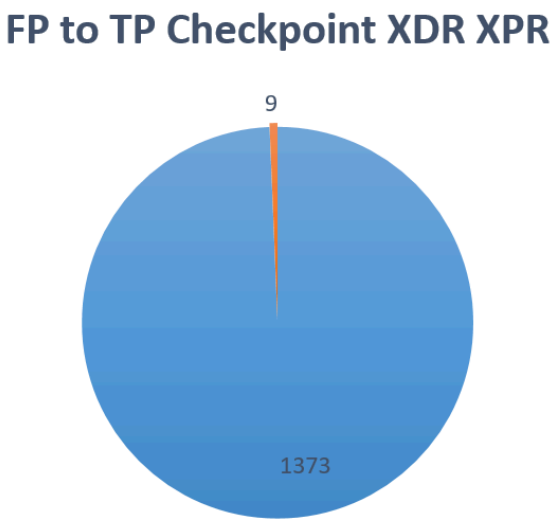


Рисунок 2.2 – Діаграма співвідношення хибних спрацювань до дійсних у системі Checkpoint

Результати дослідження показали, що Checkpoint XDR/XPR має проблеми із завеликою кількістю хибних спрацювань, що може створювати додаткове навантаження на аналітиків SOC і знижувати ефективність обробки інцидентів. Попри заявлену автоматизацію, система потребує додаткового налаштування, щоб мінімізувати помилкові спрацювання. Виявлено, що механізми самонавчання та адаптивності ШІ потребують покращення для більш точної фільтрації подій.

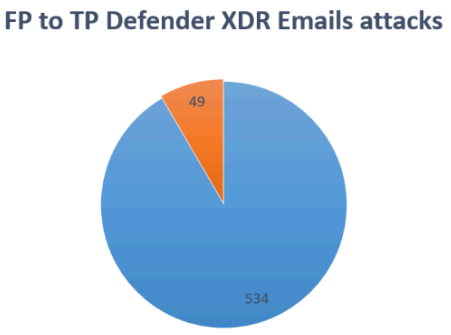
Водночас система впоралася із виявленням дійсних атак, що підтверджує її ефективність за умови правильного налаштування та доповнення іншими джерелами загроз, такими як Threat Intelligence. Для покращення продуктивності рекомендується оптимізувати політики безпеки, інтегрувати додаткові механізми навчання ШІ та зменшити кількість хибних спрацювань шляхом ретельного коригування алгоритмів виявлення.

Загалом Checkpoint XDR/XPR має значний потенціал, але потребує вдосконалення механізмів фільтрації загроз, щоб зменшити навантаження на аналітиків та покращити ефективність реагування на реальні кіберзагрози.

2.3.2 Аналіз ефективності систем штучного інтелекту для виявлення кіберзагроз: дослідження Checkpoint XDR/XPR

Microsoft Defender XDR for Exchange є комплексним рішенням для захисту електронної пошти, що використовує штучний інтелект для виявлення загроз, таких як фішингові атаки, компрометація облікових записів і поширення шкідливого програмного забезпечення. Це рішення інтегрується в екосистему Microsoft 365 і застосовує аналітику поведінки, машинне навчання та обробку природної мови (NLP) для аналізу підозрілих листів. Основна увага приділяється виявленню фішингових атак, спрямованих на компрометацію облікових записів користувачів.

У рамках дослідження система була розгорнута у корпоративному середовищі та перебувала під моніторингом протягом 52 днів. За цей період Defender XDR for Exchange зафіксував 49 дійсних загроз і 534 хибних спрацювання, що свідчить про загалом високу ефективність у виявленні реальних атак. Окремо було проаналізовано роботу лише ШІ-основаного фільтра, який використовує методи обробки природної мови для класифікації електронних листів. Його результати виявилися незадовільними: за весь період нагляду він не зафіксував жодної дійсної загрози, але створив 23 хибних спрацювання.



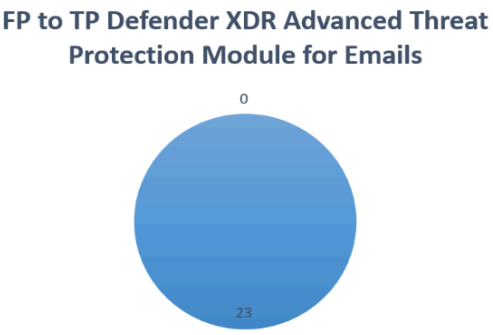


Рисунок 2.3 – Діаграма співвідношення хибних спрацювань до дійсних у системі Checkpoint

Головна проблема ШІ-основаного фільтра полягала в тому, що він аналізував лише текст листів і не враховував контекстні фактори, такі як історія комунікацій, технічні мета-дані та репутація відправника. Усі зафіксовані повідомлення, які система вважала потенційно небезпечними, виявилися очікуваними легітимними листами. Це свідчить про те, що алгоритми NLP орієнтовані на загальні патерни тексту, але не можуть ефективно розрізняти реальні загрози від стандартного ділового листування.

Натомість комбіновані методи виявлення, що враховували поведінкові фактори та технічні параметри, дозволили системі правильно зафіксувати 49 дійсних атак. Це підтверджує важливість багаторівневого аналізу загроз, оскільки покладання виключно на NLP-фільтри значно знижує ефективність виявлення.

2.3.3 Аналіз ефективності NLP-моделей у контексті виявлення кіберзагроз

Загалом Microsoft Defender XDR for Exchange демонструє високу ефективність у виявленні фішингових атак за умови використання комплексного підходу до аналізу загроз. Водночас, ШІ-алгоритми обробки природної мови, що працюють без додаткових перевірок, виявилися неефективними, що вказує на необхідність їх подальшого вдосконалення.

Обробка природної мови (NLP) відіграє ключову роль у сучасних системах аналізу загроз, дозволяючи автоматично інтерпретувати текстову інформацію та ідентифікувати потенційні ризики. Серед найбільш відомих NLP-моделей сьогодні можна виділити OpenAI GPT, Google BERT, Meta LLaMA та Microsoft Turing NLG. Вони широко застосовуються у сфері кібербезпеки для аналізу журналів подій, класифікації загроз та автоматизації обробки інцидентів. Попри високу точність у

загальних завданнях розпізнавання мови, такі моделі все ще мають обмеження при роботі з технічними контекстами та кіберзагрозами.

При тестуванні NLP-моделей із використанням реальних логів із корпоративного середовища було виявлено ряд недоліків у їхній здатності правильно інтерпретувати загрози. Зокрема, OpenAI GPT хоч і розпізнав факт атаки у внутрішній мережі компанії, однак не зміг ідентифікувати, що її джерелом був внутрішній сканер вразливостей. Як наслідок, система не змогла надати жодних реальних методів обробки інциденту, фактично знімаючи з себе відповідальність за точність аналізу.

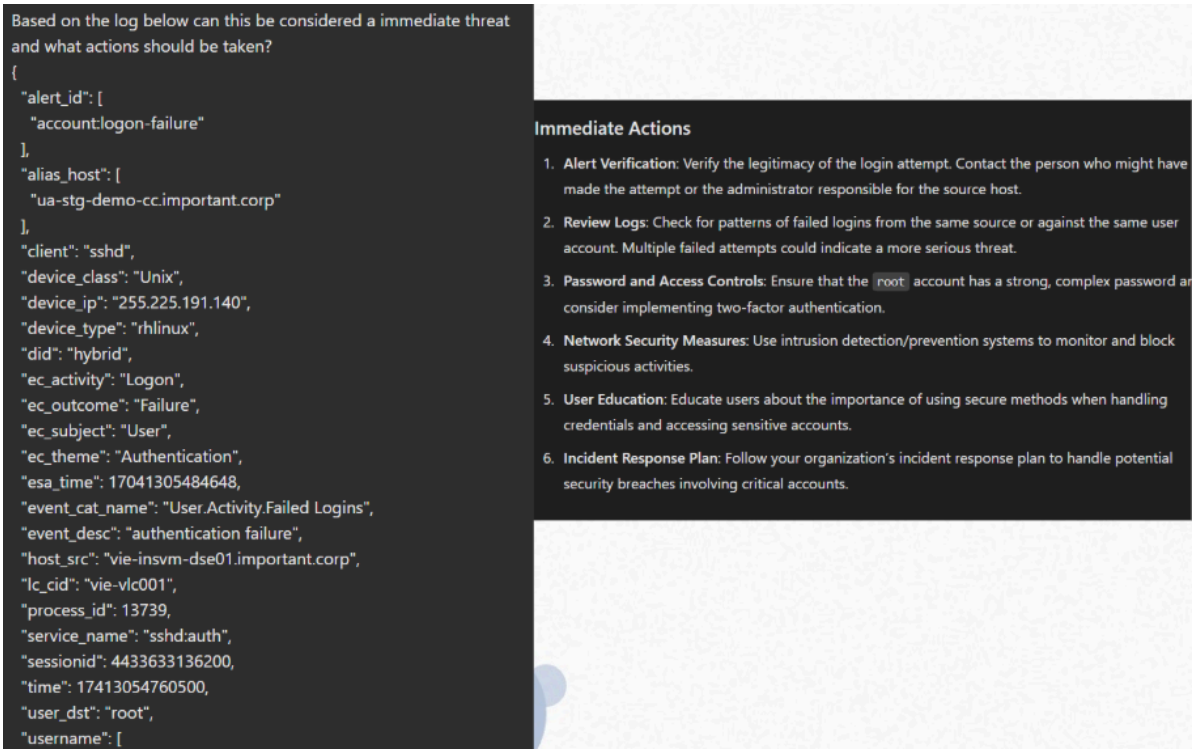


Рисунок 2.4 – Наданий лог файл разом відповіддю NLP моделі

Додаткове тестування інших NLP-моделей показало, що вони схильні до хибних спрацювань у випадках, коли їм надаються запити, які містять логічну пастку або семантичну неоднозначність. Наприклад, модель могла ідентифікувати як загрозу запис із запитом "Is the 'XSS attack' an XSS attack?", що не містить реальної загрози, але за своєю структурою вводить модель в оману.

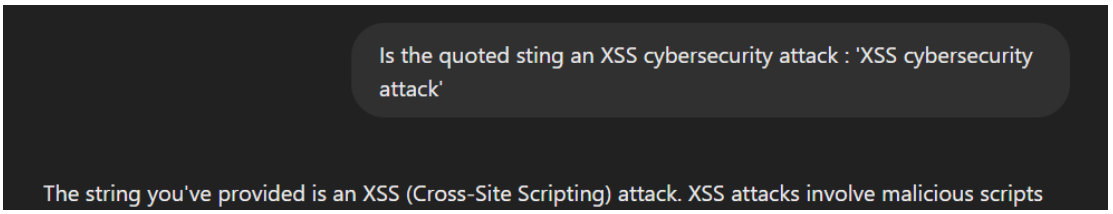


Рисунок 2.5 – Оманливий запит до NLP моделі та її відповідь

2.3.3 Аналіз рішення Microsoft Copilot for Security

Цікавою відмінністю серед рішень на базі штучного інтелекту є Microsoft Copilot for Security, який спеціалізується на екосистемі продуктів Microsoft і тісно інтегрується з такими компонентами, як Microsoft Azure, Sentinel, Defender XDR та інші. На відміну від моделей, що зосереджуються на пошуку аномалій та генерації відповідних сповіщень безпеки, Copilot for Security спроектовано спеціально для підтримки фахівців з кібербезпеки, з акцентом на прискорення аналітичного процесу, розвантаження інженерів безпеки та автоматизацію рутинних технічних задач. Його основна функція — бути ефективним помічником саме для аналітиків у центрах безпеки, які працюють у хмарній або гібридній інфраструктурі Microsoft. Саме такий напрямок, орієнтований не на повну заміну, а на посилення професійної діяльності, виглядає найбільш перспективним станом на сьогодні.

Цю думку підтверджує також спеціальна публікація Microsoft під назвою Generative AI and Security Operations Center Productivity: Evidence from Live Operations[35], в якій наведено практичні дані про ефективність використання Copilot for Security у робочому середовищі. Згідно з результатами, впровадження цього інструмента суттєво покращило продуктивність у різних сценаріях, включаючи інцидент-менеджмент, технічну підтримку та програмну розробку. Наведені дані у публікації демонструють наступне:

Технічна підтримка клієнтів: зменшення часу виконання завдань для новачків на 34%

Лабораторний експеримент з розслідування інцидентів кібербезпеки: зниження часу виконання завдань на 23%

Лабораторний експеримент з розгортання HTTP-сервера на JavaScript: скорочення часу на виконання на 55,8%

Польове дослідження розробки програмного забезпечення: збільшення кількості виконаних завдань на 26,08%

Лабораторне дослідження IT-адміністраторів: покращення точності виконання завдань на 34,53% та зменшення часу виконання на 30,69%

У сукупності ці результати вказують на реальне зростання ефективності роботи фахівців за рахунок використання Copilot for Security. Microsoft також заявляє, що їхнє рішення дозволяє досягти 22,6% відносного часу на вирішення інциденту в порівнянні з аналогічними завданнями без використання ШІ. Приклад розглянутого інциденту наведено на рисунку 2.6

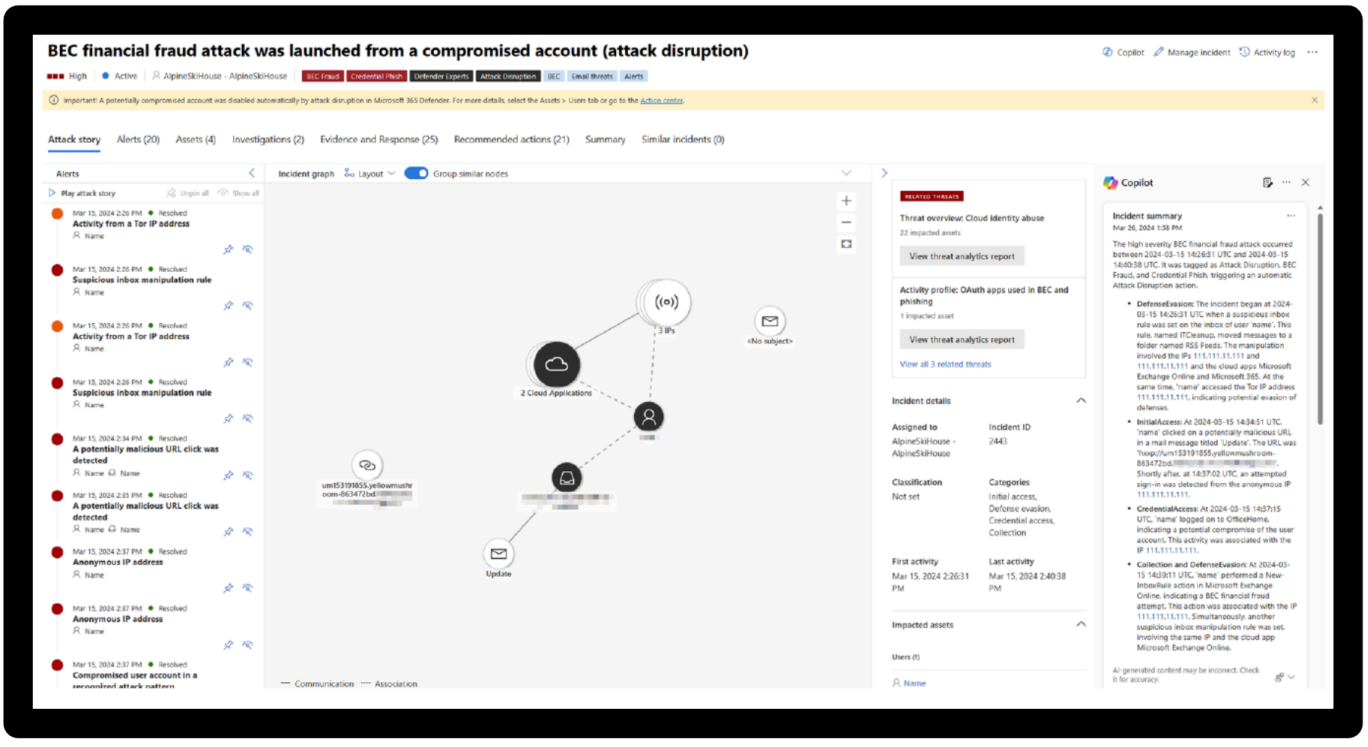
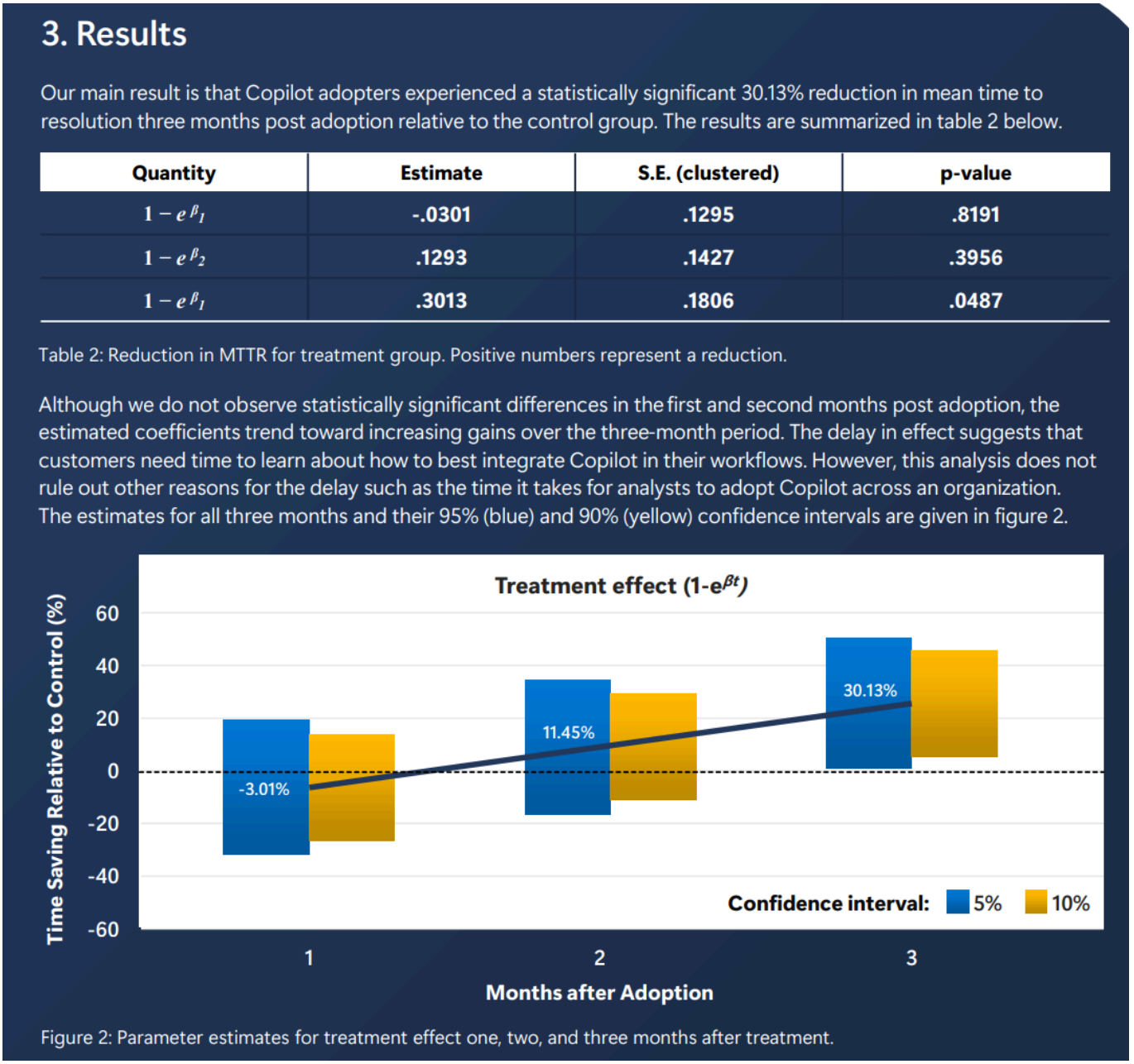


Рисунок 2.6 – Приклад підсумовування інциденту компрометації бізнес-електронної пошти у Microsoft Copilot

Ефективність прийняття рішення Copilot for Security Microsoft, у порівнянні з контрольною групою, відповідно до публікації наведено на рисунку 2.7



3. Results

Our main result is that Copilot adopters experienced a statistically significant 30.13% reduction in mean time to resolution three months post adoption relative to the control group. The results are summarized in table 2 below.

Quantity	Estimate	S.E. (clustered)	p-value
$1 - e^{\beta_1}$	-.0301	.1295	.8191
$1 - e^{\beta_2}$	.1293	.1427	.3956
$1 - e^{\beta_1}$	.3013	.1806	.0487

Table 2: Reduction in MTTR for treatment group. Positive numbers represent a reduction.

Although we do not observe statistically significant differences in the first and second months post adoption, the estimated coefficients trend toward increasing gains over the three-month period. The delay in effect suggests that customers need time to learn about how to best integrate Copilot in their workflows. However, this analysis does not rule out other reasons for the delay such as the time it takes for analysts to adopt Copilot across an organization. The estimates for all three months and their 95% (blue) and 90% (yellow) confidence intervals are given in figure 2.

Figure 2: Parameter estimates for treatment effect one, two, and three months after treatment.

Рисунок 2.7 – Ефективність прийняття рішення Copilot for Security Microsoft, у порівнянні з контрольною групою

Це означає, що завдяки Copilot for Security фахівці можуть не лише швидше реагувати на інциденти, а й краще розподіляти ресурси в умовах високого навантаження [35].

Загалом, сучасні NLP-моделі мають значний потенціал для використання у сфері кібербезпеки, але їхні можливості обмежені у випадках, коли необхідний

глибокий контекстний аналіз або визначення джерела загрози. Високий рівень хибних спрацювань та складнощі з обробкою специфічних технічних запитів залишають відкритим питання щодо їхньої практичної ефективності у виявленні та реагуванні на реальні атаки.

2.4 Аналіз можливостей LMM моделей у пошуку відповідностей технік мапінгу для спрощення підбиття підстумків

Для перевірки ефективності штучного інтелекту у пошуку відповідностей технік згідно Mapping to ATT&CK from narrative reporting [36] використано офіційний матеріал тренування Mitre. Розглянемо ефективність відповідностей, яка наведена у таблиці 2.1. Для LMM моделі ChatGPT4-o mini встановлено запит, який вибрано з офіційного матеріалу Mitre.

Таблиця 2.1

Порівняння мапінгу за допомогою штучного інтелекту до технік MITRE Attack з офіційним відповідями до тренувального мапінгу

LLM Model	MITRE Answers	Full match	Not present in LLM Models	Not present in Mitre Answers
T1566.001 – Phishing: Spearphishing Attachment	T1566.001 – Phishing: Spearphishing Attachment	T1566.001 – Phishing: Spearphishing Attachment	Execution - User Execution: Malicious File [T1204.002]	T1583.001 – Acquire Infrastructure: Domains
"spear phishing emails with malicious attachments..."	"spear phishing emails with malicious attachments..."	"spear phishing emails with malicious attachments..."	Execution - User Execution: Malicious Link [T1204.001]	"registers and leverages domains that masquerade as legitimate web services..."
T1566.002 – Phishing: Spearphishing Link	T1566.002 – Phishing: Spearphishing Link	T1566.002 – Phishing: Spearphishing Link	"...typically resulting in a POWBAT infection."	T1190 – Exploit Public Facing Application
"...and/or hyperlink..."	"...and/or hyperlink..."	"...and/or hyperlink..."		"...exploited vulnerable web services..."
T1583.001 – Acquire Infrastructure: Domains	Execution - User Execution: Malicious File [T1204.002]	T1505.003 – Server Software Component: Web Shell		T1140 – Deobfuscate/Decode Files or Information
"registers and leverages domains that masquerade as legitimate web services..."	Execution - User Execution: Malicious Link [T1204.001]	"...to install web shell, such as ANTAKA and ASPXSPY"		Possibly inferred from use of tool like WinRAR/7-Zip with password protection.
T1190 – Exploit Public Facing Application	"spear phishing emails with malicious attachments..."	T1078 – Valid Accounts		T1036 – Masquerading
"...exploited vulnerable web servers..."	"...and/or hyperlink..."	"...used stolen legitimate credential..."		"domains that masquerade as legitimate web services..."
T1059.006 – Command and Scripting Interpreter: PowerShell	T1566.001 – Phishing: Spearphishing Attachment	T1003.001 – OS Credential Dumping: LSASS Memory		T1087 – Account Discovery [likely via Mimikatz and other tools].
Implicit in "POWBAT infection" [which his PowerShell-based malware].	"spear phishing emails with malicious attachments..."	Tool like Mimikatz, Windows Credential Editor, and Pinx Dump.		T1021.002 – Remote Services: SMB/Windows Admin Shares [PsExec]
T1505.003 – Server Software Component: Web Shell	T1566.002 – Phishing: Spearphishing Link	T1046 – Network Service Scanning		"...PsExec, RemCom, and xCmdSvc..."
"...to install web shell, such as ANTAKA and ASPXSPY"	"...and/or hyperlink..."	"...port scanner, BLUETORCH"		T1572 – Protocol Tunneling
T1078 – Valid Accounts	T1505.003 – Server Software Component: Web Shell	T1059.006 – Command and Scripting Interpreter: PowerShell		"...SOCKS5 proxies..."
"...used stolen legitimate credentials..."	"...to install web shell, such as ANTAKA and ASPXSPY"	Implicit in "POWBAT infection" [which his PowerShell-based malware].		T1078.003 – Valid Accounts: Local Accounts or T1078.004 – Valid Accounts: Domain Accounts
T1078.003 – Valid Accounts: Local Accounts or T1078.004 – Valid Accounts: Domain Accounts	T1078 – Valid Accounts	T1021.004 – Remote Services: SSH		Continuing use of stolen credentials.
Continuing use of stolen credentials.	"...used stolen legitimate credentials..."	"...used SSH..."		T1076 – Remote Desktop Protocol
T1078 – Remote Desktop Protocol	T1003.001 – OS Credential Dumping: LSASS Memory	T1021.001 – Remote Services: Remote Desktop Protocol [RDP]		"APT29 has used this protocol to maintain persistence..."
"APT29 has used this protocol to maintain persistence..."	Tools like Mimikatz, Windows Credential Editor, and Pinx Dump.	"...used RDP..."		T1053.005 – Scheduled Task/Job: Scheduled Task [assumed based on known POWBAT behavior].
T1053.005 – Scheduled Task/Job: Scheduled Task [assumed based on known POWBAT behavior].	T1046 – Network Service Scanning	T1090.003 – Proxy: Multi-hop Proxy		T1003 – OS Credential Dumping [general technique].
T1003.001 – OS Credential Dumping: LSASS Memory	"...port scanner, BLUETORCH"	"...SOCKS5 proxies between infected hosts..."		T1110 – Brute Force: Credential Stuffing
Tools like Mimikatz, Windows Credential Editor, and Pinx Dump.	T1059.006 – Command and Scripting Interpreter: PowerShell	T1560.001 – Archive Collected Data: Archive via Utility		Use of tools like Hydra suggests brute-force attempts.
T1003 – OS Credential Dumping [general technique].	Implicit in "POWBAT infection" [which his PowerShell-based malware].	"...archives stolen data with compression tools such as WinRAR or 7-Zip"		T1003.001 – LSASS Memory
T1110.004 – Brute Force: Credential Stuffing	T1021.004 – Remote Services: SSH			Via Mimikatz, Windows Credential Editor.
Use of tools like Hydra suggests brute-force attempts.	"...used SSH..."			T1036 – Input Capture [possible but not explicit in the text].
T1140 – Deobfuscate/Decode Files or Information	T1021.001 – Remote Services: Remote Desktop Protocol [RDP]			T1036 – PowerShell
Possibly inferred from use of tools like WinRAR/7-Zip with password protection.	"...used RDP..."			Via POWBAT, customscripts].
T1036 – Masquerading	T1090.003 – Proxy: Multi-hop Proxy			T1071.001 – Application Layer Protocol: Web Protocols [likely via POWBAT or web shell]
"domains that masquerade as legitimate web services..."	"...SOCKS5 proxies between infected hosts..."			T1048.001 – Exfiltration Over Web Service [likely, though not explicitly mentioned]
T1003.001 – LSASS Memory	T1560.001 – Archive Collected Data: Archive via Utility			T1567.002 – Exfiltration Over Web Service: Exfiltration to Cloud Storage
[via Mimikatz, Windows Credential Editor].	"...archives stolen data with compression tools such as WinRAR or 7-Zip"			[possible implication if masquerading as legit web services]
T1056 – Input Capture [possible but not explicit in the text].				
T1059 – PowerShell				
[via POWBAT, customscripts].				
T1046 – Network Service Scanning				
"...port scanner, BLUETORCH"				
T1083 – File and Directory Discovery [likely via scripts and tools, implied].				
T1087 – Account Discovery [likely via Mimikatz and other tools].				
T1021.001 – Remote Services: Remote Desktop Protocol [RDP]				
"...used RDP..."				
T1021.004 – Remote Services: SSH				
"...used SSH..."				
T1021.002 – Remote Services: SMB/Windows Admin Shares [PsExec]				
"...PsExec, RemCom, and xCmdSvc..."				
T1572 – Protocol Tunneling				
"...SOCKS5 proxies..."				
T1560.001 – Archive Collected Data: Archive via Utility				
"...archives stolen data with compression tools such as WinRAR or 7-Zip"				
T1090.003 – Proxy: Multi-hop Proxy				
"...SOCKS5 proxies between infected hosts..."				
T1071.001 – Application Layer Protocol: Web Protocols [likely via POWBAT or web shell]				
T1048.001 – Exfiltration Over Web Service [likely, though not explicitly mentioned]				
T1567.002 – Exfiltration Over Web Service: Exfiltration to Cloud Storage				
[possible implication if masquerading as legit web services]				

Із загальних 13 технік, зазначених у вибірці, що базується на офіційному звіті Mapping to ATT&CK from Narrative Reporting, модель штучного інтелекту коректно ідентифікувала 11 технік із точним обґрунтуванням прив’язки до відповідних фрагментів тексту. Це становить майже 90% відповідності з офіційними результатами.

Окрім того, модель запропонувала додатково 18 технік, які відсутні в офіційному звіті. З них 6 технік були релевантно прив’язані до тексту, а ще 12 запропоновані на основі контексту. Важливо зазначити, що ці додаткові техніки не є хибними — вони мають змістовне підґрунтя і можуть слугувати цінним доповненням при виконанні мапінгу.

Загалом, результати роботи штучного інтелекту в межах цієї перевірки можна оцінити як відмінні.

Під час ручного аналізу, здійсненого аналітиками першого рівня цільового підприємства, середній рівень відповідності технік склав близько 60%. Це зумовлено як складністю і різноманіттям наявних технік, так і можливістю дублювання деяких з них у випадках, коли використовується як загальна, так і специфічна форма однієї і тієї ж дії.

2.5 Аналіз можливостей застосування штучного інтелекту у вирішенні інцидентів

Відповідно до проведеного дослідження було підготовлено тезу до виступу на конференції за темою "Reducing Alert Fatigue in SOC Operations: Improving Incident Response Using Automation and Artificial Intelligence" [19]. Основна увага доповіді була зосереджена на проблемі перевантаження аналітиків безпеки великою кількістю сповіщень про потенційні загрози, що часто призводить до втоми від сповіщень (alert fatigue) та зниження ефективності обробки інцидентів. У межах дослідження було розглянуто сучасні методи автоматизації процесів аналізу загроз та використання штучного інтелекту для прискорення реагування на інциденти.

Запропонований підхід включає розробку практичного рішення для зменшення часу розслідування та вирішення інциденту. Для цього було створено систему, що забезпечує автоматизований пошук інформації про інцидент, включаючи виявлені вразливості та їх використання. Крім того, система виконує підготовку підсумків інциденту на основі проведеного аналізу, що допомагає аналітикам швидко отримувати необхідні висновки без потреби в ручному опрацюванні великих обсягів даних.

Окремий модуль рішення виконує аналіз логів, виділяючи ключові аспекти події та надаючи короткі, структуровані висновки, які можуть бути використані для оперативного реагування. Це дозволяє значно скоротити час, необхідний для оцінки загрози, а також мінімізувати ймовірність пропуску критично важливих деталей під час розслідування інциденту.

2.6 Огляд моделей LLM, для обробки інцидентів інформаційної безпеки

Зростання обсягів інцидентів кібербезпеки та збільшення складності аналізу супровідної інформації обумовили потребу у нових підходах до обробки інцидентів.

Зокрема, традиційні методи аналізу логів, пошуку технічної інформації та складання підсумків розслідувань часто виявляються неефективними в умовах сучасних атак та великих обсягів даних. Поява великих мовних моделей (LLM) відкрила нові можливості для підвищення ефективності процесів обробки інцидентів за рахунок автоматизації збору, узагальнення та аналізу інформації [39].

Використання LLM в обробці інцидентів безпеки

Роль великих мовних моделей у сфері інформаційної безпеки сьогодні полягає не лише у виявленні аномалій, а й у підтримці аналітиків шляхом автоматизації рутинних та інтелектуально навантажених завдань:

Структурування великих обсягів логів та подій безпеки.

Виявлення технічних вразливостей за допомогою швидкого пошуку релевантної інформації.

Генерування звітів і висновків за результатами розслідування.

Огляд ключових моделей

GPT-4

Розробник: OpenAI

Дата випуску: 2023

Ключові особливості: Покращене розуміння складних логічних патернів, здатність обробляти великі обсяги даних (до 32 тис. токенів).

Актуальність: Добре підходить для створення аналітичних звітів на основі технічних даних інцидентів. Виявляє глибинні зв'язки між подіями.

Переваги: Точність, глибина аналізу, стійкість до маніпуляцій.

Недоліки: Висока вартість використання, потреба у значних обчислювальних ресурсах.

GPT-4-o

Розробник: OpenAI

Дата випуску: 2024

Ключові особливості: Мультимодальність, велике контекстне вікно (до 128 тис. токенів), оптимізована швидкість.

Актуальність: Ідеально підходить для швидкого пошуку технічної інформації та роботи з великими обсягами логів.

Переваги: Підтримка кількох форматів даних (текст, зображення), краща економічність.

Недоліки: Потребує продуманого застосування для вузькоспеціалізованих завдань.

GPT-4 Turbo

Розробник: OpenAI

Дата випуску: 2023

Ключові особливості: Вища швидкість та нижча вартість у порівнянні з базовою версією GPT-4.

Актуальність: Підходить для реального часу обробки запитів та попереднього аналізу великих логів без втрати контексту.

Переваги: Висока продуктивність, економічність.

Недоліки: Може поступатися точністю у складних логічних завданнях.

Claude 3 Sonnet

Розробник: Anthropic

Дата випуску: 2024

Ключові особливості: Велике контекстне вікно (200 тис. токенів), орієнтація на завдання кодування та аналізу текстів.

Актуальність: Добре підходить для узагальнення інцидентів на основі складних логів і технічних даних.

Переваги: Висока швидкість, здатність розуміти технічний контекст.

Недоліки: Може бути менш точною у дуже глибоких аналітичних завданнях.

Claude 3 Opus

Розробник: Anthropic

Дата випуску: 2024

Ключові особливості: Найвища когнітивна здатність серед моделей Claude, велике контекстне вікно до 1 млн токенів.

Актуальність: Ідеальна для формування комплексних розслідувань та аналізу величезних логів або всієї кодової бази.

Переваги: Дуже високий рівень міркувань, здатність до складного причинно-наслідкового аналізу.

Недоліки: Високі вимоги до обчислювальних ресурсів.

Deepseek-VL

Розробник: DeepSeek (Китай)

Дата випуску: 2024

Ключові особливості: Мультимодальна модель з підтримкою тексту, зображень та коду. Побудована на архітектурі LLM з можливістю виконання коду та глибокого аналізу візуального контенту.

Актуальність: Ідеальна для завдань, що потребують інтегрованого аналізу графічних і текстових даних, технічної документації або знімків екрана з систем, а також для пояснення коду та програмної логіки.

Переваги: Висока продуктивність у завданнях програмування, здатність аналізувати зображення, діаграми, таблиці. Підтримує складні технічні діалоги.

Недоліки: Обмежена доступність в порівнянні з моделями від західних розробників, іноді дає менш стабільні відповіді при генерації великого обсягу тексту. Високі проблеми у безпеці. База даних Deepseek знаходилась у вільному доступі [36].

Microsoft Copilot (на базі Azure AI)

Розробник: Microsoft

Дата випуску: 2023 (розвиток триває)

Ключові особливості: Інтеграція із середовищем Microsoft 365 та Azure, оптимізація для корпоративного використання, можливість роботи із захищеними даними.

Актуальність:

Формування узагальнених звітів безпосередньо у корпоративних середовищах.

Автоматизований пошук технічної документації, пов'язаної з виявленими інцидентами.

Стандартизоване складання підсумків розслідувань у відповідності до внутрішніх політик безпеки.

Переваги:

Вбудована інтеграція в корпоративні процеси (SharePoint, Teams, Outlook).

Висока безпека та контроль над даними.

Можливість швидкого пошуку внутрішньої та публічної технічної інформації.

Недоліки:

Орієнтованість на екосистему Microsoft.

Може вимагати налаштування для специфічних сценаріїв кібербезпеки.

Загальний огляд продуктивності моделей у вигляді графіків [38] які зображені на рисунках[2.8 – 2.13] :

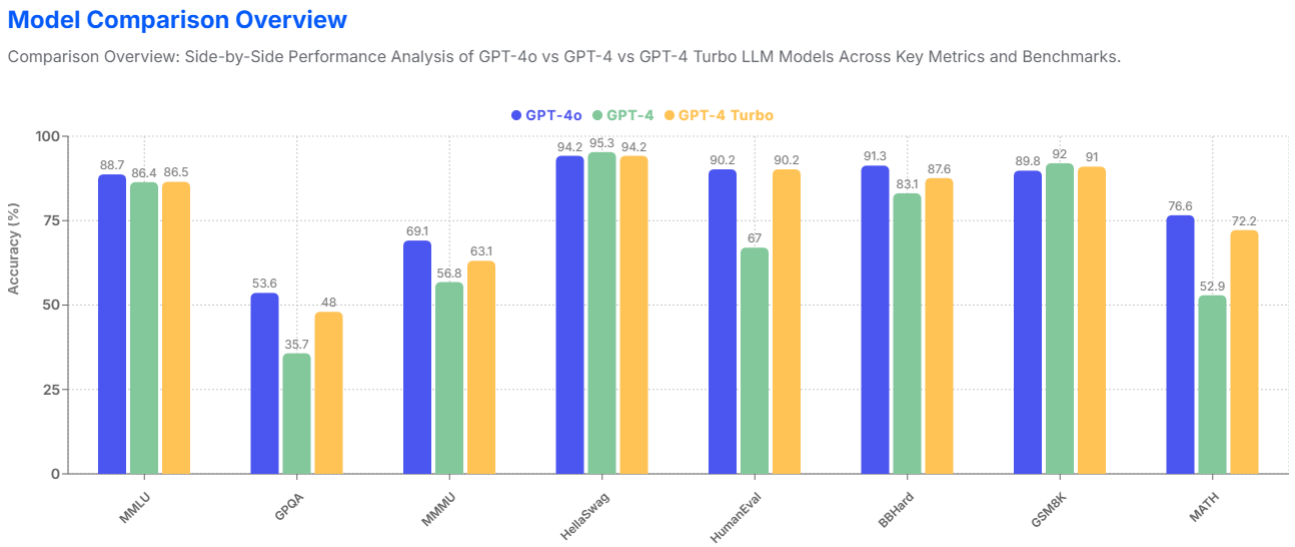


Рисунок 2.8 − Огляд порівняння моделей GPT

LLM Model Performance Overview

Performance Overview : Visualizing and Analyzing Key Metrics of Two Leading LLM Models for Performance Comparison.

MODEL	GPT-4O	GPT-4	GPT-4 TURBO
Context size	128K	32K	128K
Cutoff date	Oct 2023	November 2023	November 2023
Input/output cost	$0.005 / $0.015	$0.06 / $0.12	$0.01 / $0.03
Latency (TTFT)	0.48s	0.60s	0.60s
Throughput	80t/s	28.5t/s	28.5t/s

Рисунок 2.9 - Огляд ефективності моделей GPT

Comparing GPT-4o vs GPT-4 vs GPT-4 Turbo

A detailed comparison of GPT-4o vs GPT-4 vs GPT-4 Turbo performance and features.

BENCHMARK	GPT-4O	GPT-4	GPT-4 TURBO
MMLU	88.7%	86.4%	86.5%
GPQA	53.6%	35.7%	48%
MMMU	69.1%	56.8%	63.1%
HellaSwag	94.2%	95.3%	94.2%
HumanEval	90.2%	67%	90.2%
BBHard	91.3%	83.1%	87.6%
GSM8K	89.8%	92%	91%
MATH	76.6%	52.9%	72.2%

Рисунок 2.10 – Деталізоване порівння у бенчмарках GPT моделей

Model Comparison Overview

Comparison Overview: Side-by-Side Performance Analysis of Claude 3 Haiku vs Claude 3 Opus vs Claude 3 Sonnet LLM Models Across Key Metrics and Benchmarks.

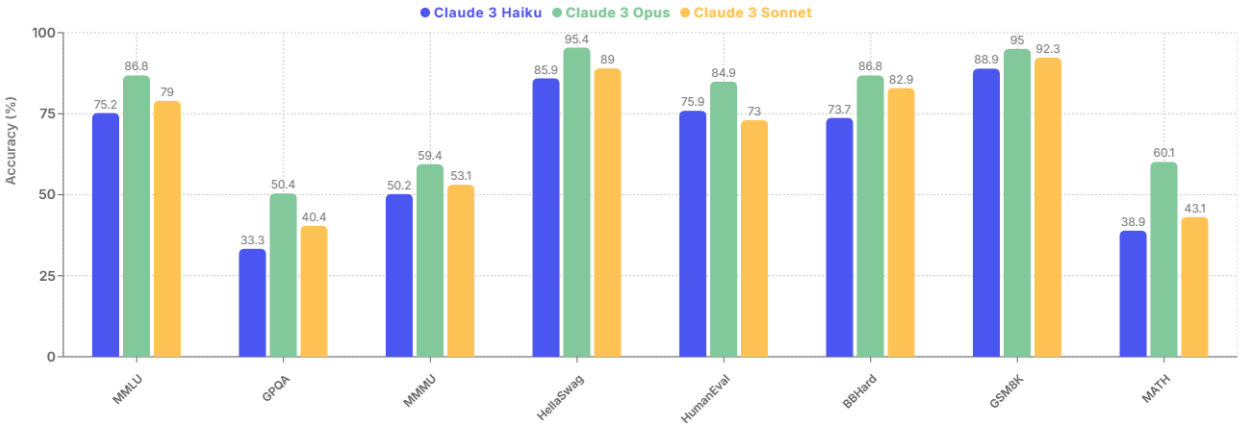


Рисунок 2.11 - Огляд порівняння моделей Claude

LLM Model Performance Overview

Performance Overview : Visualizing and Analyzing Key Metrics of Two Leading LLM Models for Performance Comparison.

MODEL	CLAUDE 3 HAIKU	CLAUDE 3 OPUS	CLAUDE 3 SONNET
Context size	200K	200K	256K
Cutoff date	March 2024	May 2024	May 2024
Input/output cost	$0.00025 / $0.00125	$0.015 / $0.075	$0.0002 / $0.0011
Latency (TTFT)	0.55s	1.66s	0.65s
Throughput	134.3t/s	23.0t/s	170.4t/s

Рисунок 2.12 - Огляд ефективності моделей Claude

Comparing Claude 3 Haiku vs Claude 3 Opus vs Claude 3 Sonnet

A detailed comparison of Claude 3 Haiku vs Claude 3 Opus vs Claude 3 Sonnet performance and features.

BENCHMARK	CLAUDE 3 HAIKU	CLAUDE 3 OPUS	CLAUDE 3 SONNET
MMLU	75.2%	86.8%	79%
GPQA	33.3%	50.4%	40.4%
MMMU	50.2%	59.4%	53.1%
HellaSwag	85.9%	95.4%	89%
HumanEval	75.9%	84.9%	73%
BBHard	73.7%	86.8%	82.9%
GSM8K	88.9%	95%	92.3%
MATH	38.9%	60.1%	43.1%

Рисунок 2.13 – Деталізоване порівння у бенчмарках Claude моделей

У таблиці 2.2 наведено узагальнене порівняння популярних моделей штучного інтелекту за трьома основними аспектами обробки інцидентів інформаційної безпеки: формування опису інциденту, пошук технічної інформації, та підбиття підсумків розслідування.

Таблиця 2.2

Порівняння LLM для обробки інцидентів інформаційної безпеки

Модель	Формування опису інциденту	Пошук технічної інформації	Підбиття підсумків розслідування
GPT-4	Добре (детальність, глибоке розуміння контексту)	Задовільно (обмежений пошук)	Відмінно (логічне структурування результатів)
GPT-4-o	Відмінно (швидкість + якість балансуються)	Відмінно (швидкий пошук і аналіз інформації)	Добре (дещо слабше стратегічне узагальнення)
Claude 3 Opus	Добре (якісний текст, особливо для бізнес-стилю)	Добре (іноді поверхнева технічна деталізація)	Відмінно (відмінне логічне оформлення підсумків)
Claude 3 Sonnet	Добре (розумний опис, але менш точний)	Добре (підходить для типових запитів)	Задовільно (обмежене узагальнення складних кейсів)
GPT-4 Turbo	Добре (економія ресурсів, стабільна якість)	Добре (адекватно для стандартних запитів)	Задовільно (узагальнення менш послідовне)
Microsoft Copilot	Добре (зручний інтегрований опис у середовищах M365)	Задовільно (пошук сильно залежить від джерел)	Добре (підходить для типових корпоративних звітів)

Результатом огляду було прийняте рішення про необхідність інтеграції декількох LLM-моделей у програмний модуль для можливості проведення відповідного тестування та порвінняя засновані на реальних кейсах. Це надасть змогу цільовій команії самостійно обирати необхідну модель зважаючи на технічні та нетехнічні вимоги.

Висновки до другого розділу

У другому розділі було здійснено комплексний аналіз можливостей використання штучного інтелекту при обробці інцидентів інформаційної безпеки. Основна увага приділялася аналізу рішень на основі ШІ, які пропонуються провідними вендорами у сфері кібербезпеки, зокрема Microsoft, Checkpoint, OpenAI, Anthropic та DeepSeek. Було встановлено, що ШІ здатний значно підвищити

ефективність реагування на інциденти, зменшити навантаження на аналітиків безпеки та автоматизувати частину рутинних завдань, таких як пошук технічної інформації, класифікація подій або формування підсумків розслідувань.

Важливою частиною дослідження стала оцінка практичної ефективності впроваджених рішень. Наприклад, результати роботи систем Microsoft Copilot for Security вказують на реальне покращення продуктивності аналітиків SOC завдяки автоматизації аналізу інцидентів та оптимізації прийняття рішень. Також було проведено глибокий аналіз хибних спрацювань у системах, таких як Checkpoint XDR/XPR, що дозволило зробити висновки щодо обмежень сучасних ШІ-систем у завданнях повноцінного виявлення загроз.

Крім того, у цьому розділі були розглянуті можливості великих мовних моделей (LLM), таких як GPT-4, Claude 3, Microsoft Copilot, Deepseek, у контексті їх застосування для структуризації логів, пошуку інформації про вразливості та стандартизованого формування описів інцидентів. Проведене дослідження засвідчило, що LLM-моделі демонструють високу ефективність у задачах узагальнення та пошуку, проте потребують людського контролю під час аналізу складних або контекстно неоднозначних кейсів.

Окремим результатом стало виявлення обмежень сучасних NLP- та LLM-рішень у завданнях первинного виявлення інцидентів — через велику кількість хибних спрацювань, відсутність глибокого розуміння контексту подій та необхідність комбінування з іншими інструментами (такі як EDR, SIEM або аналітика поведінки).

Загалом, проведене дослідження підтвердило потенціал штучного інтелекту як інструмента підсилення діяльності центрів реагування на інциденти, особливо в аспектах прискорення розслідувань, скорочення часу на обробку та покращення якості звітності. Проте, на поточному етапі розвитку, ШІ не може повністю замінити аналітиків або автоматизувати процес виявлення інцидентів без значного ризику

помилок — натомість його слід розглядати як помічника у вже верифікованих випадках.

РОЗДІЛ 3
РОЗРОБКА МЕТОДУ ОБРОБКИ ІНЦИДЕНТІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

3.1. Визначення вимог для створення програмного модуля

Зважаючи на результати аналізу, представлені у попередніх розділах, для досягнення поставленої мети роботи — покращення процесу обробки інцидентів інформаційної безпеки із використанням сучасних технологій штучного інтелекту — було сформульовано конкретний перелік цільових задач, які має вирішувати розроблений програмний модуль. Ці задачі спрямовані на підвищення ефективності обробки інцидентів, зниження часу на аналіз та мінімізацію людського фактора при виконанні рутинних операцій.

До основних задач віднесено наступні напрямки:

1. Формування узагальненої інформації на основі вхідних логів та технічних даних за допомогою NLP-моделей.

Використання моделей обробки природної мови (Natural Language Processing) дозволяє автоматизувати процес перетворення великого обсягу сирих даних у зрозумілий опис інциденту. Це сприяє пришвидшенню первинного аналізу, дозволяючи фахівцям зосередитись на прийнятті рішень замість ручного аналізу логів.

2. Автоматизований пошук технічної інформації про виявлений інцидент за допомогою AI-асистента.

У рамках цього завдання передбачається інтеграція інтелектуальної системи пошуку, яка здатна знаходити релевантні статті, CVE-записи, офіційну документацію та інші джерела технічної інформації на основі короткого опису або характерних

ознак інциденту. Це дозволяє оперативно отримувати необхідні дані для аналізу загроз і вибору стратегії реагування.

3. Підбиття стандартизованих підсумків розслідування інциденту відповідно до ключових методик життєвого циклу інциденту згідно рекомендацій найкращих практик.

Стандартизація підсумків дозволяє забезпечити єдність у представленні результатів розслідувань, спрощуючи комунікацію всередині команди безпеки та між різними підрозділами. Формування підсумкових звітів відповідно до життєвого циклу обробки інцидентів (наприклад, за моделями NIST або SANS) також полегшує подальший аналіз, виявлення слабких місць та формування рекомендацій для покращення захисту.

Вибір саме цих задач обґрунтований виявленими проблемами у функціонуванні наявних систем штучного інтелекту, зокрема значною кількістю хибних спрацювань при виявленні інцидентів. Сучасні ШІ-рішення у сфері кібербезпеки ще не здатні повноцінно замінити аналітика у прийнятті складних рішень, однак вони можуть суттєво полегшити допоміжні процеси: структурування даних, прискорення доступу до знань та формування якісних аналітичних висновків.

Таким чином, запропоновані задачі є логічною відповіддю на актуальні виклики у сфері інформаційної безпеки та спрямовані на інтеграцію штучного інтелекту у критично важливі, але рутинні етапи обробки інцидентів.

Додатковою задачею в рамках реалізації запропонованого підходу є доцільність інтеграції розробленого програмного модуля на базі штучного інтелекту безпосередньо у вже існуючу на підприємстві систему інцидент-менеджменту. Такий підхід дозволяє мінімізувати затримки у передачі інформації між окремими інформаційними системами, що особливо критично у випадку обробки інцидентів інформаційної безпеки в реальному часі. Крім того, тісна інтеграція сприяє безперервному збагаченню бази знань інцидент-менеджменту актуальними даними,

що у свою чергу підвищує якість аналітики та адаптивність роботи штучного інтелекту.

Важливим етапом інтеграції є обґрунтований вибір конкретної моделі великої мовної моделі (LLM), яка буде використовуватися для генерації відповідей та обробки запитів. Окрім класичних технічних критеріїв — таких як точність, швидкість роботи, масштабованість та можливість налаштування під специфіку завдань підприємства — слід враховувати також нетехнічні аспекти вибору моделі. Зокрема, значний вплив мають юридичні, етичні та регуляторні обмеження, що діють у юрисдикціях, де оперує підприємство.

Прикладом таких обмежень є ситуація з деякими китайськими LLM-моделями [38], які підлягають повному блокуванню або значним обмеженням у США, Італії, Південній Кореї, Австралії та Тайвані. Відповідно, вибір моделі, яка потенційно може не відповідати вимогам регуляторів або міжнародним стандартам безпеки даних (наприклад, GDPR, HIPAA, NIST 800-53 тощо), може становити серйозний ризик для цільового підприємства як у юридичному, так і в репутаційному вимірах.

Додатково, рівень інтеграції моделі у наявну IT-інфраструктуру підприємства суттєво впливає на вибір постачальника штучного інтелекту. Наприклад, якщо підприємство активно використовує екосистему рішень Microsoft для управління та безпеки IT-середовища (Azure Active Directory, Microsoft Sentinel, Microsoft Defender XDR тощо), більш логічним і безпечним рішенням буде інтеграція LLM-моделі, що належить Microsoft (наприклад, Azure OpenAI Service). Це дозволить зберегти усі дані в межах довіреного корпоративного середовища без потреби залучення сторонніх провайдерів, мінімізуючи ризики витоку чи несанкціонованого доступу до конфіденційної інформації.

Таким чином, вибір конкретної моделі для інтеграції є не лише технічним, а й стратегічним рішенням, що має бути здійснене на основі комплексної оцінки технічних характеристик, регуляторних вимог, специфіки бізнес-процесів підприємства та архітектури існуючої системи безпеки. Результатом огляду

продуктивності моделей було прийняте рішення про необхідність інтеграції декількох LLM-моделей у програмний модуль для можливості проведення відповідного тестування та порвінняня засновані на реальних кейсах, окрім DeepSeekAI. Це надасть змогу цільовій команії самостійно обирати необхідну модель зважаючи на технічні та нетехнічні вимоги.

3.2 Архітектура та технології веб-додатку для обробки інцидентів інформаційної безпеки

Веб-додаток створений із використанням сучасних технологій, орієнтованих на швидку розробку, масштабованість і інтеграцію моделей штучного інтелекту для аналізу даних інцидентів. Його функціонал зосереджений на трьох основних завданнях: формування узагальненої інформації за логами та технічними даними, автоматизованому пошуку супровідної технічної інформації та стандартизованому підбитті підсумків розслідувань інцидентів.

Клієнтська частина

Інтерфейс користувача побудований за допомогою React та TypeScript [40], що забезпечує динамічність, типову безпеку та зручність розробки. Стилізація виконана через Tailwind CSS [41], що дозволяє швидко адаптувати вигляд додатку до різних сценаріїв використання.

- React забезпечує гнучку побудову сторінок та інтерфейсних елементів на основі змін стану додатку, що важливо для інтерактивного відображення результатів аналізу інцидентів.
- TypeScript підвищує надійність за рахунок суворої типізації даних, що є критичним для роботи з результатами AI-обробки логів та запитів.
- Tailwind CSS дозволяє швидко створювати адаптивний дизайн, що підтримує різні режими перегляду інформації про інциденти.

Серверна частина

На серверній стороні використовується Next.js із можливістю серверного рендерингу (SSR) та статичної генерації (SSG) [42], що підвищує продуктивність і дозволяє краще індексувати результативні звіти про обробку інцидентів.

- Серверна логіка відповідає за отримання вхідних логів і технічних даних, їх попередню обробку та передачу у відповідні AI-модулі.
- Для формування узагальненої інформації використовуються NLP-моделі, що витягують ключові події, сутності та аномалії з текстових логів.
- Пошук технічної інформації реалізований через інтеграцію AI-асистента, який автоматично знаходить релевантні CVE-записи, технічну документацію або аналітичні статті.
- Стандартизовані підсумки розслідування формуються згідно з ключовими методиками життєвого циклу інцидентів (NIST, SANS), що забезпечує якісне оформлення результатів розслідування.

Всі отримані результати зберігаються у базі даних SQLite, що дозволяє швидко отримувати інформацію для подальшого аналізу або побудови звітів.

Загальна структура проекту

Проєкт має чітко організовану файлову структуру для забезпечення простоти розробки і масштабованості:

- /components/ — інтерфейсні компоненти React.
- /services/ — модулі роботи з API AI-сервісів і пошуку інформації.
- /utils/ — утиліти для обробки даних логів та форматування результатів.
- /pages/ — маршрути додатка з реалізацією SSR для основних сторінок аналізу інцидентів.
- /public/ — статичні ресурси.

- /db/ — конфігурація та управління базою даних SQLite.

Інтеграція з системами безпеки для обробки інцидентів

На рисунку 3.1 зображено основну ідею системної частини процесу обробки інциденту.

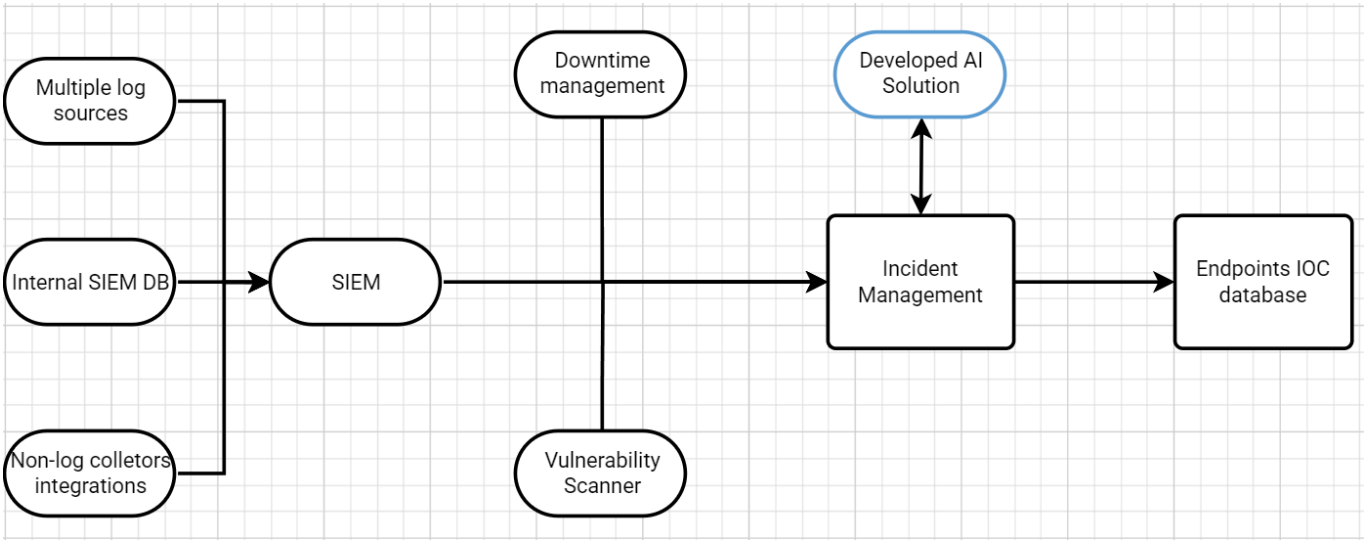


Рисунок 3.1 – Системна частина процесу обробки інциденту

SIEM збирає логи з кінцевих пристроїв [43], мережевих, хмарних та кінцевих пристроїв та на основі правил написаних аналітиками безпеки створює сповіщення у систему менеджементу інцидентів. Аналітик що буде розслідувати інцидент має змогу одразу скористатися рішенням штучного інтелекту для отримання узагальненої інформації про інцидент, де окрім пошуку інформації по відкритим джерелам може бути присутня інформація з SIEM правила та інших підключених баз даних компанії, таких як Сканер вразливостей та система менеджменту зупинки роботи обладнання. Це корисно оскільки не завжди аналітик що реагує на інцидент матиме детальне розуміння правила, оскільки правило могло бути створено іншим аналітиком.

При наявній інформації про вразливість яка експлутається можливо використати наступну функцію системи – перевірка вразливості, яка окрім загальної

короткої та точної інформації з відкритих джерел інстуту стандарів та технологій. Ця інформація включатиме в себе тип вразливості, вразливе програмне забезпечення та його версію, необхідні дії для вирішення вразливості.

При успішній інтеграції з системою управління вразливостями можна отримати інформацію чи є кінцевий хост вразливий. Це є актуальним [44] оскільки мережеве обладнання, як правило, не має інформації про цільові характеристики кінцевих точок, та може приймати легітимний інтернет трафік за спробу використання вразливості у ПЗ яке навіть не встановлене на кінцевому хості. Відповідно, отримавши цю інформацію аналітик може зменшити або збільшити пріоритет спрацювання правила та сконцентрувати увагу на тому що є більш доцільним.

В результаті проведення повноцінного дослідження інциденту розроблене рішення буде корисним для підбиття стандартизованих підсумків розслідування інциденту відповідно до ключових методик життєвого циклу інциденту згідно рекомендацій найкращих практик, що значно пришвидшить цей процес та дозволить демонструвати оброблені інциденти на подальших аудитах маючи повну картину життєвого циклу інциденту.

3.3 Опис розробленого методу

Результатом проведених досліджень став метод обробки інцидентів інформаційної безпеки з використанням LMM-моделей штучного інтелекту, який зоюражений схематично на рисунку 3.2

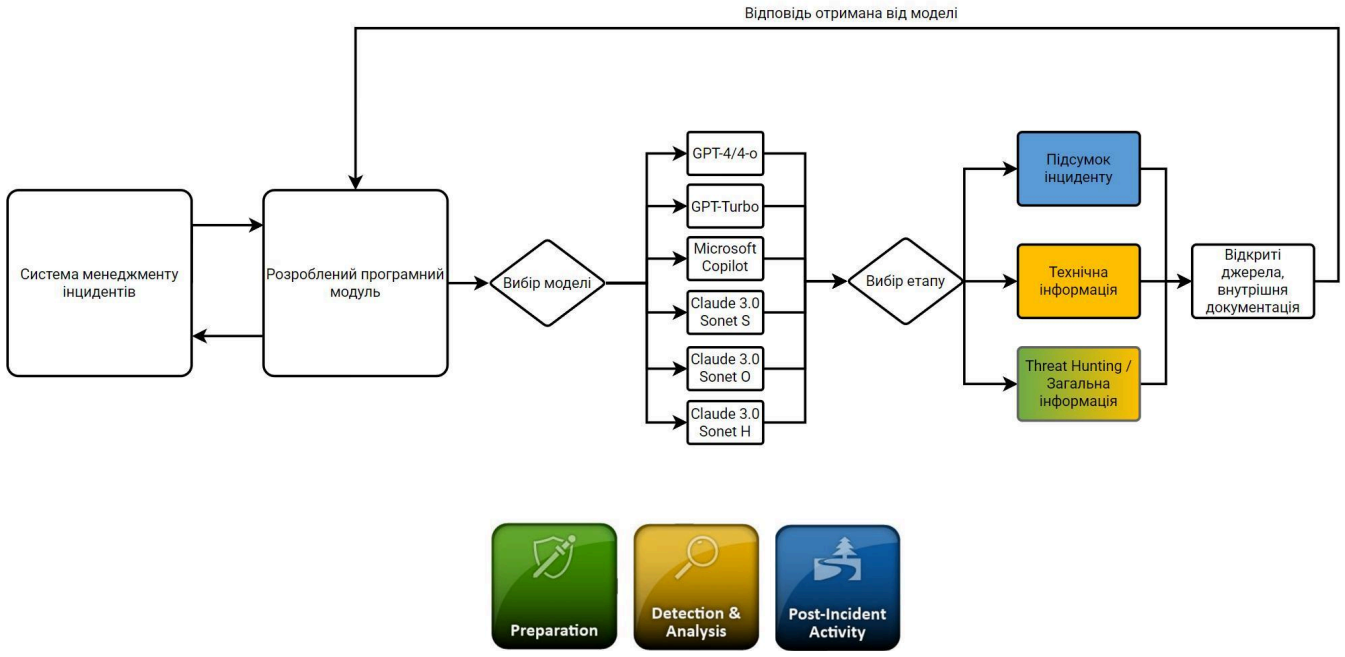


Рисунок 3.1 – Метод обробки інцидентів інформаційної безпеки з використанням LMM-моделей штучного інтелекту

Запропонований метод інтеграції штучного інтелекту у процес обробки інцидентів інформаційної безпеки базується на автоматизованій взаємодії між системою менеджменту інцидентів (наприклад, SIEM або SOAR-платформою), розробленим програмним модулем та потужними мовними моделями, що застосовуються для підтримки прийняття рішень у сфері кібербезпеки [45].

Процес починається з того, що вхідні дані щодо інциденту надходять безпосередньо з системи менеджменту інцидентів, яка вже зафіксувала підозрілу активність або попередження. Ці дані можуть включати технічні деталі, логі з мережевих пристроїв, інформацію про задіяні хости, версії програмного забезпечення, коди спрацювання IPS/IDS тощо.

Отримана інформація передається у розроблений програмний модуль, який виступає посередником між джерелом інциденту та штучним інтелектом. Цей модуль відповідає за:

- формування структурованого запиту (prompt engineering),
- вибір відповідної мовної моделі залежно від задачі,

- фільтрацію, шифрування (за потреби) та маршрутизацію трафіку через відкриті API — для взаємодії з хмарними моделями, і закриті — для внутрішніх інтеграцій з корпоративними системами.

Після обробки запиту користувачем або автоматикою, модуль здійснює вибір однієї з мовних моделей. На даному етапі можуть бути використані:

- GPT-4 або GPT-4-o — у випадках, коли необхідний глибокий аналіз логів, формування комплексного висновку або узагальнення з великого обсягу даних.
- GPT-Turbo — для пришвидшеного оброблення запитів з меншою складністю.
- Microsoft Copilot for Security — якщо інцидент пов’язаний з екосистемою Microsoft (Defender XDR, Sentinel, Intune), що забезпечує кращу контекстну інтеграцію.
- Claude 3.0 Sonet (Sonet, Opus, Hoku) — для задач, пов’язаних із генерацією пояснень, рекомендацій або класифікацією подій.

Далі обирається етап життєвого циклу інциденту, для якого потрібна підтримка. У реалізації представлені три основні напрями:

- Підсумок інциденту (Post-Incident Activity) — генерація короткого та формалізованого звіту з урахуванням стандартів звітності (наприклад, MITRE ATT\&CK, STIX 2.1).
- Технічна інформація (Detection & Analysis) — деталізація вразливості, перевірка CVE, технічні характеристики протоколів тощо.
- Threat Hunting / Загальна інформація (Preparation) — підтримка на ранніх етапах розвідки загроз, виявлення потенційних векторів атаки або аналіз подібних випадків із відкритих джерел (OSINT) та внутрішньої бази знань.

Отримана відповідь наразі додатково залишається у розробленому програмному модулі, для потенційного аналізу якості майбутніх відповідей [46]. Після цього результат повертається до системи менеджменту інцидентів — як окремий запис або аналітична примітка. Це забезпечує:

- збереження повного ланцюжка подій,
- можливість повторного аналізу з урахуванням контексту,

- ескалацію або закриття інциденту за визначеним SLA.

Таким чином, реалізований підхід поєднує можливості сучасних мовних моделей з наявною інфраструктурою підприємства, суттєво підвищуючи продуктивність аналітиків SOC, прискорюючи час реагування та забезпечуючи якісне документування інцидентів. Модель чітко відповідає фазам NIST Cybersecurity Framework [47] — Preparation, Detection & Analysis, Post-Incident Activity — і може бути масштабована відповідно до зростаючих потреб організації.

3.4 Огляд розробленого програмного модуля

Розроблений програмний модуль, з метою наглядності, створений з можливістю текстового введеняя даних та отримання відповіді безпосередньо у самому модулі, що залишає можливість його роботи як окремої системи, без додаткових модулів описаних в архітектурі.

Цей модуль має приємний та зручний інтерфейс, який дозволяє швидко перемикатися між наявними задачами як зображено на рисунку 3.2.

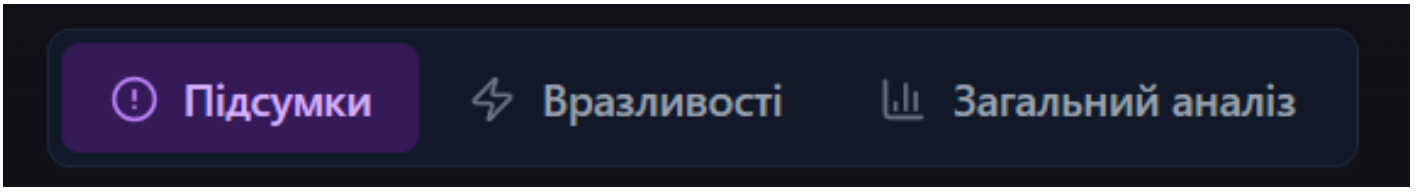


Рисунок 3.2 – Вибір необіхдної задачі у програмному модулі

Для ручного керування веб-застосунком використовується вікно вибору моделі, як на рисунку 3.3.

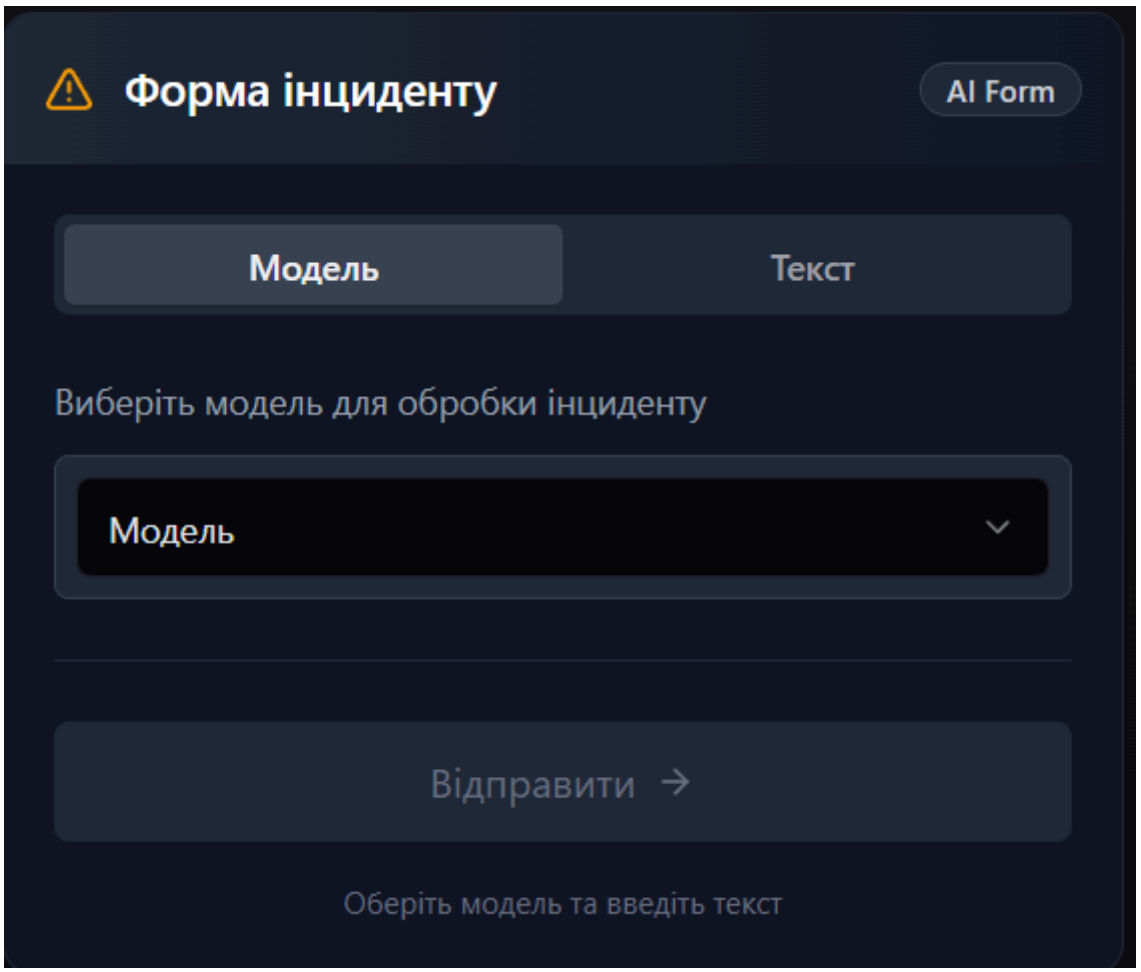


Рисунок 3.3 – Вибір необхідної моделі

У модулі при ручному введені наявне поле введення тексту як на рисунку 3.4.

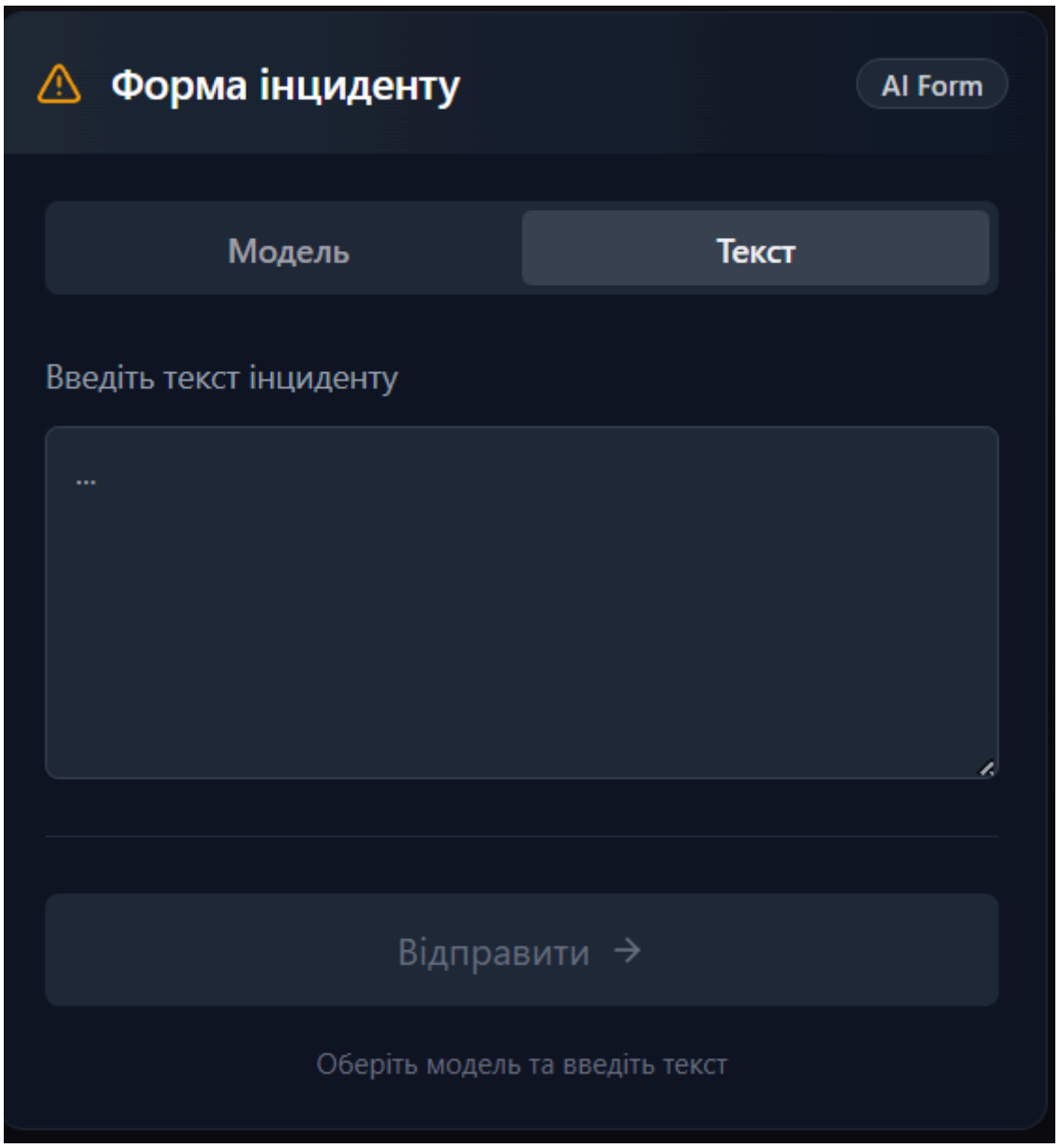


Рисунок 3.4 – Поле введення тексту

Форма ручного керування аналогічна для усіх типів задач. Загальний вигляд користувацького інтерфейсу програмного модуля зображений на рисунку 3.5.

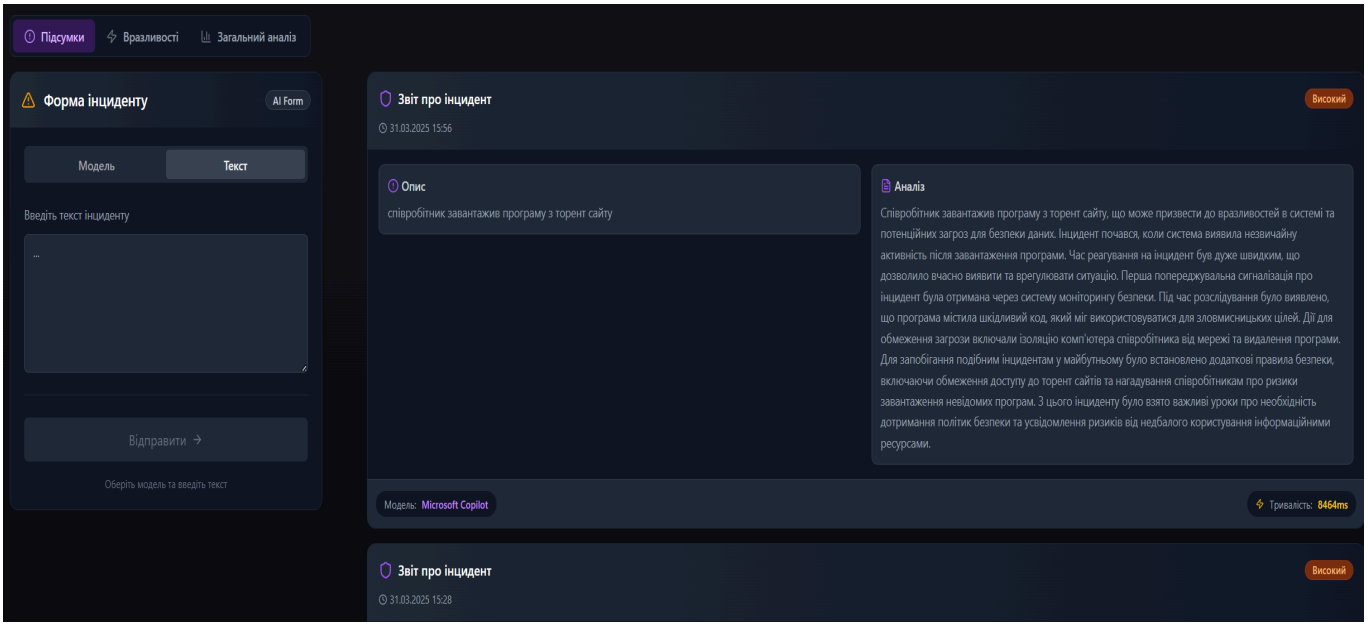


Рисунок 3.5 – Загальний вигляд користувацького інтерфейсу програмного модуля

Загальна структура передбачає відоборжаення короткого опису або ідентифікатора інциденту який розглядається, та отриману відповідь від штучного інтелекту відображається у графі аналіз. Представлена також інформація про обрану модель та тривалість відповіді. Оскільки коректна критичність інциденту впливає на час обробки [48] класифікація критичності інциденту буде спочатку автоматично визначена штучним інтелектом, але при необхідності може бути змінена аналітиком вручну.

Висновок до третього розділу

У третьому розділі було реалізовано ключове завдання дослідження — розробку методу обробки інцидентів інформаційної безпеки з використанням штучного інтелекту та його інтеграцію у програмний модуль. Запропонований підхід базується на автоматизованій взаємодії між системами менеджменту інцидентів (SIEM/SOAR), мовними моделями великого масштабу (LLM) та спеціалізованим

програмним забезпеченням, що виконує функції посередника, аналітика та генератора звітів.

Визначено вимоги до функціональності програмного модуля, зокрема автоматизоване формування узагальненої інформації з логів, пошук технічних даних про інциденти та стандартизоване підбиття підсумків розслідувань.

Розроблено архітектуру веб-додатку, що забезпечує інтеграцію з сучасними AI-моделями та корпоративною інфраструктурою безпеки.

Представлено метод обробки інцидентів, який охоплює всі ключові фази життєвого циклу інциденту згідно з NIST Cybersecurity Framework: Preparation, Detection & Analysis, Post-Incident Activity.

Реалізовано програмний модуль, що дозволяє як автоматизовану, так і ручну взаємодію з LLM-моделями, забезпечуючи гнучкість, масштабованість та відповідність вимогам безпеки.

Таким чином, результати, отримані в цьому розділі, демонструють практичну реалізацію концепції використання штучного інтелекту для підвищення ефективності обробки інцидентів інформаційної безпеки, що є важливим кроком до створення адаптивних, інтелектуально підтримуваних систем кіберзахисту.

РОЗДІЛ 4
АНАЛІЗ РЕЗУЛЬТАТІВ ВИКОРИСТАННЯ ПРОГРАМНОГО МОДУЛЯ ОБРОБКИ ІНЦИДЕНТІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

4.1 Перевірка результатів на прикладі оброблених інцидентів

Для тестування ефективності обрано розслідувані інциденти на підприємстві.

Як описано в першому розділі, згідно спеціальної публікації NIST SP 800-61r2 [49] у підсумку вирішення інциденту має відповідати плану дій наведеному на рисунку 4.1, що відповідає підсумкам які мають бути описані аналітиком.

	Action	Completed
	Detection and Analysis	
1.	Determine whether an incident has occurred	
1.1	Analyze the precursors and indicators	
1.2	Look for correlating information	
1.3	Perform research (e.g., search engines, knowledge base)	
1.4	As soon as the handler believes an incident has occurred, begin documenting the investigation and gathering evidence	
2.	Prioritize handling the incident based on the relevant factors (functional impact, information impact, recoverability effort, etc.)	
3.	Report the incident to the appropriate internal personnel and external organizations	
	Containment, Eradication, and Recovery	
4.	Acquire, preserve, secure, and document evidence	
5.	Contain the incident	
6.	Eradicate the incident	
6.1	Identify and mitigate all vulnerabilities that were exploited	
6.2	Remove malware, inappropriate materials, and other components	
6.3	If more affected hosts are discovered (e.g., new malware infections), repeat the Detection and Analysis steps (1.1, 1.2) to identify all other affected hosts, then contain (5) and eradicate (6) the incident for them	
7.	Recover from the incident	
7.1	Return affected systems to an operationally ready state	
7.2	Confirm that the affected systems are functioning normally	
7.3	If necessary, implement additional monitoring to look for future related activity	
	Post-Incident Activity	
8.	Create a follow-up report	
9.	Hold a lessons learned meeting (mandatory for major incidents, optional otherwise)	

Рисунок 4.1 – Контрольний список обробки інцидентів[L]

Розглянемо приклад реального інциденту інформаційної безпеки, що стався на цільовому підприємстві та продемонстрував ефективність поєднання технічних засобів захисту, автоматизованих рішень і людського контролю. У межах даного випадку було зафіксовано фішингову атаку, в результаті якої один із працівників, отримавши електронного листа, що імітував повідомлення від знайомого контакта, відкрив вміщене посилання на нібито офіційний документ для підпису. Посилання вело на фальшиву вебсторінку, стилізовану під портал авторизації Microsoft Office 365. Не підозрюючи загрози, користувач ввів облікові дані й пройшов другий фактор аутентифікації, що дозволило зловмисникам отримати повний доступ до його корпоративного хмарного середовища, включаючи сервіси Office, OneDrive та інші пов’язані ресурси. Після отримання доступу зловмисники переглянули декілька внутрішніх документів до того моменту, коли обліковий запис був автоматично заблокований, сесії завершені, а фактори автентифікації – скинуті.

Цей інцидент засвідчив належну роботу напівавтоматизованого процесу реагування, інтегрованого з SIEM-системою, поштовим фільтром і службою безпеки. Лист, надісланий одночасно декільком користувачам, був швидко помічений як фішинговий і переданий до центру операційної безпеки. З певною затримкою, але в межах прийнятного вікна реагування, системи поштової безпеки проаналізували посилання та класифікували його як шкідливе. Після успішного входу до облікового запису з нетипових IP-адрес та зміненого браузерного агента система зафіксувала аномалії та ініціювала тимчасове блокування облікового запису, що зупинило подальші дії зловмисника. Після ручного втручання фахівців SOC було підтверджено факт компрометації, здійснено повний аналіз активності з підозрілих адрес, вилучено шкідливі листи з поштових скриньок інших працівників, заблоковано відправника та передано фішинговий зразок до вендора поштового фільтра. Фішингову URL-адресу також заблоковано на проксі-сервері й системі CASB, а зібрані індикатори компрометації додано до системи управління

інцидентами. Було ініційовано індивідуальне навчання користувача, який потрапив на фішинговий гачок, а також започатковано кампанію з підвищення обізнаності серед персоналу.

Підсумковий звіт, підготовлений аналітиком SOC, містив підтвердження інциденту, опис вектора атаки, дії зловмисників, хронологію подій та перелік вжитих заходів, зокрема: блокування сеансів, скидання пароля та MFA, анулювання доступів, блокування посилань, вилучення шкідливого контенту, створення запитів на оновлення фільтрації та запуск внутрішнього інструктажу. Було також запропоновано подальші дії щодо блокування хостинг-доменів, які використовуються для розміщення подібних фішингових сторінок. У фінальному висновку наголошено на важливості вдосконалення систем поштової фільтрації та підвищення обізнаності користувачів для зниження ймовірності подібних інцидентів у майбутньому. Висновок від аналітика у перекладі з англійської : “Ми отримали повідомлення від користувача про фішинговий електронний лист.

Під час нашого розслідування було виявлено, що це був фішинговий електронний лист. Електронний лист містить фішингову URL-адресу, яка веде на фальшиву сторінку входу в Office 365 та вимагає облікових даних.

Під час розслідування ми отримали додаткове сповіщення про підозрілий вхід користувача, який стався після натискання URL-адреси у цьому фішинговому електронному листі. Зловмисники змогли отримати доступ до його облікового запису та переглянули 3 файли, що зберігалися в хмарі, перш ніж ми заблокували обліковий запис користувача. Його сеанси скасовано. Скидання MFA, скидання пароля. Створено запит для служби підтримки для допомоги з новим налаштуванням.

Фішингову URL-адресу заблоковано на проксі-сервере, відправника заблоковано на поштовому шлюзі. Діапазони IP-адрес зловмисників заблоковано на CASB. Повідомлення видалено з поштової скриньки користувача. IOC додано до IM. Електронний лист повідомлено до вендорів. Було створено завдання з виправлення ситуації, щоб перевірити можливість блокування доменів хостингу додатків. Було призначено навчання з фішингу.

Висновок: покращення обізнаності користувачів та покращення системи фільтрації електронної пошти.”

Надамо системі усі наявні записи з системи менеджменту інцидентів розробленому програмному модулю та перевіримо наданий висновок, на рисунку 4.2 можливо побачити що у даному випадку використовувалась модель Microsoft Copilot і час відповіді склав приблизно шість секнуд. Система автоматично визначила рівень важливості інциденту як високий, що відповідає важливості яка була визначена аналітиком після результатів обробки.

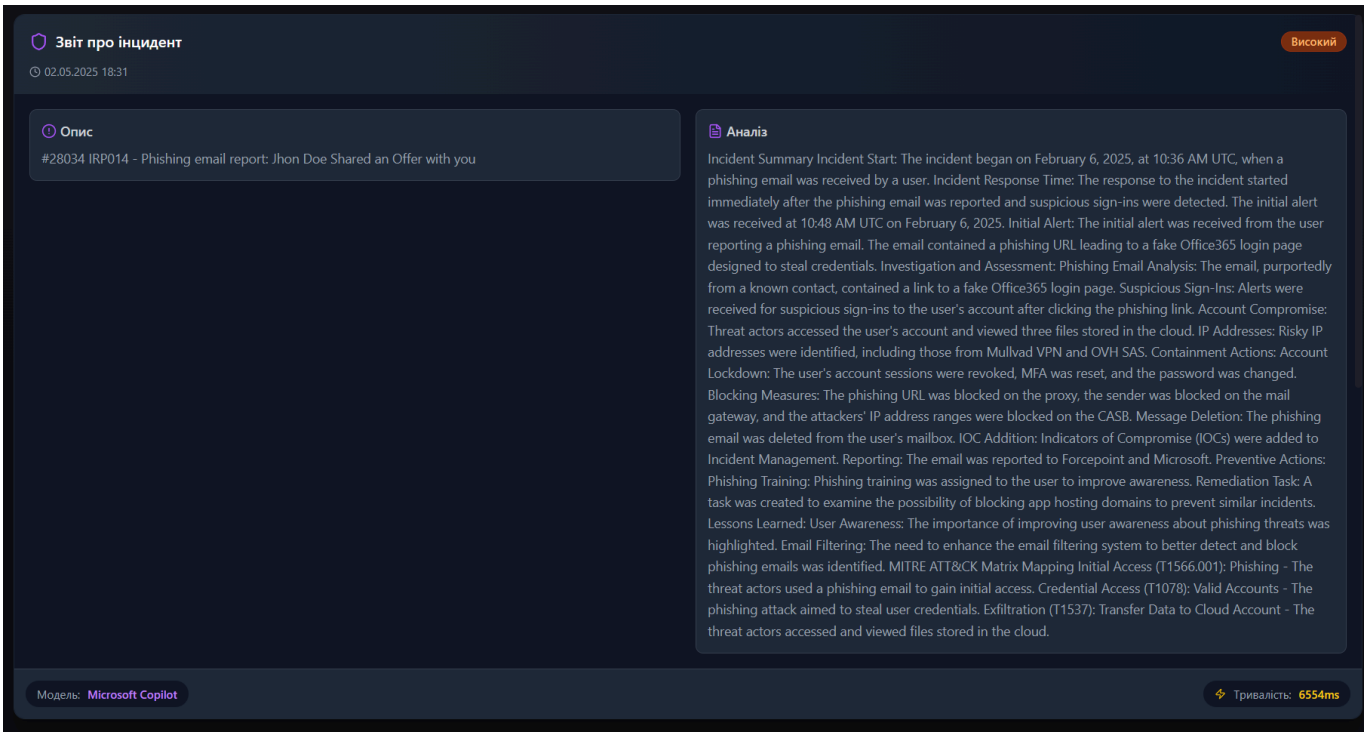


Рисунок 4.2 – Отримана відповідь від програмного модуля

У перекладі маємо наступну відповідь від програмного модуля : «

Короткий опис інциденту

Початок інциденту: Інцидент розпочався 6 лютого 2025 року о 10:36 UTC, коли користувач отримав фішинговий електронний лист.

Час реагування на інцидент: Реакція на інцидент розпочалася одразу після повідомлення про фішинговий електронний лист та виявлення підозрілих входів. Початкове сповіщення було отримано о 10:48 UTC 6 лютого 2025 року.

Початкове сповіщення: Початкове сповіщення було отримано від користувача, який повідомив про фішинговий електронний лист. Електронний лист містив фішингову URL-адресу, що вела на підроблену сторінку входу Office365, призначену для крадіжки облікових даних.

Розслідування та оцінка:

Аналіз фішингового електронного листа: Електронний лист, нібито від відомого контакту, містив посилання на підроблену сторінку входу Office365.

Підозрілі входи: Було отримано сповіщення про підозрілі входи до облікового запису користувача після натискання фішингового посилання.

Компрометація облікового запису: Зловмисники отримали доступ до облікового запису користувача та переглянули три файли, що зберігаються в хмарі.

IP-адреси: Було виявлено ризиковані IP-адреси, зокрема від Mullvad VPN та OVH SAS.

Заходи щодо стримування:

Блокування облікового запису: Сеанси облікового запису користувача було скасовано, багатофакторну автентифікацію (MFA) було скинуто, а пароль змінено.

Заходи блокування: Фішингову URL-адресу було заблоковано на проксі-сервері, відправника було заблоковано на поштовому шлюзі, а діапазони IP-адрес зловмисників було заблоковано на CASB.

Видалення повідомлення: Фішинговий електронний лист було видалено з поштової скриньки користувача.

Додавання IOC: Індикатори компрометації (IOC) було додано до управління інцидентами.

Стримання: Про електронний лист було повідомлено до служб захисту поштового шлюзу.

Профілактичні дії:

Навчання фішингу: Користувачеві було призначено навчання фішингу для підвищення обізнаності.

Завдання з виправлення: Було створено завдання для вивчення можливості блокування доменів хостингу додатків для запобігання подібним інцидентам.

Винесені уроки:

Обізнаність користувачів: Було підкреслено важливість підвищення обізнаності користувачів про фішингові загрози.

Фільтрація електронної пошти: Було визначено потребу вдосконалення системи фільтрації електронної пошти для кращого виявлення та блокування фішингових листів.

Матричне зіставлення MITRE ATT&CK

Початковий доступ (T1566.001): Фішинг - Зловмисники використовували фішинговий електронний лист для отримання початкового доступу.

Доступ до облікових даних (T1078): Дійсні облікові записи - Фішингова атака була спрямована на викрадення облікових даних користувачів.

Викрадення (T1537): Передача даних до хмарного облікового запису - Зловмисники отримали доступ до файлів, що зберігаються в хмарі, та переглядали їх.»

Наведений висновок, згенерований програмним модулем на основі моделі Microsoft Copilot, є кращим за попередній текстовий опис, оскільки має ряд вагомих переваг, що відповідають сучасним вимогам до обробки та представлення інцидентів інформаційної безпеки.

По-перше, структура відповіді є чітко стандартизованою, з логічно розмежованими розділами: короткий опис інциденту, розслідування, заходи стримування, профілактика, уроки та зіставлення з MITRE ATT&CK. Такий підхід не лише полегшує розуміння змісту для аналітиків і керівництва, але й забезпечує відповідність практикам сучасного управління безпековими інцидентами, включаючи NIST, ISO/IEC 27035 та інші фреймворки.

По-друге, автоматизований звіт дозволяє уникнути суб'єктивності, яка неминуче присутня у ручному описі. Використання ШІ-моделі гарантує консистентність термінології, точність у відображенні часових міток, а також підвищену об'єктивність при класифікації рівня серйозності інциденту. У цьому

випадку, ШІ-класифікація важливості як "висока" співпала з оцінкою аналітика, що підтверджує надійність моделі у прийнятті рішень.

Третьою перевагою є інтеграція з таксономією MITRE ATT&CK, що є важливою складовою сучасного аналізу загроз. Ручний опис часто не включає цей елемент або робить це непослідовно. Автоматизоване включення технік атак у відповідності до ATT&CK підвищує цінність звіту для подальшого аналізу, кореляції з іншими інцидентами, а також звітності перед зовнішніми аудиторами чи вендорами.

Крім того, автоматизований висновок включає детальний опис як технічних аспектів інциденту (IP-адреси, конкретні дії зловмисника, використані інструменти), так і заходів стримування (блокування сесій, скидання MFA, вилучення листів, блокування на CASB), що уніфіковано представлено в зручному для аналізу форматі. У ручному описі ця інформація розкидана, що ускладнює швидке ознайомлення з інцидентом.

Варто також зазначити, що час генерації такого висновку (приблизно 6 секунд) дозволяє оперативно формувати документи, які раніше могли б займати години праці аналітика. У конкретному випадку, розслідування та стримання інциденту зайняло загалом 17 хвилин з моменту першого надходження сповіщення про інцидент, час на підбиття підсумків та категоризацію інциденту склав ще 4 хвилини, таким чином використовуючи розроблене програмне рішення повний час вирішення інциденту міг бути зменшений на більше ніж 20%. Це критично важливо у ситуаціях з високою динамікою атак, коли швидкість реагування має вирішальне значення. До того ж, автоматизований звіт легко масштабувати – він може бути одразу інтегрований у системи документообігу, навчання персоналу чи постінцидентний аналіз без потреби в ручному редагуванні.

Загалом, основна перевага запропонованого висновку полягає у його стандартизованості, повноті, об'єктивності та швидкості формування. Це значно покращує якість розслідувань, підтримує послідовність у звітності, дозволяє

накопичувати порівнювану статистику інцидентів і надає можливість будувати ефективні стратегії превентивного захисту.

Розглянемо інший приклад інциденту інформаційної безпеки, який дозволяє продемонструвати додаткові можливості розробленого програмного модуля. У цьому випадку йдеться про спрацювання системи запобігання вторгненням, яка зафіксувала потенційно небезпечну активність, а саме спробу експлуатації відомої вразливості CVE-2017-0016. Згідно з інформацією, опублікованою Національним інститутом стандартів і технологій США (NIST) [50], дана вразливість дозволяє віддалене виконання коду шляхом надсилання спеціально сформованих мережевих пакетів протоколів SMBv2 або SMBv3. Це створює потенційну загрозу для систем, які не мають відповідних оновлень безпеки. З технічної точки зору, експлуатація CVE-2017-0016 можлива лише у системах, що працюють на конкретних версіях Windows, таких як Windows 10 Gold, 1511, 1607; Windows 8.1; Windows RT 8.1; Windows Server 2012 R2 та Windows Server 2016.

Аналітик, використовуючи інструменти програмного модуля, отримав змогу оперативно оцінити, чи є цільова система дійсно вразливою до зазначеної експлуатації. Завдяки попередньому знанню інфраструктури підприємства, або шляхом інтеграції з внутрішніми системами інвентаризації активів і системами управління вразливостями, було встановлено, що жодна з систем, які перебувають в експлуатації на підприємстві, не використовує вищезазначені застарілі версії операційних систем. Це дозволило аналітику уникнути витрати часу на детальне вивчення потенційно неактуального інциденту та спрямувати зусилля на аналіз інших подій, які могли б становити реальну загрозу в даний момент.

Подальший аналіз перехопленого мережевого пакету також підтвердив відсутність ризику: виявилося, що пакет не містить елементів формату SMB, які були б необхідні для реалізації експлуатації CVE-2017-0016. Таким чином, попри те, що система запобігання вторгненням ініціювала спрацювання, фактична загроза була відсутня. Інформація про цей випадок була передана вендору системи захисту для уточнення та покращення точності сигнатур. У процесі інциденту власник

пристрою, з якого було зафіксовано підозрілу активність, не помітив жодних збоїв у роботі, що додатково підтвердило відсутність негативного впливу.

Підсумковий висновок аналітика, перекладений із англійської мови, підтверджує зазначене: аналітична команда була попереджена про заблокований трафік, який система IPS класифікувала як потенційно небезпечний. Подія була зафіксована 8 квітня 2025 року о 09:55:17 GMT. Було вказано, що задіяні пристрої не є вразливими до CVE-2017-0016, а аналіз мережевих пакетів не виявив ознак активної експлуатації вразливості або іншої шкідливої поведінки. Зазначалося, що це єдина зафіксована спроба, яка могла бути одноразовою дією, спричиненою роботою активного мережевого сканера. Така активність є типовою для автоматизованих засобів моніторингу, які постійно ініціюють численні з’єднання з різними хостами.

Процес обробки цього інциденту був досить швидким. Загальний час реагування склав 14 хвилин, з яких близько двох хвилин було витрачено на підтвердження версій операційних систем на хості за допомогою відповідних засобів управління інфраструктурою, тоді як ще три хвилини знадобилося на складання фінального висновку.

Висновок аналітика у перекладі : «Нас попередили про заблокований трафік IPS, що надходить від вихідного пристрою до хоста призначення.

Час події - вівторок, 8 квітня 2025 р., 09:55:17 GMT

Назва спрацьованого захисту: CVE-2017-0016, і задіяні пристрої не вразливі до нього.

Перевірка пакетів також не дала жодних результатів, які б підтвердили ознаки експлуатації CVE або будь-якої шкідливої активності.

Спостерігалося лише одне запобігання IPS, не варто робити виключення IPS, оскільки це може бути одноразова дія.

Власник вихідного пристрою не знає, чи це запобігання мало якийсь вплив на роботу, оскільки це діяльність сканера, який постійно працює та виконує багато подібних підключень.

Інших запобігань не було, це перший випадок такої активності.»

Надамо програмному модулю інформацію виключно про першочергове спрацювання, та попросимо надати інформацію про вразливість, версію цільового хосту та її вразливість, яка зображена на рисунку 4.3. Після попросимо систему створити висновок згідно повного аналізу який зображена на рисунку 4.4.

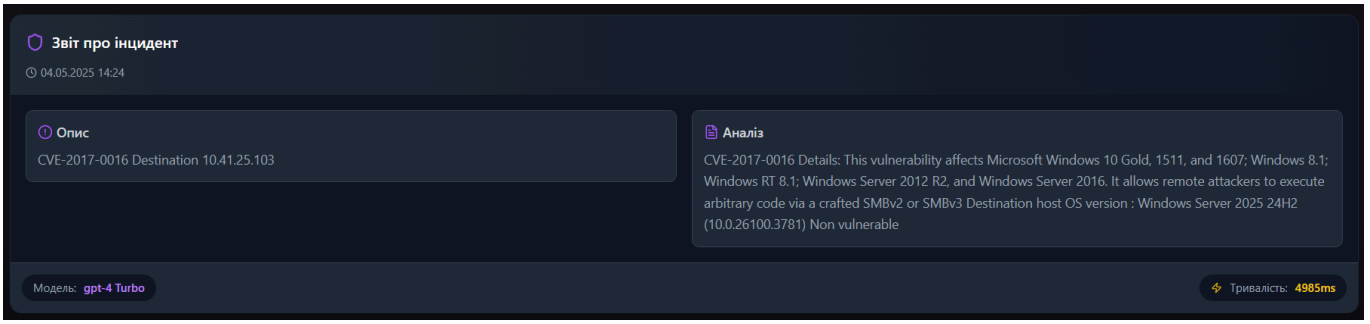


Рисунок 4.3 – Отримана відповідь від програмного модуля щодо виявленої вразливості

Згідно отриманої інформації з програмного модуля, можливо було б за менше 5 секунд визначити неможливість експлуатації даної вразливості. У випадку відсутності інших спрацювань по цільовому хосту важливість даного спрацювання як низьку. Що позитивно вплине на якість роботи та пріоритезацію завдань у випадку високої навантаженості центру реагування на інциденти інформаційної безпеки.

Розглянемо підготовлений висновок щодо інциденту на рисунку та деталі зображені на рисунку 4.4

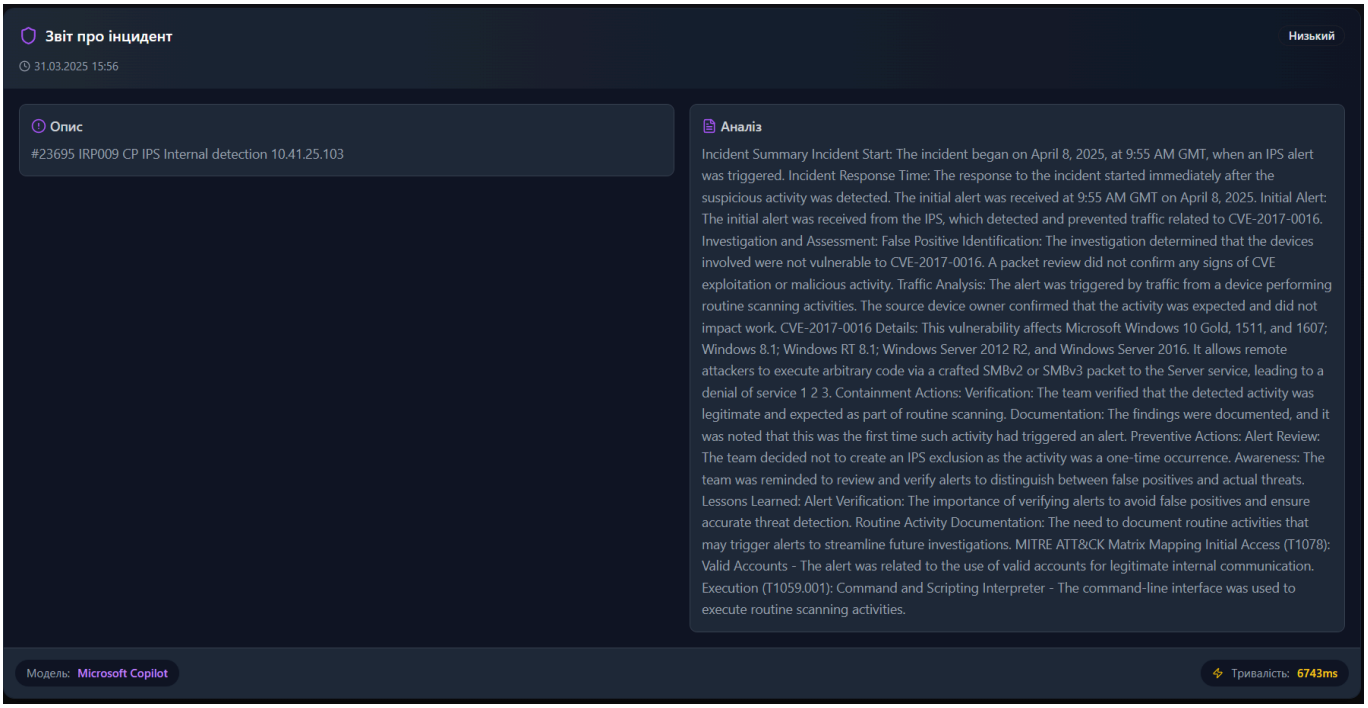


Рисунок 4.4 – Отримана відповідь від програмного модуля щодо підсумків інциденту

У перекладі отримали : «Зведення про інцидент

Початок інциденту: Інцидент розпочався 8 квітня 2025 року о 9:55 ранку за Гринвічем, коли було спрацьовано сповіщення IPS.

Час реагування на інцидент: Реакція на інцидент розпочалася одразу після виявлення підозрілої активності. Початкове сповіщення було отримано о 9:55 ранку за Гринвічем 8 квітня 2025 року.

Початкове сповіщення було отримано від IPS, яка виявила та запобігла трафіку, пов'язаному з CVE-2017-0016. Розслідування та оцінка: Хибнопозитивна ідентифікація: Розслідування визначило, що задіяні пристрої не були вразливими до CVE-2017-0016.

Перевірка пакетів не підтвердила жодних ознак експлуатації CVE або шкідливої активності.

Аналіз трафіку: Сповіщення було спрацьовано трафіком від пристрою, який виконував рутинне сканування.

Власник вихідного пристрою підтвердив, що активність була очікуваною та не вплинула на роботу. Деталі CVE-2017-0016: Ця вразливість впливає на Microsoft Windows 10 Gold, 1511 та 1607; Windows 8.1; Windows RT 8.1; Windows Server 2012 R2 та Windows Server 2016. Це дозволяє віддаленим зловмисникам виконувати довільний код через створений пакет SMBv2 або SMBv3 до служби Server, що призводить до відмови в обслуговуванні.

Дії щодо стримування: Перевірка: Команда підтвердила, що виявлена активність була легітимною та очікуваною в рамках звичайного сканування.

Документація: Результати були задокументовані, і було зазначено, що це був перший випадок, коли така активність викликала сповіщення.

Запобіжні дії: Перевірка сповіщень: Команда вирішила не створювати виключення IPS, оскільки активність була одноразовою.

Обізнаність: Команді нагадали про необхідність переглядати та перевіряти сповіщення, щоб розрізняти хибнопозитивні результати та фактичні загрози.

Винесені уроки: Перевірка сповіщень: Важливість перевірки сповіщень, щоб уникнути хибнопозитивних результатів та забезпечити точне виявлення загроз.

Документування звичайної діяльності: Необхідність документування рутинної діяльності, яка може викликати сповіщення, для оптимізації майбутніх розслідувань.

Матричне зіставлення MITRE ATT&CK

Початковий доступ (T1078): Дійсні облікові записи - Сповіщення було пов'язане з використанням дійсних облікових записів для легітимного внутрішнього зв'язку.

Виконання (T1059.001): Інтерпретатор команд і сценаріїв – інтерфейс командного рядка використовувався для виконання рутинних дій сканування.»

Розширений аналіз наведеної ситуації демонструє, наскільки ефективною може бути автоматизація обробки інцидентів інформаційної безпеки за допомогою штучного інтелекту. У даному випадку програмному модулю було надано лише ключову інформацію про спрацювання системи запобігання вторгнень (IPS), пов’язане з відомою вразливістю CVE-2017-0016. На основі цієї обмеженої

інформації, система протягом менш ніж п’яти секунд змогла ідентифікувати, що зазначене спрацювання наразі не становить реальної загрози для цільової інфраструктури.

Це стало можливим завдяки швидкому аналізу технічних характеристик вразливості, порівнянню їх із версією операційної системи хоста та перевірці, чи підпадає вона під умови можливого експлуатування.

Отриманий у результаті висновок ШІ містить чітку структуровану відповідь, яка включає в себе не лише основні технічні деталі вразливості, а й оцінку релевантності цієї загрози для конкретної ситуації. Оскільки було вказано, що задіяні пристрої не використовують операційні системи, які згідно з базою знань NIST підпадають під дію CVE-2017-0016.

Таким чином розроблений програмний модуль дозволяє уникнути надмірного навантаження на аналітиків під час великої кількості паралельних інцидентів. Замість того, щоб кожного разу вручну перевіряти технічні деталі CVE, версії ОС та проводити трафік-аналіз, аналітик може отримати готовий звіт і швидко прийняти рішення.

По-друге, скорочення часу реагування — у даному випадку до п’яти секунд на початкову перевірку — значно підвищує ефективність роботи Центру реагування на інциденти (SOC). Це критично важливо в умовах обмежених ресурсів, коли пріоритезація загроз відіграє ключову роль у забезпеченні безперервного захисту організації. За умови використання розробленого програмного модуля економія часу на повне вирішення інциденту складал б більше 35%.

Загалом, ефективність цього підходу полягає у здатності системи швидко обробляти навіть неструктуровану або мінімальну вхідну інформацію, проводити технічну верифікацію згідно відкритих баз даних що сприяє прийняттю вірних рішень у найкоротші строки. Це є показником зрілості інтеграції штучного інтелекту в інфраструктуру кіберзахисту.

Середнім показником часу написання звітності щодо інциденту у вибірці є 19% та ще 10% витрачено на пошук і документування інформації про цільові хости

та версії встановленого на них програмного забезпечення. Таким чином використання програмного модуля з інтеграцією у системи менеджменту інцидентів дозволило б скоротити час закриття інциденту в середньому 29%. Цей показник є більшим за аналогічний у рішенні Майкрософт, через відмінність підходів, інтеграцію з іншими наявними системами та відмінним списком задач, що вирішуються.

Висновок до четвертого розділу

У четвертому розділі було виконано ключове завдання дослідження — аналіз результатів використання розробленого методу обробки інцидентів інформаційної безпеки, реалізованого у вигляді програмного модуля з інтеграцією сучасних мовних моделей штучного інтелекту.

На основі реальних кейсів, що охоплювали фішингову атаку та хибнопозитивне спрацювання системи запобігання вторгненням, було продемонстровано ефективність запропонованого рішення. Проведене тестування засвідчило, що програмний модуль:

Забезпечує стандартизовану, структуровану та повну звітність, що відповідає сучасним фреймворкам (NIST, MITRE ATT&CK, ISO/IEC 27035);

Скорочує час реагування та підготовки звітів у середньому на 29%, що є критично важливим у високонавантажених умовах роботи центрів реагування на інциденти (SOC);

Покращує пріоритезацію інцидентів, дозволяючи швидко відсіяти хибнопозитивні спрацювання та зосередитися на дійсно критичних загрозах;

Має можливість інтеграції з внутрішніми системами підприємства, що дозволяє оперативно отримувати інформацію про вразливості, версії ПЗ та інші параметри безпеки.

Таким чином, результати, отримані в цьому розділі, підтверджують високу ефективність і практичну доцільність впровадження розробленого методу в реальні процеси кіберзахисту. Це дозволяє не лише підвищити якість реагування на інциденти, а й створити основу для подальшої автоматизації та вдосконалення систем інформаційної безпеки підприємства.

ВИСНОВОК

Результатом виконання кваліфікаційної роботи є розробка методу обробки інцидентів інформаційної безпеки з використанням штучного інтелекту та його програмна реалізація у вигляді модуля, інтегрованого із системою менеджменту інцидентів. Запропонований підхід дозволяє підвищити ефективність реагування на інциденти, автоматизувати окремі етапи їхнього аналізу та скоротити середній час на усунення загроз.

У першому розділі було проведено аналіз нормативної та законодавчої бази з обробки інцидентів у сфері інформаційної безпеки. Вивчено положення міжнародних стандартів, таких як NIST SP 800-61 та ISO/IEC 27035, а також враховано вимоги українського законодавства, зокрема проєкту Закону №11290. За результатами аналізу для подальшої роботи було обрано модель життєвого циклу інцидентів за стандартом NIST, що дозволяє структурувати процес реагування на інциденти згідно з міжнародною практикою.

У другому розділі досліджено поточні можливості застосування штучного інтелекту в процесі обробки інцидентів. Було проаналізовано рішення провідних вендорів (зокрема Microsoft, CheckPoint та окремо модель OpenAI з задачею виявлення інциденту) та встановлено, що основне застосування ШІ має бути зосереджене на автоматизації аналізу логів, формуванні підсумків та допоміжній аналітиці. Водночас доведено, що на поточному етапі розвитку технологій штучний інтелект ще не є готовим до повноцінного самостійного виявлення інцидентів через високу кількість хибнопозитивних спрацювань.

У третьому розділі розроблено новий метод обробки інцидентів, який інтегрує мовні моделі штучного інтелекту (GPT-4, Claude 3, Copilot тощо) у процес формування підсумку інциденту, пошуку технічної інформації та підтримки процесу Threat Hunting. Метод базується на модульній архітектурі, яка дозволяє здійснювати передачу вхідних даних із системи менеджменту інцидентів за допомогою відкритих API до моделей у хмарному середовищі, а після обробки — повертати результати

назад до системи управління інцидентами для їх збереження та подальшого використання.

У четвертому розділі було проведено аналіз ефективності застосування розробленого модуля на основі експериментального впровадження. Результати свідчать про зменшення середньої тривалості вирішення інцидентів приблизно на 29%, а також зниження навантаження на аналітиків першого рівня. Було встановлено, що запропонований підхід дозволяє швидше формувати технічні довідки, витягати релевантну інформацію з відкритих джерел, а також значно покращити якість підсумкової документації з інциденту.

Враховуючи поставлену мету кваліфікаційної роботи, були виконані всі основні завдання:

Проаналізовано ключові стандарти обробки інцидентів інформаційної безпеки;

Досліджено сучасні рішення на основі штучного інтелекту у сфері кібербезпеки;

Розроблено метод обробки інцидентів із залученням мовних моделей ШІ;

Реалізовано програмний модуль та проаналізовано результати його використання в умовах, наближених до практичних.

Таким чином, робота демонструє перспективність впровадження штучного інтелекту на етапах аналізу та документування інцидентів, забезпечуючи поєднання нормативного підходу та інноваційних технологій для підвищення ефективності інформаційної безпеки в організації.

СПИСОК ДЖЕРЕЛ

1. William S. Cryptography and Network Security: Principles and Practice. Pearson Education, Limited, 2002. 681 с.

2. Wołoszyn J. Integrating Artificial Intelligence in Cybersecurity Detection and Response. Dydaktyka Informatyki. 2024. Т. 19. С. 209–217. URL: https://doi.org/10.15584/di.2024.19.17 (дата звернення: 12.05.2025).

3. Важливий крок для досягнення захисту Держави в кіберпросторі : Ухвалено законопроект 11290. URL: https://cip.gov.ua/ua/news/vazhlivii-krok-dlya-posilennya-zakhistu-derzhavi-v-kiberprostori-ukhvaleno-zakonoproyekt-11290 (дата звернення: 12.05.2025).

4. Олександр Потій : Настав час відходу від КСЗІ. URL: https://www.cip.gov.ua/ua/news/oleksandr-potii-nastav-chas-vidkhodu-vid-kszi (дата звернення: 12.05.2025).

5. Scarfone K. A., Grance T., Masone K. Computer security incident handling guide. Gaithersburg, MD : National Institute of Standards and Technology, 2008. URL: https://doi.org/10.6028/nist.sp.800-61r1 (дата звернення: 12.05.2025).

6. Ahmad A., Maynard S. B., Shanks G. A case analysis of information systems and security incident responses. International Journal of Information Management. 2015. Т. 35, № 6. С. 717–723. URL: https://doi.org/10.1016/j.ijinfomgt.2015.08.001 (дата звернення: 12.05.2025).

7. Toward viable information security reporting systems / F. Olav Sveen та ін. Information Management & Computer Security. 2007. Т. 15, № 5. С. 408–419. URL: https://doi.org/10.1108/09685220710831143 (дата звернення: 12.05.2025).

8. Gonzalez J. J. Towards a Cyber Security Reporting System – A Quality Improvement Process. Lecture Notes in Computer Science. Berlin, Heidelberg, 2005. С. 368–380. URL: https://doi.org/10.1007/11563228_28 (дата звернення: 12.05.2025).

9. Tatu T., Ament C., Jaeger L. Lessons Learned from an Information Security Incident: A Practical Recommendation to Involve Employees in Information Security. Hawaii International Conference on System Sciences. 2018. URL: https://doi.org/10.24251/hicss.2018.471 (дата звернення: 12.05.2025).

10. Bruschi D., Diomede N. A framework for assessing AI ethics with applications to cybersecurity. AI and Ethics. 2022. URL: https://doi.org/10.1007/s43681-022-00162-8 (дата звернення: 14.05.2025).

11. How integration of cyber security management and incident response enables organizational learning / A. Ahmad та ін. Journal of the Association for Information Science and Technology. 2019. Т. 71, № 8. С. 939–953. URL: https://doi.org/10.1002/asi.24311 (дата звернення: 12.05.2025).

12. An Effective Cybersecurity Training Model to Support an Organizational Awareness Program / R. Sabillon та ін. Journal of Cases on Information Technology. 2019. Т. 21, № 3. С. 26–39. URL: https://doi.org/10.4018/jcit.2019070102 (дата звернення: 14.05.2025).

13. Artificial Intelligence and Cybersecurity: Past, Presence, and Future / T. C. Truong та ін. Advances in Intelligent Systems and Computing. Singapore, 2020. С. 351–363. URL: https://doi.org/10.1007/978-981-15-0199-9_30 (дата звернення: 14.05.2025).

14. A framework for incident response management in the petroleum industry / M. G. Jaatun та ін. International Journal of Critical Infrastructure Protection. 2009. Т. 2, № 1-2. С. 26–37. URL: https://doi.org/10.1016/j.ijcip.2009.02.004 (дата звернення: 12.05.2025).

15. Camacho N. G. The Role of AI in Cybersecurity: Addressing Threats in the Digital Age. Journal of Artificial Intelligence General science (JAIGS) ISSN:3006-4023. 2024. Т. 3, № 1. С. 143–154. URL: https://doi.org/10.60087/jaigs.v3i1.75 (дата звернення: 14.05.2025).

16. CHATGPT: A DOUBLE-EDGED SWORD IN CYBERSECURITY - EVALUATING RISKS AND RECOMMENDATIONS FOR SAFER AI INTEGRATION / M. G. Patel та ін. ShodhKosh: Journal of Visual and Performing Arts. 2024. Т. 5, № 5. URL: https://doi.org/10.29121/shodhkosh.v5.i5.2024.1956 (дата звернення: 14.05.2025).

17. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection. Knowledge and Information Systems. 2025. URL: https://doi.org/10.1007/s10115-025-02366-w (дата звернення: 14.05.2025).

18. Gonzalez R. Artificial Intelligence in Cybersecurity. American Journal of Rising Scholar Activities. 2022. Т. 1, № 1. URL: https://doi.org/10.7771/2692-4161.1005 (дата звернення: 14.05.2025).

19. Костюченко В., Табаченко Д., Білоконь І. Зменшення втоми від тривог в SOC: Покращення реагування на інциденти за допомогою автоматизації та штучного інтелекту. VIII Міжнародна науково-практична конференція «Проблеми кібербезпеки інформаційно-комунікаційних систем» (PCSICS). 2025. С. 109-110

20. Michael K., Abbas R., Roussos G. AI in Cybersecurity: The Paradox. IEEE Transactions on Technology and Society. 2023. Т. 4, № 2. С. 104–109. URL: https://doi.org/10.1109/tts.2023.3280109 (дата звернення: 14.05.2025).

21. Threats and Opportunities with AI-based Cyber Security Intrusion Detection: A Review / B. Dash та ін. International Journal of Software Engineering & Applications. 2022. Т. 13, № 5. С. 13–21. URL: https://doi.org/10.5121/ijsea.2022.13502 (дата звернення: 14.05.2025).

22. Punia V., Aggarwal G., Shivam. Impact of Artificial Intelligence (AI) in Cybersecurity. Recent Advances in Computational Intelligence and Cyber Security. London, 2024. С. 183–193. URL: https://doi.org/10.1201/9781003518587-18 (дата звернення: 14.05.2025).

23. Puthal D., Mohanty S. Cybersecurity Issues in AI. IEEE Consumer Electronics Magazine. 2021. Т. 10, № 4. С. 33–35. URL: https://doi.org/10.1109/mce.2021.3066828 (дата звернення: 14.05.2025).

24. Qumer S. M., Ikrama S. Poppy Gustafsson: redefining cybersecurity through AI. The Case For Women. 2022. С. 1–38. URL: https://doi.org/10.1108/cfw.2022.000001 (дата звернення: 14.05.2025).

25. Sarker I. H., Furhad M. H., Nowrozy R. AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. SN Computer Science. 2021. Т. 2, № 3. URL: https://doi.org/10.1007/s42979-021-00557-0 (дата звернення: 14.05.2025).

26. Sharma P., Dash B., Ansari M. F. Anti-Phishing Techniques – A Review of Cyber Defense Mechanisms. IJARCCE. 2022. Т. 11, № 7. URL: https://doi.org/10.17148/ijarcce.2022.11728 (дата звернення: 14.05.2025).

27. Markevych M., Dawson M. A Review of Enhancing Intrusion Detection Systems for Cybersecurity Using Artificial Intelligence (AI). International conference KNOWLEDGE-BASED ORGANIZATION. 2023. Т. 29, № 3. С. 30–37. URL: https://doi.org/10.2478/kbo-2023-0072 (дата звернення: 14.05.2025).

28. Mohamed N. Artificial Intelligence in Cybersecurity: A Review of Solutions for APT-Exploited Vulnerabilities. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), м. Kamand, India, 24–28 черв. 2024 р. 2024. С. 1–7. URL: https://doi.org/10.1109/icccnt61001.2024.10724084 (дата звернення: 14.05.2025).

29. Bishop M. Computer Security: Art and Science. Addison-Wesley Professional, 2002. 1200 с.

30. Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron. Deep Learning. Das umfassende Handbuch: Grundlagen, aktuelle Verfahren und Algorithmen, neue Forschungsansätze. MITP Verlags GmbH, 2018.

31. Martin J. H., Jurafsky D. Speech and Language Processing (2nd Edition). 2-ге вид. Prentice Hall, 2006. 944 с.

32. Mitnick K. D., Simon W. L. Art of Deception: Controlling the Human Element of Security. Wiley & Sons, Incorporated, John, 2001.

33. Mogotsi I. C. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze: Introduction to information retrieval. Information Retrieval. 2009. Т. 13, № 2. С. 192–195. URL: https://doi.org/10.1007/s10791-009-9115-y (дата звернення: 12.05.2025).

34. Patterson C. M., Nurse J. R. C., Franqueira V. N. L. Learning from cyber security incidents: A systematic review and future research agenda. Computers & Security. 2023. С. 103309. URL: https://doi.org/10.1016/j.cose.2023.103309 (дата звернення: 12.05.2025).

35. Generative AI and Security Operations Center Productivity: Evidence from Live Operations. URL: https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Generative-AI-and-Security-Operations-Center-Productivity-Evidence-from-Live-Operations_v2.5-FINAL.pdf (дата звернення: 12.05.2025).

36. Learn more about ATT&CK training. URL: https://attack.mitre.org/resources/learn-more-about-attack/training/cti/ (дата звернення: 12.05.2025).

37. The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review / M. F. Ansari та ін. IJARCCE. 2022. Т. 11, № 9. URL: https://doi.org/10.17148/ijarcce.2022.11912 (дата звернення: 14.05.2025)

37. DeepSeek AI database exposed over 1 billion records. URL: https://thehackernews.com/2025/01/deepseek-ai-database-exposed-over-1.html?_m=3n.009a.3579.bi0ao45cy8.2lle (дата звернення: 12.05.2025).

38. Which countries have banned DeepSeek and why? Al Jazeera. URL: https://www.aljazeera.com/news/2025/2/6/which-countries-have-banned-deepseek-and-why (дата звернення: 12.05.2025).

39. Performance. URL: https://lomereiter.github.io/2015/03/29/performance.html (дата звернення: 12.05.2025).

40. Nguyen, Nguyen. "Creating a modern web user interface using react and typescript." (2022). URL: https://www.theseus.fi/bitstream/handle/10024/745669/Thesis.pdf (дата звернення: 12.05.2025).

41 Gerchev, Ivaylo. Tailwind CSS. SitePoint Pty Ltd, 2022 С. 33–35. URL: https://books.google.com.ua/books?hl=en&lr=&id=GczDEAAAQBAJ&oi=fnd&pg=PT3&dq=Tailwind&ots=hTDe9QRFEh&sig=rDfVBTg6xJ9QYTD-KLi3bOvPjJA&redir_esc=y#v=onepage&q=Tailwind&f=false

42. Vasyliev, D. O., and M. Suknov. "Comparative analysis of the content rendering techniques for web application development." (2023) С. 174-175. URL: https://dspace.khadi.kharkov.ua/bitstreams/f04ff3b7-b262-4c7b-804c-a5a01c7d3b4f (дата звернення: 12.05.2025).

43. Murthy M. P. V. S. N. Network Forensics and Incident Response Tool with AI-Assisted Threat Analysis. International Journal for Research in Applied Science and Engineering Technology. 2025. Т. 13, № 3. С. 1002–1008. URL: https://doi.org/10.22214/ijraset.2025.67471 (дата звернення: 14.05.2025).

44. Thompson E. C. Continuous Monitoring of Incident Response Program. Cybersecurity Incident Response. Berkeley, CA, 2018. С. 125–135. URL: https://doi.org/10.1007/978-1-4842-3870-7_10 (дата звернення: 14.05.2025).

45. Nunnaguppala, Laxmi Sarat Chandra. "Leveraging AI In Cloud SIEM And SOAR: Real-World Applications For Enhancing SOC And IRT Effectiveness." International Journal for Innovative Engineering and Management Research 10.08 (2021): С. 376-393. DOI: 10.48047/IJIEMR/V10/ISSUE 08/61 (дата звернення: 14.05.2025).

46. Shtok, Anna, et al. "Learning from the past: answering new questions with past answers." Proceedings of the 21st international conference on World Wide Web. 2012 URL:https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=17ba9019ec7eb1f03d64f4d80e47376a4f6f8583 (дата звернення: 14.05.2025)..

47. Shen, Lei. "The NIST cybersecurity framework: Overview and potential impacts." Scitech Lawyer 10.4 (2014): C. 16. URL: https://doi.org/10.1007/978-3-030-03638-6_23 (дата звернення: 14.05.2025).

48. Kure, Halima Ibrahim, et al. "Asset criticality and risk prediction for an effective cybersecurity risk management of cyber-physical system." Neural Computing and Applications 34.1 (2022): C. 493-514. https://doi.org/10.1007/s00521-021-06400-0 (дата звернення: 12.05.2025).

49. Computer Security Incident Handling Guide (NIST SP 800-61 Rev. 2). URL: https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-61r2.pdf (дата звернення: 12.05.2025).

50. CVE-2017-0016. URL: https://nvd.nist.gov/vuln/detail/CVE-2017-0016 (дата звернення: 12.05.2025).

ДОДАТОК А

ДОДАТОК Б

ДОДАТОК В
СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Тези наукових конференцій

Костюченко В., Табаченко Д., Білоконь І. Зменшення втоми від тривог в SOC: Покращення реагування на інциденти за допомогою автоматизації та штучного інтелекту. VIII Міжнародна науково-практична конференція «Проблеми кібербезпеки інформаційно-комунікаційних систем» (PCSICS). 2025. С. 109-110