МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ КІБЕРБЕЗПЕКИ ТА ЗАХИСТУ ІНФОРМАЦІЇ
КАФЕДРА УПРАВЛІННЯ КІБЕРБЕЗПЕКОЮ ТА ЗАХИСТОМ ІНФОРМАЦІЇ

КВАЛІФІКАЦІЙНА РОБОТА

на тему: “**ЗАСОБИ ШТУЧНОГО ІНТЕЛЕКТУ В ОРГАНІЗАЦІЇ ТА ЗДІЙСНЕННІ КІБЕРАТАК**”

на здобуття освітнього ступеня бакалавра
зі спеціальності 125 Кібербезпека
освітньої програми Управління інформаційною та кібернетичною безпекою

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Богдан ЯРМОЛЕНКО
(підпис) *Ім’я, ПРІЗВИЩЕ здобувача*

Виконала: здобувач вищої освіти гр. УБД-42

Богдан ЯРМОЛЕНКО
Ім’я, ПРІЗВИЩЕ

Керівник: Тетяна МУЖАНОВА
к. держ. упр., доцент Ім’я, ПРІЗВИЩЕ

Рецензент:
Ім’я, ПРІЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут кібербезпеки та захисту інформації

Кафедра управління кібербезпекою та захистом інформації

Ступінь вищої освіти бакалавр

Спеціальність 125 Кібербезпека

Освітня програма Управління інформаційною та кібернетичною безпекою

ЗАТВЕРДЖУЮ
Завідувач кафедри УКБЗІ
_____ Світлана ЛЕГОМІНОВА
“____”_____2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Ярмоленку Богдану Віталійовичу
(прізвище, ім’я, по батькові здобувача)

1. Тема кваліфікаційної роботи “Засоби штучного інтелекту в організації та здійсненні кібератак”, керівник кваліфікаційної роботи МУЖАНОВА Тетяна, к. держ. упр., доцент,
(ІПРІЗВИЩЕ, Ім’я,, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій від “24“ лютого 2025 р. №56.

2. Строк подання кваліфікаційної роботи “20” травня 2025р.

3. Вихідні дані до кваліфікаційної роботи: *штучний інтелект (ШІ), ШІ в організації кібератак, протидія атакам із застосуванням ШІ.*

4. Перелік питань, які мають бути розроблені:
4.1. Дослідити теоретичні основи застосування штучного інтелекту у сфері кібербезпеки.
4.2. Проаналізувати особливості застосування ШІ у зловмисній кібердіяльності.
4.3. Вивчити засади протидії загрозам із використанням штучного інтелекту.

5. Перелік ілюстративного матеріалу: *презентація PowerPoint*

6. Дата видачі завдання “11“ березня 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Етапи кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
1.	Визначення об’єкту, предмету, мети та завдань дослідження.	18.03.2025	
2.	Збір та аналіз літератури.	29.03.2025	
3.	Дослідження теоретичних основ застосування штучного інтелекту у сфері кібербезпеки.	08.04.2025	
4.	Аналіз особливостей застосування ШІ у зловмисній кібердіяльності.	22.04.2025	
5.	Вивчення засад протидії загрозам із використанням штучного інтелекту.	08.05.2025	
6.	Формулювання висновків за результатами проведеного дослідження.	20.05.2025	
7.	Оформлення роботи.	22.05.2025	
8.	Оформлення презентації.	29.05.2025	
9.	Отримання рецензії на роботу.	03.06.2025	
10.	Захист в ЕК.	__.06.2025	

Здобувач вищої освіти _____ Богдан ЯРМОЛЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

Керівник
кваліфікаційної роботи _____ Тетяна МУЖАНОВА
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ КІБЕРБЕЗПЕКИ ТА ЗАХИСТУ ІНФОРМАЦІЇ

ПОДАННЯ ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня бакалавра

Направляється здобувач Ярмоленко Б.В. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 125 Кібербезпека
(код, найменування спеціальності)
освітньої програми Управління інформаційною та кібернетичною безпекою
(назва)
на тему: “Засоби штучного інтелекту в організації та здійсненні кібератак”

Кваліфікаційна робота і рецензія додаються.

Директор ННІКБЗІ _____ Євгенія ІВАНЧЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

Висновок керівника кваліфікаційної роботи

Здобувач ЯРМОЛЕНКО Богдан у кваліфікаційній роботі дослідив теоретичні основи застосування штучного інтелекту у сфері кібербезпеки; проаналізував особливості застосування ШІ у зловмисній кібердіяльності; вивчив засади протидії загрозам із використанням штучного інтелекту.

Під час підготовки кваліфікаційної роботи ЯРМОЛЕНКО Богдан показав хорошу теоретичну та практичну підготовку, довів володіння науково-дослідницькими методами, вміння самостійно знаходити шляхи вирішення наукових проблем, проявив себе як організований, відповідальний виконавець. Результати дослідження апробовані на Всеукраїнській науково-практичній конференції “Стратегії кіберстійкості: управління ризиками та безперервність бізнесу” 27 лютого 2025 року.

Все це дозволяє оцінити кваліфікаційну роботу здобувача ЯРМОЛЕНКА Богдана на позитивну оцінку та присвоїти йому кваліфікацію бакалавра з кібербезпеки за освітньою програмою Управління інформаційною та кібернетичною безпекою.

Керівник кваліфікаційної роботи _____ Тетяна МУЖАНОВА
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

“____“_____2025 року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач Ярмоленко Б.В. допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедри
управління кібербезпекою та
захистом інформації _____ Світлана ЛЕГОМІНОВА
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

ВІДГУК РЕЦЕНЗЕНТА

на кваліфікаційну бакалаврську роботу

здобувача вищої освіти ЯРМОЛЕНКА Богдана
на тему “Засоби штучного інтелекту в організації та здійсненні кібератак”

Актуальність. У сучасних умовах кіберзагрози для підприємств і організацій стають дедалі масштабнішими та складнішими. Однією з основних причин такої ситуації є все ширше застосування для організації та проведення кібератак інноваційних технологій, зокрема штучного інтелекту. З огляду на просунуті можливості щодо аналізу великих даних, маскування зловмисних дій і автоматизації рутинних процесів, ШІ дозволяє проводити масштабні і складні для виявлення атаки, що значно збільшує навантаження на фахівців з кібербезпеки і вимагає постійної адаптації діючих підходів кіберзахисту.

Враховуючи ці аспекти, дослідження ролі засобів штучного інтелекту в організації та здійсненні кібератак є актуальним науковим завданням.

Позитивні сторони.

1. У роботі досліджено теоретичні основи застосування штучного інтелекту у сфері кібербезпеки; проаналізовано особливості застосування ШІ у зловмисній кібердіяльності; вивчено засади протидії загрозам із використанням штучного інтелекту.

2. Кваліфікаційна робота оформлена відповідно до вимог. Матеріал дослідження викладено відповідно до плану, зроблено логічні висновки. Ключові положення роботи представлено у вигляді рисунків і таблиць.

3. Здобувач опрацював достатню джерельну базу: близько 60 публікацій, в тому числі англомовні статті й наукові праці вітчизняних учених.

4. Робота має практичну спрямованість, містить аналіз особливостей і методів застосування ШІ для організації кібератак, а також для їх запобігання і протидії.

Недоліки.

Доцільно було б глибше проаналізувати передовий досвід нормативного регулювання проблем застосування ШІ Європейського Союзу та США.

Однак, вищезгадані зауваження не впливають на загальну позитивну оцінку кваліфікаційної роботи.

Висновок: Кваліфікаційна робота виконана на достатньому науково-методичному рівні і заслуговує позитивної оцінки, а здобувач ЯРМОЛЕНКО Богдан заслуговує присвоєння кваліфікації бакалавра з кібербезпеки за освітньою програмою Управління інформаційною та кібернетичною безпекою.

Рецензент: _____ _____

підпис Ім’я, ПРІЗВИЩЕ

РЕФЕРАТ

Кваліфікаційна робота присвячена дослідженню ролі засобів штучного інтелекту в організації та здійсненні кібератак. Робота складається зі вступу, трьох розділів, що містять 14 рисунків і 3 таблиці, висновків і списку використаних джерел із 58 найменувань. Загальний обсяг роботи становить 75 аркушів, з яких 5 аркушів займає список використаних джерел.

Метою роботи є дослідження ролі засобів штучного інтелекту в організації та здійсненні кібератак.

Об'єктом дослідження є засади організації та здійснення кібератак.

Предмет дослідження – роль засобів штучного інтелекту в організації та здійсненні кібератак.

Методи дослідження. Для вирішення поставленого наукового завдання використано методи аналізу та синтезу, порівняння, системного підходу, класифікації і моделювання.

Як результат у роботі досліджено теоретичні основи застосування штучного інтелекту у сфері кібербезпеки; проаналізовано особливості застосування ШІ у зловмисній кібердіяльності; вивчено засади протидії загрозам із використанням штучного інтелекту.

Галузь застосування. Описані підходи можуть бути використані для підвищення рівня запобігання і протидії кіберзагрозам із використанням штучного інтелекту на підприємствах та в організаціях України.

Ключові слова: ШТУЧНИЙ ІНТЕЛЕКТ (ШІ), ШІ В ОРГАНІЗАЦІЇ КІБЕРАТАК, ПРОТИДІЯ АТАКАМ ІЗ ЗАСТОСУВАННЯМ ШІ.

ABSTRACT

The qualification work is devoted to the study of the role of artificial intelligence tools in the organization and implementation of cyberattacks. The work consists of an introduction, three chapters containing 14 figures and 3 tables, conclusions and a list of references which consists of 58 items. The total volume of the work is 75 pages, of which 5 pages are occupied by the list of references.

The purpose of the study is to investigate the role of artificial intelligence tools in the organization and implementation of cyberattacks.

The object of the study is the principles of organizing and implementing cyberattacks.

The subject of the study is the role of artificial intelligence tools in organizing and implementing cyberattacks.

Research methods. To solve the scientific problem, the methods of analysis and synthesis, comparison, system approach, classification and modeling were used.

As a result, the theoretical foundation of the application of artificial intelligence in cybersecurity were studied in the work; the features of the use of AI in malicious cyberactivity were analyzed; the principles of countering threats using artificial intelligence have been studied.

Field of application. The described approaches can be used to increase the level of prevention and countering cyber threats using artificial intelligence at enterprises and organizations in Ukraine.

Keywords: ARTIFICIAL INTELLIGENCE (AI), AI IN ORGANIZING AND IMPLEMENTING CYBERATTACKS, COUNTERING ATTACKS USING AI.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ І СКОРОЧЕНЬ 9
ВСТУП 10
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У СФЕРІ КІБЕРБЕЗПЕКИ 12
1.1 Поняття й основні риси штучного інтелекту 12
1.2 Основні підходи до класифікації систем ШІ 16
1.3 Технології ШІ, що використовуються у сфері кібербезпеки 21
1.4 Особливості і види сучасних загроз кібербезпеці 24
1.5 Вплив ШІ на підходи до забезпечення кібербезпеки 28
Висновки до розділу 1 32
РОЗДІЛ 2. ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ШІ У ЗЛОВМИСНІЙ КІБЕРДІЯЛЬНОСТІ 34
2.1 Стратегії кібератак з використанням алгоритмів ШІ 34
2.2 Автоматизація шкідливих процесів: фішинг, спам, розсилки 37
2.3 Генеративні моделі (GAN, LLM) як інструменти атак 40
2.4 Застосування глибокого навчання в атаках на системи розпізнавання 43
2.5 Атаки на основі обхідних алгоритмів: маніпуляція даними і системами 45
2.6 Приклади реальних кейсів кібератак із застосуванням ШІ 48
Висновки до розділу 2 50
РОЗДІЛ 3. ПРОТИДІЯ ЗАГРОЗАМ ІЗ ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ 52
3.1 Огляд сучасних систем захисту, що використовують ШІ 52
3.2 Використання ШІ у виявленні аномальної активності й кіберзагроз 57
3.3 Інтелектуальні системи моніторингу і реагування на інциденти 58
3.4 Роль машинного навчання у побудові превентивного кіберзахисту 61
3.5 Проблеми етичності та ризики використання ШІ у кібербезпеці 63
3.6 Рекомендації щодо протидії кібератакам із ШІ на рівні підприємства та держави 66
Висновки до розділу 3 68
ВИСНОВКИ 69
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 71

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ І СКОРОЧЕНЬ

МН	Машинне навчання	Machine Learning
ШІ	Штучний інтелект	Artificial Intelligence
APT	Advanced Persistent Threats	Ускладнені постійні атаки
ANN	Artificial Neural Network	Штучна нейронна мережа
GAN	Generative Adversarial Network	Генеративна змагальна мережа
DL	Deep Learning	Глибоке навчання
IDS	Intrusion Detection System	Система виявлення вторгнень
IPS	Intrusion Prevention System	Система запобігання вторгненням
LLM	Large Language Model	Велика мовна модель
ML	Machine Learning	Машинне навчання
NLP	Natural Language Processing	Обробка природної мови
SIEM	Security Information and Event Management	Управління інформацією та подіями безпеки
SOC	Security Operations Center	Центр операцій безпеки
SOAR	Security Orchestration, Automation and Response	Оркестрація, автоматизація і реагування на інциденти безпеки

ВСТУП

Актуальність теми. Сучасні цифрові технології, які стали фундаментом успішного функціонування економіки, державного управління, оборонної сфери і повсякденного життя громадян, нерідко стають інструментом реалізації складних і добре організованих кіберзагроз.

Сьогодні особливо стрімко зростає роль штучного інтелекту, який широко використовується як у забезпеченні кіберзахисту, так і в здійсненні зловмисної кібердіяльності. Завдяки своїй здатності аналізувати великі масиви даних, адаптувати поведінку атак, маскувати шкідливу активність та автоматизувати процеси розробки зловмисного програмного забезпечення, ШІ значно ускладнює роботу команд кібербезпеки і вимагає постійного вдосконалення діючих підходів кіберзахисту відповідно до розвитку можливостей ШІ.

З огляду на зазначене, дослідження ролі засобів штучного інтелекту в організації та здійсненні кібератак є актуальним науковим завданням.

Мета роботи полягає у дослідженні ролі засобів штучного інтелекту в організації та здійсненні кібератак.

Об'єкт дослідження – засади організації та здійснення кібератак.

Предмет дослідження – роль засобів штучного інтелекту в організації та здійсненні кібератак.

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

1. Дослідити теоретичні основи застосування штучного інтелекту у сфері кібербезпеки.

2. Проаналізувати особливості застосування ШІ у зловмисній кібердіяльності.

3. Вивчити засади протидії загрозам із використанням штучного інтелекту.

Методи дослідження. Для вирішення означеного вище наукового завдання в роботі використано методи аналізу та синтезу, порівняння, системного підходу, класифікації і моделювання.

Практичне значення одержаних результатів. Результати дослідження сприятимуть обґрунтованому вибору і впровадженню ефективних методів та

інструментів запобігання і протидії кібератакам із застосуванням ШІ, а також при підготовці методичних матеріалів для проведення тренінгів із даної тематики.

Апробація результатів кваліфікаційної роботи відбулася на Всеукраїнській науково-практичній конференції “Стратегії кіберстійкості: управління ризиками та безперервність бізнесу” 27 лютого 2025 року.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У СФЕРІ КІБЕРБЕЗПЕКИ

1.1 Поняття й основні риси штучного інтелекту

Штучний інтелект (ШІ) є однією з найперспективніших технологій сучасності, яка змінює не лише спосіб обробки інформації, але й концепцію взаємодії людини з цифровим середовищем. Його суть полягає у здатності технічних систем імітувати елементи людського інтелекту, зокрема навчання, планування, розуміння мови, аналіз ситуацій, прийняття рішень. У контексті кібербезпеки ШІ виступає як інструмент подвійного призначення: з одного боку, він підсилює захисні механізми інформаційних систем, а з іншого - надає нові можливості зловмисникам для обходу існуючих бар'єрів.

Термін «Artificial Intelligence» був уперше використаний Джоном Маккарті в 1956 році під час Дартмутської конференції, яка фактично започаткувала наукову галузь, орієнтовану на створення машин, здатних думати. Із того часу ШІ пройшов кілька хвиль інтенсивного розвитку, періоди розчарування та відродження. Сучасний етап його еволюції пов'язаний з досягненнями в області обчислювальної потужності, доступності великих обсягів даних (Big Data) і вдосконаленням алгоритмів машинного навчання.

Наукова спільнота пропонує кілька підходів до класифікації штучного інтелекту. Поширеним є поділ за рівнем когнітивної складності: на слабкий, сильний та суперінтелект. Слабкий ШІ (Narrow AI) виконує конкретно визначені функції, як-от розпізнавання зображень, переклад мови або фільтрація електронної пошти. Він уже є реальністю та активно застосовується в багатьох сферах, зокрема в безпеці. Сильний ШІ (General AI), який міг би мислити та навчатися на рівні людини, наразі залишається гіпотетичним. Ще більш складною є концепція суперінтелекту (Superintelligence) - надлюдського розуму, що теоретично здатен вийти з-під контролю [1].

Технології ШІ класифікуються також за функціональними характеристиками. Найважливішими серед них є машинне навчання, МН (Machine Learning, ML), глибоке навчання (Deep Learning, DL), обробка природної мови ((Natural Language Processing, NLP), експертні системи, нейронні мережі, нечітка логіка (fuzzy logic) та генеративні моделі.

МН - це основна технологія, яка дозволяє системам знаходити закономірності в даних і вдосконалювати свої алгоритми на основі досвіду. Його підвиди включають наглядове, без нагляду та навчання з підкріпленням. Глибоке навчання ґрунтується на використанні багатошарових штучних нейронних мереж, здатних виявляти складні зв'язки в неструктурованих даних.

Окрім базової класифікації ШІ за функціональними і когнітивними характеристиками, доцільно також розглядати його в контексті міждисциплінарного впливу. Сучасні дослідження демонструють, що ШІ поступово втрачає статус «вузької технології» і переходить у площину інфраструктурного явища, яке проникає в економіку, медицину, право, освіту, державне управління. У межах кібербезпеки це означає, що захист інформації більше не може розглядатися лише як сукупність технічних заходів, він повинен вхоплючати соціальні, правові та когнітивні аспекти, які враховують поведінку користувача, мотивацію атакуючих, реакцію суспільства на загрози.

Історично склалося, що дослідження ШІ здійснювалися в кількох наукових школах. У США домінує прикладна модель розвитку ШІ, орієнтована на ефективність і швидку інтеграцію в бізнес-процеси. Європейські науковці часто підходять до проблеми з філософсько-гуманітарної позиції, акцентуючи на етичності, правових межах, прозорості алгоритмів. У Китаї активно розвиваються національні ініціативи створення «цифрового громадянина», де ШІ інтегрується з біометрією, фінансами, системами контролю [2].

Не менш важливою є порівняльна характеристика ШІ із суміжними технологіями, наприклад, із традиційними експертними системами. Якщо останні базуються на чітких логічних правилах, то сучасні системи ШІ працюють на основі імовірнісних оцінок, самонавчання, адаптації. Це відкриває нові

горизонти, але водночас ускладнює контроль над системами, зокрема в безпековому контексті, де ціна помилки може бути надто високою.

Особливу увагу в останні роки привертають генеративні моделі - інструменти, що можуть створювати нову інформацію, яка має схожі характеристики з навчальними даними. Це, зокрема, генеративні змагальні мережі (Generative Adversarial Network, GAN), автокодувальники, а також великі мовні моделі (Large Language Model, LLM), серед яких GPT, Claude, Gemini. Вони дозволяють генерувати текст, зображення, відео, програмний код, що може використовуватись як у творчих, так і в шкідливих цілях.

Кожна з вищезазначених технологій уже сьогодні активно впроваджується у галузь кібербезпеки. Машинне навчання використовується в системах виявлення вторгнень (Intrusion Detection System, IDS), захисту від шкідливого ПЗ, виявлення фішингу та аномалій у поведінці користувачів. Глибокі нейронні мережі допомагають класифікувати трафік, прогнозувати вектори атак, виявляти складні схеми введення в оману. NLP використовується для аналізу повідомлень е-пошти, чатів технічної підтримки, соціальних мереж на предмет потенційних загроз.

Особливо актуальними є застосування генеративних моделей у зловмисній діяльності. Наприклад, GAN можуть створювати фальшиві документи, діпфейки або підроблені сторінки авторизації. LLM використовуються для створення переконливих фішингових повідомлень, автоматичного спілкування з жертвами або генерації соціально-інженерних сценаріїв атак [3].

Усе це формує нову парадигму в розвитку загроз - кібератаки, керовані ШІ, які не тільки маскуються краще, ніж традиційні, але й адаптуються в режимі реального часу.

Наукова література та сучасні аналітичні звіти провідних IT-компаній (Microsoft, IBM, CheckPoint, OpenAI підкреслюють двозначну роль ШІ: це водночас інструмент захисту й джерело нових ризиків. Саме ця подвійність робить критично важливим глибоке розуміння принципів класифікації, функціонування та застосування ШІ в безпековому контексті.

Широке впровадження ШІ актуалізує не лише технічні, а й етичні та правові виклики. Зокрема, піднімається питання прозорості рішень, прийнятих системами ШІ, відповідальності за наслідки їхніх дій, а також недопущення упередженості та дискримінації. Так звані «чорні скриньки» (black box models) - моделі, логіку дій яких важко пояснити людині, викликають занепокоєння в контексті безпеки, адже дії таких систем можуть бути непередбачуваними.

У сфері кібербезпеки також набуває особливої ваги те, що ШІ може блокувати дії користувача, аналізувати поведінку або навіть обмежувати доступ до критичних систем, що вимагає чіткої регламентації меж його повноважень.

На рівні держав і міжнародних організацій ведеться активна робота зі створення нормативної бази. У Європейському Союзі одним із ключових документів став Artificial Intelligence Act - проект регламенту щодо гармонізації правил у сфері штучного інтелекту [9]. Він передбачає класифікацію систем ШІ за рівнем ризику: від мінімального (ігрові алгоритми) до неприпустимого (соціальне оцінювання громадян, як у Китаї). У США акцент робиться на інноваційності, проте діють рекомендації NIST щодо прозорості, безпеки та надійності ШІ.

Україна також зробила перші кроки в цьому напрямку: в урядових стратегіях цифрової трансформації закладено принципи етичного використання ШІ. У 2021 році Мінцифри затвердило Концепцію розвитку ШІ в Україні, яка передбачає створення законодавства з урахуванням європейських підходів, що особливо важливо в контексті наближення до стандартів ЄС [4].

Окрім нормативного регулювання, важливим напрямом є стандартизація ШІ. Наприклад, ISO/IEC JTC 1/SC 42 (Artificial Intelligence) займається формалізацією понять, процесів розробки та тестування інтелектуальних систем, включаючи моделі ризиків у сфері кібербезпеки [9]. Усе це підтверджує, що застосування ШІ неможливе без морального, правового і соціального осмислення, особливо коли йдеться про його роль у захисті (або підриві) критично важливих інформаційних систем.

Таким чином, ШІ є не просто набором алгоритмів, а новим типом технологічного мислення, що трансформує підходи до аналізу ризиків, захисту даних та організації протидії загрозам. Поглиблене вивчення його технологій є першочерговим завданням для фахівців з кібербезпеки в умовах стрімкої діджиталізації та загострення гібридних конфліктів.

1.2 Основні підходи до класифікації систем ШІ

Питання класифікації систем штучного інтелекту є одним із базових у формуванні теоретичних засад ШІ. Відповідна класифікація дозволяє систематизувати різноманітні підходи до побудови, навчання та застосування інтелектуальних систем, визначити їхню ефективність, гнучкість, потенційні ризики та обмеження. У контексті кібербезпеки класифікація допомагає не лише зрозуміти, які системи можуть бути застосовані для захисту, а й виявити, які типи ШІ найчастіше використовуються у зловмисних цілях.

Найпоширенішим є підхід до класифікації за рівнем інтелектуальної складності, відповідно до якого системи ШІ умовно поділяють на три типи (Рис.1.1):

Рис. 1.1. Класифікація AI за рівнем когнітивної складності

Слабкий штучний інтелект (Weak AI / Narrow AI). Системи слабкого ШІ орієнтовані на вирішення конкретно визначених завдань і не мають самосвідомості або здатності до універсального міркування. Прикладами слабкого ШІ є системи рекомендацій, голосові помічники, інтелектуальні чат-боти, пошукові алгоритми, фільтри спаму тощо. Слабкий ШІ вже широко

застосовується у сфері інформаційної безпеки: для аналізу логів, фільтрації фішингових листів, моніторингу активності користувачів тощо.

Серед характерних рис слабкого ШІ:

- обмеженість домену застосування;
- висока точність при повторюваних завданнях;
- нездатність до генералізації знань.

Сильний штучний інтелект (Strong AI / General AI). Це гіпотетичний тип систем, здатний розуміти, навчатися та приймати рішення в будь-якій сфері, так само як людина. Такий ШІ має бути самосвідомим, адаптивним, здатним до самоаналізу та еволюції власної поведінки. На сьогодні сильний ШІ залишається науковою концепцією, хоча провідні дослідницькі організації (DeepMind, OpenAI, Anthropic) активно досліджують потенційну архітектуру таких систем.

Сильний ШІ теоретично може не лише допомагати в захисті інформації, а й створювати повністю автономні системи безпеки. Однак, через відсутність механізмів повного контролю, існує ризик некоректних рішень, що виходять за межі моральних або правових норм.

Суперінтелект (Superintelligence). Поняття суперінтелекту тісно пов’язане з футурологічними сценаріями. Йдеться про створення машинного інтелекту, що перевершує найрозвинутіший людський розум у всіх аспектах. Такий інтелект буде не лише швидшим або точнішим, але і глибшим у плані стратегічного мислення, етичного аналізу, самонавчання. Автор ідеї суперінтелекту Нік Бостром вважає, що саме цей тип ШІ може становити найбільшу загрозу для людства, якщо не буде вчасно врегульовано його поведінку [5].

Хоча суперінтелект не є предметом поточних практичних реалізацій, дослідження у цій галузі впливають на формування етичних норм і концепцій «безпечного ШІ» (AI safety).

Окрім рівня інтелекту, важливою є класифікація ШІ за функціональним призначенням, яка дозволяє зрозуміти, як саме інтелектуальні системи

застосовуються в інформаційній безпеці. Найважливішими серед них є машинне навчання (ML), глибоке навчання (DL), обробка природної мови (NLP) (Рис. 1.2).

Рис. 1.2. Основні види систем ШІ за функціональними характеристиками

Відповідно до іншого підходу системи AI поділяють на такі функціональні типи (Рис. 1.3):

Рис. 1.3. Функціональні типи систем ШІ

1. Аналітичні системи - моделі, що виконують завдання обробки та інтерпретації даних, виявлення аномалій, кореляції подій. Використовуються в SIEM-системах, у розвідці загроз (Threat Intelligence), аналітиці логів.

2. Прогнозуючі системи - ШІ, що будують прогнози щодо потенційних атак, поведінки користувача або вразливостей. Часто поєднуються з моделями поведінкової аналітики (User Behavior Analytics, UBA).

3. Генеративні системи - створюють новий контент: текст, зображення, код. Мають як захисне (генерація запитів, інтерфейсів), так і атакувальне застосування (соціальна інженерія, діпфейки).

4. Реагуючі системи - здатні автоматично здійснювати дії у відповідь на виявлені загрози: ізоляція вузла, повідомлення адміністратора, активація політик безпеки.

Цей функціональний поділ особливо важливий для проектування кібербезпекових рішень, що потребують адаптивності, швидкості реагування і автономії [5].

Класифікація ШІ за архітектурною побудовою

Відповідно до архітектури, що визначає, як інформація передається та обробляється всередині системи, системи ШІ поділяють на (Рис. 1.4):

Рис. 1.4. Види систем ШІ за архітектурою

- Символічні системи (Symbolic AI) - працюють на основі логічних правил, фактів і дедукції. Широко використовувалися у ранніх експертних системах.
- Нейронні мережі (Artificial Neural Networks) - моделі, натхненні біологічною будовою мозку. Добре працюють із неструктурованими даними (зображення, текст, звук).
- Гібридні моделі - поєднують символічний підхід із нейромережами. Вони дозволяють використовувати переваги обох: логіку та навчання.
- Еволюційні алгоритми - включають генетичні алгоритми та інші біоінспіровані моделі, що використовуються для оптимізації рішень.

У сфері інформаційної безпеки гібридні моделі є найбільш перспективними, оскільки вони можуть забезпечити як гнучкість навчання**,** так і інтерпретованість рішень**,** що критично важливо для довіри до систем.

Класифікація за типом навчання

Ця класифікація стосується способу, яким ШІ набуває знань:

- Наглядове навчання (Supervised Learning) - модель тренується на розмічених даних. Використовується для класифікації шкідливого ПЗ, визначення типу атак.
- Навчання без нагляду (Unsupervised Learning) - шукає приховані закономірності у даних без розмітки. Актуально для виявлення нових типів аномалій або невідомих загроз.
- Навчання з підкріпленням (Reinforcement Learning) - агент діє в середовищі та отримує винагороди/штрафи за свої дії. Використовується в адаптивних системах реагування.
- Напівконтрольоване навчання (Semi-supervised) - поєднує марковані та немарковані дані. Оптимальний варіант, коли повна розмітка занадто затратна.
- Самонавчання (Self-supervised Learning) - використовує внутрішню структуру даних для створення навчальних прикладів. Є основою сучасних LLM, зокрема GPT [6].

У сфері кібербезпеки важливо, що сучасні моделі часто комбінують кілька типів навчання для забезпечення більшої гнучкості, адаптивності й точності**.**

У наукових колах існують різні моделі систематизації ШІ. Наприклад, IBM Research пропонує класифікувати системи за принципом “intent-based intelligence”, тобто за цільовою спрямованістю системи: діагностична, стратегічна, операційна, тактична [12]. Google DeepMind розробляє власну класифікацію, де виділяє автономні, інтерактивні та реактивні моделі [13]. NIST (National Institute of Standards and Technology) у США пропонує використовувати концепцію «Explainable AI» (XAI) - здатність моделі пояснювати свої рішення як додаткову ознаку класифікації.

Класифікації ШІ активно використовуються при побудові систем кіберзахисту корпоративного рівня (Cisco SecureX, IBM QRadar); систем моніторингу поведінки користувачів (UEBA від Varonis, Exabeam); **і**нтелектуальних фаєрволів**,** які комбінують МН та експертні правила (Fortinet, Palo Alto); генеративних моделей для тестування уразливостей (AI fuzzing).

Правильне розуміння класифікації систем ШІ має вирішальне значення для вибору ефективної моделі захисту інформаційної системи. У кібербезпеці не існує універсального рішення, оскільки ефективність залежить від конкретного типу загроз, структури даних, рівня доступу до обчислювальних ресурсів та швидкості реагування.

Наприклад, для виявлення фішингу найчастіше застосовуються наглядові моделі на базі NLP, які класифікують вміст повідомлень; для аналізу мережевого трафіку - безнаглядові або гібридні моделі**,** що дозволяють виявити нетипову активність; навчання з підкріпленням дедалі активніше використовується у SOC-системах для автоматичного коригування політик захисту.

Кожен із підходів до класифікації дає змогу краще адаптувати систему ШІ до поставлених цілей, забезпечити гнучкість і масштабованість, а також підвищити точність виявлення загроз при мінімізації хибних спрацювань.

Визначення типу ШІ також критично важливе з точки зору етики, відповідальності та нормативного регулювання**.** Наприклад, системи, які діють автономно (реагують без участі оператора), мають бути високопередбачуваними й підлягати перевірці з боку людини. Це безпосередньо впливає на рівень довіри до рішень AI в умовах реагування на інциденти.

1.3 Технології ШІ, що використовуються у сфері кібербезпеки

Штучний інтелект (ШІ), як комплекс сучасних технологій, уже активно інтегрований у практику забезпечення кібербезпеки. Його застосування дозволяє значно підвищити ефективність виявлення атак, скоротити час реагування на інциденти та автоматизувати процеси аналізу великих обсягів даних.

Як свідчить статистика, впровадження ШІ у кібербезпеку організацій і підприємств дає їм значні переваги. Так, згідно з аналітичним звітом IBM "Cost of a Data Breach Report 2023" організації, які використовували рішення на базі штучного інтелекту, скоротили середній час виявлення порушення безпеки з 277 днів до 221 дня, а середня економія витрат на ліквідацію інцидентів становила до 1,76 мільйона доларів США в порівнянні з компаніями, які не використовували ШІ [8].

Дослідження компанії Gartner 2024 року показали, що понад 60% платформ безпеки підприємств будуть мати інтегровані функції машинного навчання або глибокого навчання для виявлення та реагування на загрози до 2025 року. Ці дані підтверджують, що інтеграція ШІ у системи захисту є не просто тенденцією, але й об'єктивною потребою для ефективної протидії сучасним цифровим загрозам.

Розглянемо основні технології ШІ, які вже використовуються або мають перспективи використання у сфері кібербезпеки.

Машинне навчання (Machine Learning)

МН є базовим напрямом розвитку ШІ. У кібербезпеці алгоритми МН застосовуються для виявлення аномалій в поведінці користувачів і систем; ідентифікації шкідливого ПЗ за характеристиками файлів або поведінкою програм; класифікації подій безпеки (розмежування нормальної та загрозливої активності); прогнозування можливих векторів атак на основі історичних даних.

Моделі, що найчастіше використовуються: дерева рішень, метод опорних векторів (SVM), випадкові ліси (Random Forest), градієнтний бустинг (XGBoost) [9].

Глибоке навчання (Deep Learning)

Глибокі нейронні мережі застосовуються для обробки великих обсягів неструктурованих даних, наприклад розпізнавання шкідливого коду за структурою байт-кодів; автоматичного виявлення вторгнень у складних мережевих потоках; виявлення фішингових сайтів за візуальними ознаками (скріншоти сайтів).

Особливість глибокого навчання полягає у високій здатності до виявлення складних шаблонів і нетипових аномалій, що значно перевершує традиційні підходи [10].

Обробка природної мови (Natural Language Processing, NLP)

Технології NLP дозволяють автоматизувати аналіз текстових повідомлень (електронна пошта, чати, форуми) на предмет потенційних загроз; виявлення фішингових атак**,** шкідливих посилань, спроб соціальної інженерії; розпізнавання емоційних чи маніпулятивних ознак у текстах.

Багато сучасних систем SOC використовують NLP-модулі для фільтрації великих обсягів повідомлень і ранжування інцидентів за пріоритетністю.

Генеративні моделі (Generative AI)

Генеративні моделі типу GAN (Generative Adversarial Networks) та LLM (Large Language Models) знаходять все більше застосування для моделювання можливих сценаріїв атак; створення фіктивних даних, які використовуються у тренуваннях систем безпеки (data augmentation); автоматичного генерування тестових сценаріїв у системах захисту. Проте важливо пам’ятати, що ті самі технології використовуються і у зловмисних цілях [11].

Автоматизація аналізу інцидентів (SOAR)

Системи автоматизації реагування (Security Orchestration, Automation and Response, SOAR) все більше інтегрують ШІ для автоматичного збору контексту інцидентів; пріоритезації загроз за ступенем критичності; запуску автоматичних сценаріїв реагування на основі розпізнаних патернів. Таким чином, застосування ШІ дозволяє перетворити аналіз безпеки з реактивної моделі (реагування після атаки) у превентивну та проактивну модель захисту.

Розглянемо кілька прикладів, де застосування ШІ довело свою ефективність.

Рання діагностика атак програм-вимагачів (ransomware): використовуючи поведінкове машинне навчання, деякі антивірусні рішення змогли запобігти шифруванню файлів ще на початкових стадіях зараження.

- Виявлення складних APT-кампаній**:** у кількох випадках Darktrace та Vectra AI виявили кіберрозвідку в корпоративних мережах задовго до того, як спрацювали традиційні сигнатурні системи.
- Автоматизація обробки фішингових атак**:** системи з NLP виявляють фішингові повідомлення навіть тоді, коли їх зміст змінено з метою обходу класичних правил фільтрації.
- SOAR-платформи на базі ШІ скоротили час реагування на інциденти з декількох годин до кількох хвилин завдяки автоматизації частини процесів [12].

1.4 Особливості і види сучасних загроз кібербезпеці

У сучасній цифровій реальності кібербезпека перетворилася на одну з ключових проблем для держав, бізнесу та окремих громадян. Водночас із розвитком ІТ збільшується кількість і складність загроз, що ставлять під сумнів стабільність, конфіденційність і доступність цифрової інформації.

На користь вище наведених положень свідчать результати статистики, зокрема згідно з дослідженнями Cybersecurity Ventures (2024) світові збитки від кіберзлочинності досягнуть $10,5 трильйона на рік до 2025 року; кількість атак із використанням ШІ зростає на 30% щороку. Практика показала, що середньостатистична компанія потребує понад 200 днів для виявлення складної атаки без залучення інструментів ШІ [13].

Як показало дослідження, зростання сучасних загроз кібербезпеці викликано низкою різних чинників. Насамперед це стрімка цифровізація усіх сфер життя, яка охоплює розвиток Інтернету речей (IoT), хмарних обчислень, мобільних технологій, створює величезні обсяги даних і збільшує кількість потенційних точок атаки.

Еволюція методів зловмисного впливу має наслідком створення умов, за яких кібератаки стали більш складними і цілеспрямованими з переважно цільовим характером впливу. Зловмисники використовують соціальну інженерію, шкідливі програми з елементами ШІ, засоби автоматизації нападу.

Нинішній ситуації у кібербезпеці притаманна доступність інструментів атаки, адже навіть недосвідчені кіберзлочинці можуть легко придбати на "темних" ринках готові рішення для проведення фішингових кампаній, написання шкідливого ПЗ або обходу систем захисту.

Значний вплив на розвиток галузі має широке застосування технологій ШІ та Big Data як для легітимних цілей захисту, так і для реалізації зловмисних атак. Саме ШІ дозволяє оптимізувати зусилля злочинців, прискорити підбір вразливостей, автоматизувати вторгнення, охопити великі цільові групи.

Глобалізація кіберзагроз є сучасною стійкою тенденцією, яка проявляється в тому, що кібератаки набувають міжнародного характеру, зачіпаючи одночасно великі масиви інфраструктури в різних країнах. Негативним наслідком глобалізації є відсутність кордонів у кіберпросторі, що ускладнює можливості виявлення й покарання кіберзловмисників [14].

Основні категорії загроз кібербезпеці (Рис. 1.5) охоплюють:

Рис. 1.5. Основні категорії кіберзагроз

- Шкідливе програмне забезпечення (Malware): віруси, трояни, програми-вимагачі, шпигунське ПЗ.

- Фішинг і соціальна інженерія, які мають на меті введення користувачів в оману з метою отримання конфіденційної інформації.
- Атаки на інфраструктуру охоплюють DDoS-атаки, підрив мережевої доступності, атаки на хмарні сервіси.
- Порушення конфіденційності даних шляхом крадіжки персональних даних, фінансових реквізитів, об’єктів інтелектуальної власності.
- Атаки на ланцюги постачання (Supply Chain Attacks) передбачають компрометацію програмного забезпечення або апаратних засобів на етапі виробництва або доставки.

Крім традиційних, виділяють також низку новітніх загроз, пов'язаних із використанням ШІ, серед яких автоматизовані фішингові атаки, які використовують персоналізовані листи, створені за допомогою NLP; діпфейки - підробні відео та аудіо для компрометації осіб чи організацій; так звані змагальні атаки (Adversarial attacks), які передбачають введення змінених даних для обману системи розпізнавання або захисту. Такі загрози стають дедалі складнішими для виявлення традиційними методами, що потребує застосування адаптивних і самонавчальних систем безпеки [15].

У сучасних умовах посилилася роль багатьох загроз, які раніше мали маргінальний характер або взагалі не існували, зокрема кібершпигунство, кібертероризм, інформаційні війни та гібридні загрози, зловмисне використання Інтернету речей, атаки на системи ШІ тощо.

Про зростання обсягів кібершпигунства свідчать дані про зростання кількості атак, спрямованих на отримання стратегічної інформації про державні органи, оборонні структури, великі корпорації. Такі атаки часто маскуються під звичайний трафік або офіційні запити. Терористичні організації дедалі частіше застосовують кібератаки для залякування населення, саботажу інфраструктури або дестабілізації соціальних процесів.

Ведення інформаційного протиборства і використання гібридних методів зловмисного впливу щорічно розширюють свої масштаби, маючи на меті

маніпулювання суспільною думкою, підрив довіри до урядів і бізнесу, нагнітання соціальної напруги, провокування конфліктів тощо. Нерідко вирішальну роль у таких випадках відіграє ШІ, який генерує переконливі фальшиві новини та маніпулятивний контент.

Зловживання Інтернетом речей (IoT) є наслідком того, що пристрої IoT часто мають низький рівень захисту, що робить їх легкою ціллю для атак ботнетів, типовим прикладом яких є атака Mirai. Атаки на системи ШІ передбачають порушення цілісності моделей ШІ через «отруєння» даних навчання (Data Poisoning), викрадення моделей (model stealing) та обхід механізмів розпізнавання [16].

Наведемо кілька конкретних прикладів масштабних кібератак останніх років.

Однією з найбільших атак на ланцюги постачання є SolarWinds 2020 року. Кібератака SolarWinds була атакою на ланцюжок постачання програмного забезпечення за участю платформи SolarWinds Orion, під час якої кіберзлочинці отримали доступ до систем SolarWinds і розгорнули троянські оновлення програми Orion. Це, у свою чергу, дозволило їм встановлювати приховане шкідливе ПЗ в мережах клієнтів SolarWinds. Злом SolarWinds був розкритий кількома компаніями з кібербезпеки спільно з Агентством з кібербезпеки та безпеки інфраструктури США (CISA) у грудні 2020 року [17].

Атака шкідливого ПЗ на американську трубопровідну систему Colonial Pipeline, що сталася у травні 2021 року, зупинила роботу всіх трубопроводів системи на п'ять днів. Внаслідок атаки президент США Джо Байден оголосив надзвичайний стан. За оцінками преси, це була найбільша успішна кібератака на нафтову інфраструктуру США, яка була здійснена групою хакерів DarkSide [18].

У 2021 році кілька постачальників керованих послуг (MSP) та їхні клієнти стали жертвами атаки програм-вимагачів, здійсненої групою REvil, що призвело до масового простою понад 1000 компаній. Атака була здійснена з використанням уразливості у VSA (Virtual System Administrator), програмному пакеті віддаленого моніторингу й управління від компанії Kaseya [19].

Фішингові кампанії під час пандемії COVID-19 є типовим прикладом, коли зловмисники активно використовували страх і невизначеність суспільства для розсилки шкідливих листів, маскування під державні установи або організації охорони здоров'я.

Слід відзначити, що глобальні прогнози щодо розвитку ландшафту кіберзагроз також викликають занепокоєння. Експерти аналітичних центрів, таких як Світовий економічний форум, прогнозують, що у найближчі 5–10 років атаки на AI та автоматизовані системи будуть зростати експоненційно; поява квантових обчислень призведе до потенційної загрози для традиційних криптографічних алгоритмів; основною мішенню кібератак стане критична інфраструктура (енергетика, водопостачання, транспорт). Окремим викликом для систем кіберзахисту залишатимуться складні постійні загрози (Advanced Persistent Threats, APT), які розвиваються приховано протягом тривалого часу і використовують складні інструменти ШІ, зокрема для маскування.

1.5 Вплив ШІ на підходи до забезпечення кібербезпеки

ШІ кардинально змінює традиційне уявлення про кібербезпеку, трансформуючи не тільки інструменти захисту, а й самі принципи побудови безпечних цифрових середовищ. Розвиток інтелектуальних систем вимагає перегляду підходів до виявлення, запобігання та реагування на загрози.

Традиційна модель кіберзахисту базується на виявленні відомих загроз на основі сигнатур (signature-based detection) та реалізації заздалегідь визначених політик доступу. Проте в умовах динамічного середовища загроз і появи нових векторів атак такі підходи стають недостатньо ефективними.

Саме ШІ є засобом підвищення рівня кібербезпеки шляхом переходу від реактивного захисту до проактивного прогнозування й запобігання загрозам, від ручного аналізу інцидентів до автоматизованого реагування на основі моделей навчання, від статичних політик до динамічної адаптації систем безпеки на основі змін середовища. ШІ дозволяє суттєво скоротити час виявлення атак

(Mean Time to Detect, MTTD) та час реагування (Mean Time to Repair, MTTR), а також зменшити навантаження на команди безпеки [20].

Завдяки ШІ виникли нові концепції в кібербезпеці. Коротко охарактеризуємо основні з них.

Архітектура адаптивної безпеки (Adaptive Security Architecture) передбачає використання моделей**,** що навчаються у режимі реального часу і здатні змінювати свої параметри в залежності від поведінки користувачів, пристроїв або додатків.

Архітектура нульової довіри (Zero Trust Architecture) просуває принцип «нікому не довіряти, всіх перевіряти», який посилюється завдяки застосуванню AI для динамічної оцінки ризиків кожного запиту на доступ.

Розвідка загроз (Threat Hunting) за допомогою ШІ використовує інтелектуальних агентів для активного пошуку загроз навіть у відсутності явних ознак компрометації.

Пояснюваність AI (Explainable AI, XAI) - це концепція, згідно з якою модель ШІ має бути здатною пояснити отриманий нею результат на зрозумілому для людини рівні. Потреба в поясненні рішень систем ШІ у сфері кібербезпеки набуває особливої актуальності для мінімізації ризику помилкових або упереджених рішень, які, в кінцевому рахунку, приймає кіберфахівець.

Реагування на інциденти на основі ШІ (AI-driven Incident Response) охоплює використання інтелектуальних систем, які автоматично формують сценарії реагування на основі типу інциденту, минулого досвіду і прогнозованого розвитку атаки.

Крім перелічених вище, також варто згадати інші інноваційні концепції кібербезпеки на основі ШІ.

Прогностична розвідка загроз кібербезпеці (Predictive Security Intelligence), сутність якої полягає в тому, що інтелектуальні системи аналізують поточні тренди атак, вразливості та геополітичні події для побудови прогнозів щодо можливих атак на конкретну інфраструктуру [21].

Незалежний кіберзахист (Autonomous Cyber Defense) втілює ідею цілком автономних агентів безпеки, які самі виявляють, аналізують і усувають загрози без втручання людини.

Когнітивний (пізнавальний) операційний центр безпеки (Cognitive Security Operations Center, SOC) – це SOC нового покоління, де рішення приймаються не лише на основі правил, а й завдяки когнітивному аналізу великих даних.

Технології введення в оману на основі ШІ (AI-Enabled Deception Technologies): інтелектуальні системи створюють фальшиві активи, обманюючи зловмисників і скеровуючи їх у контрольовані пастки.

Застосування ШІ в галузі кібербезпеки призводить до значних позитивних змін у її основних компонентах, які представлені у таблиці 1.1.

Таблиця 1.1.

Вплив AI на основні компоненти кібербезпеки

Компонент кібербезпеки	До впровадження ШІ	Після впровадження ШІ
Виявлення загроз	Реактивне, на основі сигнатур	Проактивне, поведінкове, на основі МН
Аналіз інцидентів	Ручний аналіз, повільна обробка	Автоматизований, в режимі реального часу
Реагування на атаки	Переважно вручну, з затримками	Автоматизоване, з миттєвим реагуванням
Прогнозування загроз	Засноване на історичних даних	Прогноз на основі трендів, моделей, навчання
Контроль доступу	Фіксовані правила	Динамічна адаптація політик доступу
Фішинг і соцінженерія	Виявлення вручну або за шаблоном	Аналіз тексту NLP, виявлення прихованих атак
Оцінка ризиків	Статична оцінка	Динамічна, на основі даних в реальному часі

Водночас, слід розуміти, що впровадження ШІ у забезпечення кібербезпеки поряд з численними перевагами спричиняє появу низки викликів і обмежень (Рис. 1.6).

ПЕРЕВАГИ	НЕДОЛІКИ
Швидкість обробки даних	Залежність від обсягів і якості даних
Висока точність виявлення загроз	Вразливість моделей ШІ до атак
Адаптивність до нових загроз	Етичні та правові проблеми
Автоматизація операцій	Брак прозорості й довіри
Зменшення впливу людського фактора	Ризик помилкових спрацювань

Рис. 1.6. Переваги і ризики використання ШІ в кібербезпеці

Отже, основними перевагами інтеграції ШІ у кібербезпеку є:

- висока швидкість обробки даних (аналіз тисяч подій за секунду);
- підвищена точність виявлення загроз, в тому числі нових і складних атак;
- адаптивність до нових типів загроз без необхідності перепрограмування;
- автоматизація стандартних операцій з аналізу та реагування;
- мінімізація впливу людського чинника внаслідок зменшення кількості помилок через перевтому або необізнаність персоналу.

Попри велику перспективність, застосування ШІ в забезпеченні кіберзахисту ІКС супроводжується низкою проблем, серед яких:

- залежність від обсягів і якості даних - полягає в тому, що недостатні, неправильні або упереджені дані можуть призвести до серйозних помилок у рішеннях;
- вразливість самих моделей ШІ до атак в ході навчання (poisoning attacks), що може призвести до маніпулювання навчальними даними або введення "отруєних" даних;

- етичні і правові аспекти використання ШІ, наприклад автономне ухвалення рішень ШІ про блокування користувачів або надання доступу може мати як етичні, так і правові наслідки;
- проблема прозорості і довіри до рішень ШІ - проявляється у значних труднощах, пов’язаних із поясненням прийнятих ШІ рішень для людей і виникненням ризиків для ефективного управління системами;
- ризик хибних позитивних або негативних спрацювань [22].

Висновки до розділу 1

У першому розділі проведено теоретичне обґрунтування ролі та напрямів застосування технологій штучного інтелекту у сфері кібербезпеки. Здійснено аналіз сутності, етапів розвитку і класифікації ШІ, а також досліджено основні загрози, що виникають у цифрову епоху, та трансформацію концепцій інформаційного захисту під впливом інтелектуальних систем.

Встановлено, що ШІ, від часу його зародження, зазнав суттєвих змін: від експертних систем до глибоких нейронних мереж і генеративних моделей нового покоління. Класифікація ШІ за рівнем когнітивної складності, архітектурою та методами навчання дозволяє систематизувати існуючі підходи до його використання в сфері кібербезпеки.

У ході дослідження виявлено основні технології ШІ, які використовуються в кібербезпеці: машинне навчання, глибоке навчання, обробка природної мови, генеративні моделі та автоматизація аналізу інцидентів. Показано, що інтеграція цих технологій дозволяє суттєво підвищити ефективність виявлення, прогнозування та запобігання кіберзагрозам.

Окрему увагу приділено аналізу сучасних загроз кібербезпеці, серед яких особливе місце займають атаки із застосуванням ШІ: автоматизовані фішингові кампанії, створення діпфейків, атаки на моделі машинного навчання тощо. Встановлено, що цифровізація суспільства значно розширила площу

потенційних атак і ускладнила завдання забезпечення конфіденційності, цілісності та доступності даних.

У результаті аналізу зроблено висновок, що вплив ШІ на кібербезпеку має подвійний характер: з одного боку, інтелектуальні системи підвищують рівень захисту, з іншого - породжують нові ризики і виклики, пов'язані із залежністю від якості даних, прозорістю рішень, етичністю і легітимністю застосування. У зв'язку з цим концепція кібербезпеки в сучасних умовах вимагає комплексної інтеграції інтелектуальних технологій з обов'язковим урахуванням правових, етичних і соціальних аспектів.

РОЗДІЛ 2. ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ШІ У ЗЛОВМИСНІЙ КІБЕРДІЯЛЬНОСТІ

Швидкий розвиток технологій ШІ відкрив нові можливості не лише для підвищення ефективності кіберзахисту, але й для проведення зловмисних дій у цифровому середовищі. Висока адаптивність, здатність до самонавчання, генерація контенту та автоматизація складних процесів роблять алгоритми ШІ потужним інструментом у руках кіберзлочинців.

У розділі здійснено глибокий аналіз методів використання ШІ у шкідливих кіберактивностях. Основною метою є виявлення механізмів, інструментів і стратегій застосування ШІ у процесах фішингу, генерації шкідливого ПЗ, обходу систем захисту та проведення складних багатоступеневих атак.

Для досягнення поставленої мети використані такі методи дослідження: аналіз конкретних кейсів кібератак із застосуванням інтелектуальних технологій; порівняльний аналіз ефективності атак із традиційними методами; метод моделювання сценаріїв атак на основі сучасних генеративних алгоритмів; бібліографічний метод, який охоплював вивчення сучасних наукових і практичних джерел щодо тенденцій розвитку кіберзагроз.

2.1 Стратегії кібератак з використанням алгоритмів ШІ

Штучний інтелект змінює традиційну картину кіберзагроз, дозволяючи здійснювати більш складні, адаптивні та масовані атаки. На відміну від класичних підходів, алгоритми ШІ дають змогу автоматизувати процеси збору інформації, створення шкідливого контенту, ухвалення рішень про напрямок атак та обхід засобів захисту.

Основою застосування ШІ у кіберзлочинності є здатність систем:

- Аналізувати великі обсяги даних для побудови профілів жертв.
- Автоматично створювати високоякісний фішинговий контент.
- Виявляти вразливості в інформаційних системах без втручання людини.

- Моделювати поведінку для обходу систем виявлення атак (IDS/IPS).

Проведення дослідження стратегій атак із ШІ ґрунтується на використанні методів аналізу інцидентів, порівняння технологій і моделювання можливих сценаріїв загроз.

Для аналізу використано методи: аналіз кейсів для вивчення реальних прикладів атак із застосуванням інтелектуальних технологій; метод порівняння для зіставлення традиційних та інтелектуально-орієнтованих атак за критеріями складності, ефективності й масштабу; метод моделювання для розробки гіпотетичних сценаріїв атак на базі існуючих інструментів ШІ; огляд літературних джерел: аналіз аналітичних звітів (Microsoft, IBM, Cybersecurity Ventures) [23].

На основі аналізу виявлено кілька основних стратегій використання ШІ для зловмисної діяльності.

Автоматизоване профілювання жертв. Алгоритми МН аналізують публічні джерела (соціальні мережі, реєстри доменів, відкриті бази даних) для побудови профілю жертви. Це дозволяє адаптувати атаки під конкретну особу чи організацію. Прикладом такої стратегії є атаки на топ-менеджерів компаній через персоналізовані фішингові листи, що базуються на їхній публічній активності.

Генерація фішингового контенту. Застосування великих мовних моделей (наприклад, ChatGPT, Gemini) для створення фішингових листів без граматичних помилок, із правильною стилістикою та персоналізацією. Прикладом є генерація запрошень на фіктивні бізнес-заходи від імені справжніх компаній для отримання даних користувачів.

Створення обхідних сценаріїв атак. Змагальне машинне навчання (Adversarial Machine Learning, AML) використовується для створення даних, що вводять в оману системи виявлення атак. Наприклад, змінюють структуру мережевого трафіку або шкідливих файлів так, що вони виглядають легітимними для антивірусів. У якості прикладу можна навести формування шкідливих файлів, що проходять перевірку на безпеку завдяки спеціально зміненій сигнатурі.

Автоматизація сканування та експлуатації вразливостей. AI дозволяє автоматично проводити сканування мереж, виявляти слабкі місця і пропонувати найкращі вектори атаки на основі аналізу ситуації. Прикладом є автоматичні фреймворки на основі навчання з підкріпленням (Reinforcement Learning) для динамічного тестування мереж на наявність вразливостей.

Генерація шкідливого ПЗ. Генеративні моделі (GAN) застосовуються для створення варіацій шкідливого програмного забезпечення, яке важко розпізнати традиційними методами. Прикладом є створення шкідливого коду для атаки на корпоративні системи управління даними [24].

Додаткові сучасні приклади атак із застосуванням ШІ.

Атака з використанням діпфейків проти банківської системи (2020 рік). Група кіберзлочинців використала технології генерації діпфейків для створення аудіо-файлів із голосом генерального директора міжнародної компанії. Використавши хибну телефонну розмову, зловмисники переконали банківських співробітників перевести понад 35 мільйонів доларів на підконтрольні їм рахунки [25].

Автоматизовані фішингові кампанії із застосуванням ШІ у COVID-19 період. Під час пандемії значна кількість фішингових атак була створена автоматичними генеративними моделями на тематику COVID-19, що підвищило рівень довіри жертв і успішність атак.

Атаки змагального МН, спрямовані проти систем розпізнавання обличчя. Дослідники показали, що модифікація кількох пікселів на обличчі дозволяє системам безпеки ідентифікувати людину як іншу особу або взагалі не впізнати її, що створює загрозу обходу біометричних систем аутентифікації.

Основні типи кібератак із використанням ШІ та їх характеристики подано в таблиці 2.1.

Таблиця 2.1.

Порівняльна характеристика кібератак, що використовують ШІ

Тип атаки	Застосування ШІ	Приклад використання	Особливості та ризики
Фішинг	Генерація текстів, персоналізація	ChatGPT створює листи з реалістичним контекстом	Висока правдоподібність, складність фільтрації
Соціальна інженерія	Моделювання поведінки, генерація сценаріїв	LLM симулює діалог із жертвою	Автоматизоване введення в оману
Діпфейки	GAN генерує фейкове відео/аудіо	Підробка виступів чи голосових команд	Удар по репутації, обхід біометрії
Обхід систем виявлення	Адаптивне навчання на сигнатурах	Атаки, які змінюють форму для уникнення IDS/IPS	Складність виявлення нових загроз
Атаки на ML-моделі (poisoning)	Маніпуляція даними для навчання	Отруєння навчального набору даних	Виведення з ладу захисних моделей
Анонімізація атак	Маскування активності через обхід патернів	Зміна трафіку для уникнення шаблонів виявлення	Зниження ефективності класичних засобів

2.2 Автоматизація шкідливих процесів: фішинг, спам, розсилки

Фішинг, спам та інші методи масових розсилок були і залишаються одним із найбільш ефективних інструментів кіберзлочинців. Проте з появою та

розвитком технологій ШІ шкідливі процеси отримали новий рівень автоматизації, персоналізації та масштабування.

Автоматизація на основі ШІ дозволяє:

- генерувати персоналізовані повідомлення для кожної потенційної жертви;
- модифікувати зміст атак у реальному часі залежно від реакцій отримувача;
- аналізувати поведінку користувачів після першої взаємодії та адаптувати подальші спроби атаки.

Фішинг і спам, підсилені ШІ, стають дедалі важчими для виявлення навіть сучасними системами кіберзахисту [26].

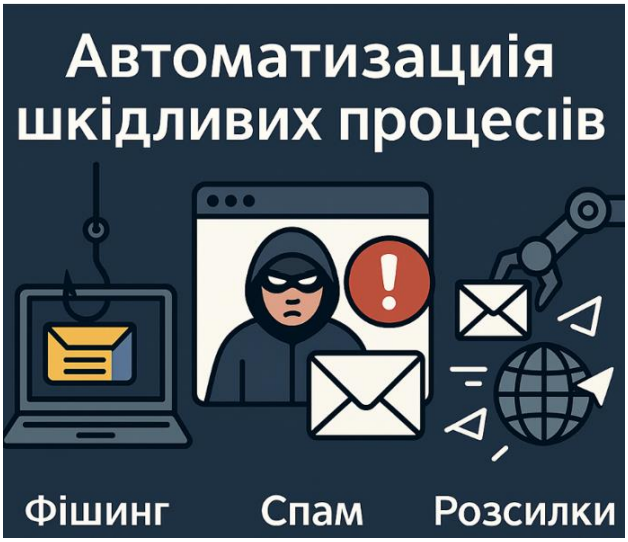

Рис. 2.1. Приклади автоматизації шкідливих процесів

Інтелектуалізація фішингових атак. Однією з ключових тенденцій є інтелектуалізація фішингових атак. Завдяки алгоритмам обробки природної мови (NLP), зловмисники можуть створювати фішингові листи, які точно імітують стиль офіційної кореспонденції конкретних компаній або організацій. Наприклад, електронні листи від імені банків або служб доставки адаптуються під регіональні мовні особливості та персональні дані жертви, що підвищує їхню правдоподібність та ефективність.

Генерація масових атак із персоналізацією. Ще одним напрямом розвитку є генерація масових атак із персоналізацією. Використання великих мовних моделей, таких як GPT, дає змогу створювати тисячі унікальних версій одного і того самого фішингового повідомлення. Це дозволяє значно знизити

ймовірність їхнього виявлення антивірусними фільтрами. Наприклад, може бути автоматично згенеровано серію персоналізованих листів до працівників великої корпорації із вкладеними шкідливими посиланнями.

Спам-боти нового покоління. ШІ також активно використовується для створення ботів нового покоління. Такі спам-боти можуть вести діалог із жертвами через електронну пошту, месенджери або соціальні мережі, адаптуючи свої відповіді до реакцій користувача. Наприклад, у месенджерах WhatsApp або Telegram бот може підтримувати розмову, поступово спонукаючи людину натиснути на шкідливе посилання або розкрити конфіденційну інформацію.

Адаптивне планування розсилок. Іншим важливим аспектом є адаптивне планування фішингових розсилок. Інтелектуальні системи аналізують часові пояси, поведінкові патерни користувачів і обирають найоптимальніший момент для надсилання листів. Завдяки цьому фішингові повідомлення потрапляють до користувачів саме в години їхньої найвищої активності, наприклад, у розпал робочого дня, коли люди найбільш сконцентровані на корпоративній пошті.

Обхід фільтрів і антиспам-систем. Нарешті, значну увагу зловмисники приділяють створенню повідомлень, здатних обійти фільтри спаму та системи захисту. Для цього використовуються методи зміни структури тексту, вставляння прихованих символів або навіть кодування основного змісту повідомлення у вигляді зображення. Таким чином, фішингові листи залишаються непоміченими для автоматичних систем захисту, але зберігають повну інформативність для людини [27].

Загалом, автоматизація шкідливих процесів за допомогою ШІ відкриває новий рівень складності для систем кіберзахисту, вимагаючи застосування адаптивних і проактивних заходів реагування.

Реальні приклади автоматизованих шкідливих кампаній

Фішинг через автоматизовані чат-боти (2022). Під час атак на платформи електронної комерції, зловмисники застосовували автоматизовані чат-боти на основі NLP, які проводили користувача через процес "верифікації

акаунта". Бот імітував службу підтримки, задаючи уточнювальні питання і отримуючи персональні дані без жодної участі людини [28].

Масові розсилки під час Black Friday (2021–2023). Зафіксовані кампанії, де нейронні мережі створювали тисячі унікальних пропозицій про фіктивні розпродажі. Кожен лист був адаптований до геолокації, історії покупок і навіть мовних уподобань користувача [29].

Використання GAN для обходу спам-фільтрів. Генеративні змагальні мережі застосовувалися для створення фальшивого контенту (зображення, шифровані повідомлення), який обминав традиційні фільтри поштових сервісів і виявлявся лише після ретельного ручного аналізу [30].

2.3 Генеративні моделі (GAN, LLM) як інструменти атак

Генеративні моделі є одним із найбільш динамічних і водночас суперечливих напрямків розвитку AI. Вони не тільки відкрили нові горизонти у творчості, медицині та IT, а й створили значні ризики для кібербезпеки.

До основних типів генеративних моделей належать:

- **Генеративні змагальні мережі (GANs)** - моделі, які навчаються створювати реалістичні дані (зображення, текст, відео) через протиставлення двох нейронних мереж: генератора і дискримінатора [31] (Рис. 2.2).

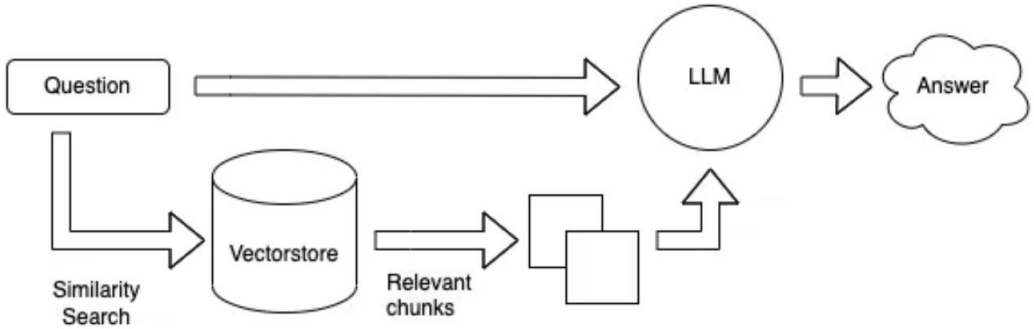

Рис. 2.2. Принцип дії великих мовних моделей (LLMs)

- **Великі мовні моделі (LLMs)** - такі як GPT, Claude, Gemini, які генерують зв'язний текст на основі заданих підказок (prompts) (Рис. 2.3).

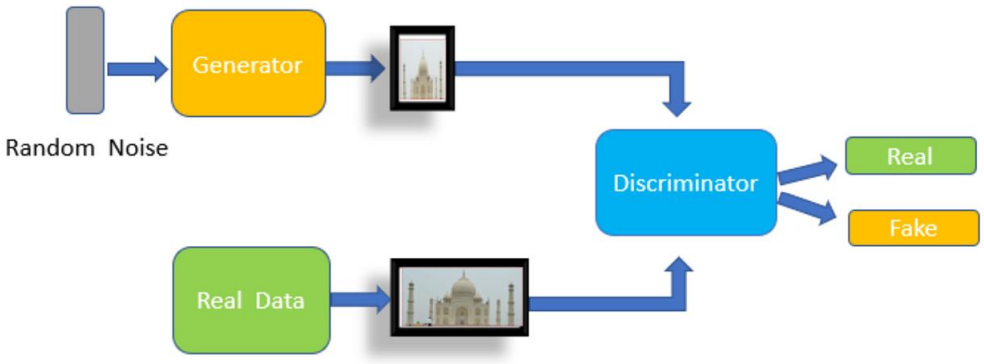

Рис. 2.3. Принцип дії генеративних змагальних мереж (GANs)

У контексті кібербезпеки ці технології набувають нового значення, оскільки їх можна використовувати як для автоматизації атак, так і для обходу традиційних засобів захисту.

Розглянемо основні способи використання генеративних моделей у кібератаках.

Створення фішингових листів і чат-ботів. Одним із найпоширеніших напрямів є створення фішингових повідомлень і чат-ботів на основі LLM. Такі моделі здатні генерувати персоналізовані повідомлення, які враховують мовні звички жертви, дотримуються офіційного стилю тієї чи іншої компанії та адаптуються в реальному часі відповідно до реакцій користувача. Наприклад, у фішинговій переписці модель може автоматично формувати відповіді на запити жертви, продовжуючи комунікацію і збільшуючи ймовірність досягнення мети атаки [32].

Генерація діпфейків для атак соціальної інженерії. Ще одним критично важливим напрямом є генерація діпфейків (deepfake) для соціальної інженерії. Застосування GAN дозволяє створювати фальшиві відео- та аудіоматеріали, які важко відрізнити від реальних. Такі відеодзвінки можуть імітувати голос і зовнішність керівника компанії, використовуючись для фальшивого погодження фінансових операцій. Наприклад, у резонансному випадку атаки на міжнародний

банк було використано діпфейк-відео, в якому «керівник» дав вказівку на переказ значної суми коштів [33].

Створення підробних вебсайтів і документів. Також генеративні моделі використовуються для створення фальшивих вебсайтів та документів. З їх допомогою формуються сторінки фішингових сайтів, які візуально і функціонально імітують реальні платформи, наприклад, банківські сайти чи корпоративні системи авторизації. Окрім того, LLM і GAN здатні генерувати переконливі фальшиві документи, такі як контракти, рахунки-фактури чи службові листи, що використовуються у BEC-атаках (Business Email Compromise). Прикладом є автоматизоване створення копії банківської сторінки з використанням генеративного шаблонізатора.

Атаки на системи виявлення шкідливого ПЗ. Ще одним важливим напрямом застосування генеративних моделей ШІ є обхід систем виявлення зловмисних програм і антивірусного захисту**.** Завдяки можливості динамічно змінювати вміст шкідливих файлів або повідомлень, зловмисники здатні створювати ПЗ, яке обходить класичні сигнатурні фільтри. Наприклад, може бути створений документ, що змінює свої характеристики при кожному скануванні системою безпеки, уникаючи виявлення.

Практичні дослідження підтверджують ефективність таких підходів. У 2023 році Google DeepMind спільно з Університетом Каліфорнії продемонстрували, що навіть відкриті генеративні моделі можна використати для створення фальшивого біомедичного контенту. Це свідчить про потенційну загрозу для медичних і фінансових систем, які покладаються на цифрову достовірність інформації. У свою чергу, протягом 2023–2024 років масовим явищем стали фішингові кампанії, здійснені з використанням ChatGPT. Зловмисники застосовували техніки prompt injection для створення персоналізованих повідомлень, орієнтованих на співробітників корпоративних структур.

У підсумку слід зазначити, що генеративні моделі відіграють дедалі вагомішу роль у реалізації складних, динамічних та персоналізованих кібератак, ускладнюючи завдання сучасних систем кіберзахисту.

2.4 Застосування глибокого навчання в атаках на системи розпізнавання

Глибоке навчання (Deep Learning) стало основою для побудови сучасних систем розпізнавання: облич, мови, поведінки користувачів, а також обробки біометричних даних. Проте ті самі алгоритми глибокого навчання використовуються зловмисниками для обходу або компрометації таких систем.

Відмінності у дії МН та DL показані на рис. 2.4.

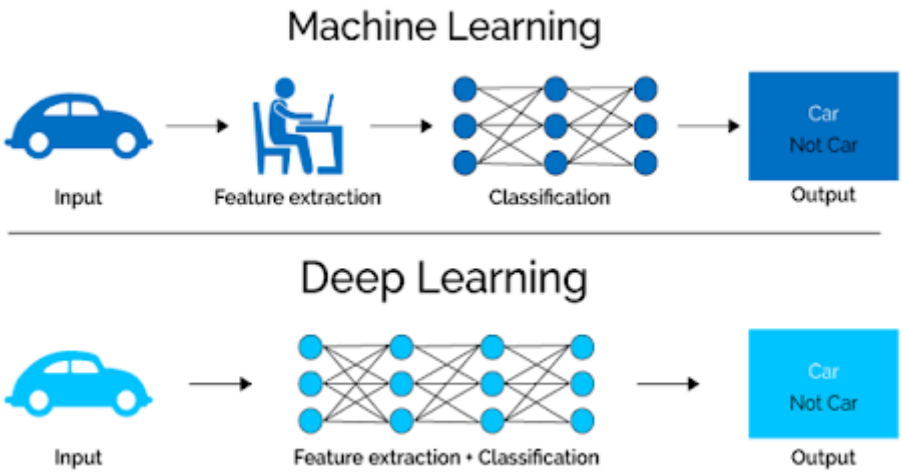

Рис. 2.4. Відмінності у дії машинного навчання та глибокого навчання

Суть атаки з використанням DL полягає у маніпулюванні вхідними даними або створенні спеціально підготовлених прикладів, що призводить до неправильної роботи моделей розпізнавання.

У рамках дослідження було застосовано комплексний підхід до аналізу сучасних методів атак на системи розпізнавання, які базуються на DL. Основну увагу було приділено атакам у сферах комп'ютерного зору (Computer Vision) та обробки природної мови (NLP), оскільки саме ці напрямки найчастіше використовуються у критичних системах безпеки та ідентифікації. Розглянемо основні механізми атак на системи розпізнавання [34].

Атаки шляхом внесення мікроскопічних змін (Adversarial Attacks). Одним із найпоширеніших способів обходу систем є так звані змагальні атаки,

суть яких полягає у внесенні мікроскопічних змін до вхідних даних. Ці зміни є майже непомітними для людського ока, проте вони змінюють результати роботи нейронної мережі. Наприклад, незначна модифікація кількох пікселів на фотографії може змусити систему розпізнати зображену особу як зовсім іншу, чим активно користуються зловмисники.

Атаки на біометричні системи. Окрему загрозу становлять атаки на біометричні системи. Нейронні мережі, які забезпечують розпізнавання облич, райдужної оболонки ока, голосу чи інших біометричних ознак, можуть бути обмануті за допомогою змінених або підроблених даних. Наприклад, використання 3D-масок або відео- і аудіо-діпфейків дозволяє зловмисникам проходити ідентифікацію замість справжніх користувачів.

Підробка текстової інформації в NLP-системах. Не менш небезпечними є атаки на системи обробки текстової інформації. У системах, що використовують для верифікації документів або виявлення шахрайства, можна застосовувати тексти, які зовні виглядають нормальними, але є модифікованими таким чином, щоб система не змогла виявити підробку. Генеративні моделі дозволяють створювати фальшиві відгуки, які маскуються під справжні та впливають на довіру до онлайн-платформ.

Атаки з розпізнаванням в аудіо- та відеопотоках. Значну загрозу становлять і маніпуляції з аудіо- та відеоданими. Сучасні системи розпізнавання мови та емоцій можуть бути обдурені шляхом заміни окремих елементів сигналу. Наприклад, зміна інтонації або додавання "невидимих шумів" може суттєво вплинути на точність класифікації або змусити систему зробити неправильний висновок.

У практичному контексті широкого розголосу набула атака Fast Gradient Sign Method (FGSM)**,** яка дозволяє швидко створювати зразки для тестування стійкості системи до adversarial впливу. Дослідження, проведені у 2023 році, продемонстрували вразливість таких комерційних рішень, як Amazon Rekognition і Microsoft Azure Face API, до спеціально модифікованих зображень.

Фізичні змагальні атаки. Серед новітніх загроз також слід виокремити фізичні змагальні атаки**.** На відміну від цифрових атак, у цьому випадку застосовуються фізичні об’єкти або елементи, які змінюють зовнішній вигляд особи в очах системи. Наприклад, наклеювання спеціальних стікерів на обличчя або використання одягу з певними візерунками може призвести до того, що система відеоспостереження не зможе правильно ідентифікувати людину. Такі результати були підтверджені у дослідженні OpenAI у 2024 році.

Атаки в обхід. Особливо небезпечними є backdoor-атаки**,** коли під час навчання або оновлення моделі у неї навмисно впроваджують прихований тригер. У присутності цього тригера система починає діяти некоректно. Наприклад, система розпізнавання облич може пропускати певних осіб, якщо ті носять визначений предмет, як-от шарф певного кольору.

Атаки «отруєння» даних (Data Poisoning). Ще одним вектором атак є **атаки типу** «отруєння» даних, коли у навчальний набір підкидаються змінені або фальшиві дані. У результаті модель вчиться на спотвореній інформації, що значно знижує її ефективність у реальному середовищі. Типовим прикладом є випадки, коли антивірусна система навчається на "забруднених" даних і починає пропускати шкідливе програмне забезпечення [35].

Загалом, атаки на системи розпізнавання стають дедалі складнішими, охоплюючи не лише цифрову, а й фізичну складову. У таких умовах надзвичайно важливо вивчати не лише захисні алгоритми, але й потенційні вектори атак, які можуть бути використані зловмисниками для обходу навіть найсучасніших систем безпеки.

2.5 Атаки на основі обхідних алгоритмів: маніпуляція даними і системами

З розвитком систем ШІ і автоматизованих засобів аналізу загроз, зловмисники почали розробляти нові стратегії атак, спрямовані не на пряме знищення чи втручання, а на обхід захисту шляхом маніпуляції даними або

логікою роботи систем**.** Такі обхідні атаки часто є невидимими для традиційних засобів виявлення загроз і становлять особливу небезпеку для складних IT-інфраструктур.

Обхідні алгоритми (evasion techniques) дозволяють:

- приховувати сліди присутності в системі;
- модифікувати поведінку шкідливого ПЗ так, щоб воно виглядало легітимним;
- уникати спрацювання сигнатур і поведінкових детекторів.

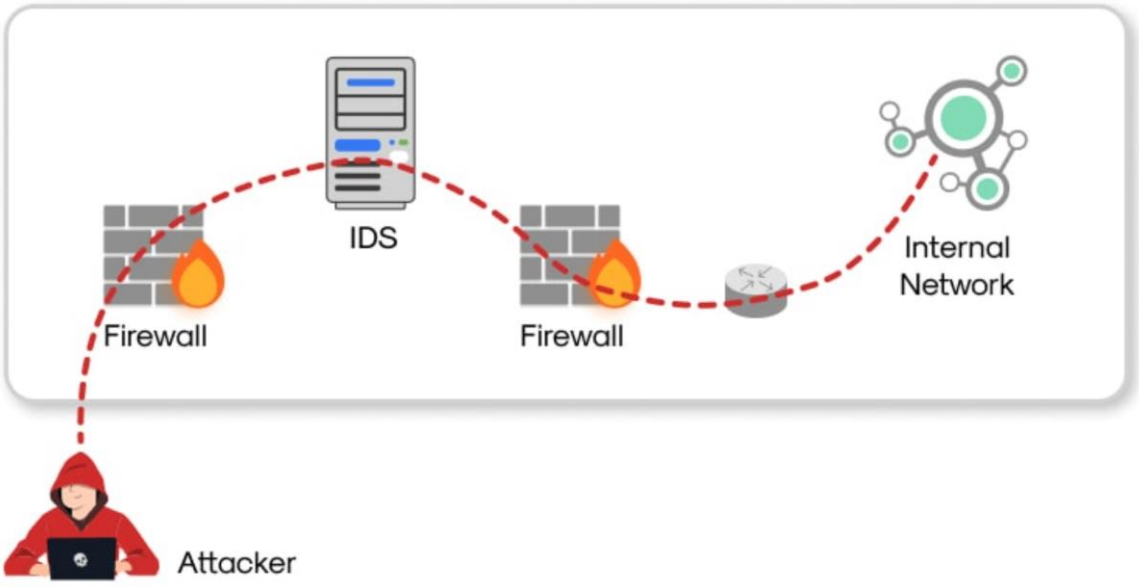

Рис. 2.5. Принцип використання обхідних алгоритмів

Однією з актуальних проблем у сфері кібербезпеки є здатність зловмисників успішно обходити наявні системи захисту. У розділі розглянуто основні види атак, що базуються на методах приховування зловмисної активності від засобів виявлення. Для вивчення цього явища було застосовано метод функціонального аналізу алгоритмів обходу, огляд відкритих інструментів маскування, аналіз прикладів реальних атак, а також проведено оцінку ефективності існуючих систем захисту в умовах застосування зазначених технік [36].

Маніпуляція вхідними даними. Одним із найпоширеніших методів є маніпуляція вхідними даними**.** Шкідливе програмне забезпечення може змінювати власну структуру, назви функцій або шифруватися, а також

розбиватися на кілька частин, які запускаються по черзі. Усе це дозволяє уникнути виявлення за допомогою сигнатурних або евристичних методів аналізу. Типовим прикладом є перейменування основного файлу шкідливого ПЗ та його запуск через проміжне легітимне програмне забезпечення.

Динамічна поведінка у середовищі. Ще однією поширеною технікою є динамічна поведінка у середовищі**.** Зловмисне програмне забезпечення часто спроектоване так, щоб залишатися пасивним, якщо виявляє ознаки аналізу (наприклад, пісочницю чи віртуальне середовище). Воно активується лише за умови, якщо функціонує у справжньому середовищі користувача, і лише при відповідності певним умовам, наприклад, IP-адресі, геолокації чи поведінці користувача. Таке ПЗ здатне роками залишатися непоміченим у системі.

Обхід аналітичних систем ШІ. З розвитком систем ШІ почастішали атаки на аналітичні системи ШІ. У цьому випадку зловмисники навмисне модифікують дані так, щоб аналітичні інструменти робили хибні висновки. Наприклад, змінюється частота появи шкідливих запитів або розсіюються ознаки атаки у часі, аби жоден окремий етап не викликав підозри. Це дозволяє створити повноцінну атаку, яка виглядає як серія незначних дій.

Обхід аутентифікації. Іншим критично важливим вектором є обхід аутентифікації**.** Використовуючи підроблені токени, зламані сесії або бічні канали, зловмисники отримують доступ до систем без необхідності проходження стандартної перевірки. Наприклад, компрометація токена OAuth у хмарному сервісі дозволяє обійти багатофакторну автентифікацію і діяти від імені користувача.

Використання легітимних інструментів. Окремо варто відзначити техніку Living off the Land (LotL), яка передбачає використання легітимних інструментів операційної системи для проведення атак. У цьому випадку зловмисники використовують вбудовані засоби, такі як PowerShell, Task Scheduler, PsExec та інші утиліти, щоб обійти контроль над встановленням стороннього ПЗ. Наприклад, за допомогою PowerShell можна розгорнути шкідливий код без запису на диск, так зване безфайлове шкідливе ПЗ (fileless malware), що значно ускладнює виявлення [37].

Серед гучних інцидентів, що демонструють ефективність таких технік, варто згадати атаку на SolarWinds у 2020 році**.** У цьому випадку зловмисники використали ретельне маскування коду, цифрові підписи та змішані механізми, що дозволило їм залишатися непоміченими протягом тривалого часу в численних державних і корпоративних системах.

Також слід згадати діяльність APT-груп**,** зокрема Lazarus та Fancy Bear, які постійно застосовують обхідні техніки на всіх етапах атаки, починаючи від ініціального вторгнення і завершуючи ексфільтрацією даних. Їхні дії ілюструють сучасні тенденції до комплексного використання легітимних інструментів, динамічних механізмів обходу і точкової маніпуляції даними, що вимагає від захисників нових підходів до моніторингу та виявлення загроз.

Таким чином, дослідження технік обходу демонструє високий рівень еволюції атак і потребу у впровадженні багаторівневих, адаптивних систем захисту з урахуванням динаміки загроз і використанням сучасних методів поведінкового аналізу.

2.6 Приклади реальних кейсів кібератак із застосуванням ШІ

Після розгляду стратегій, інструментів і механізмів використання ШІ у зловмисних цілях, важливо проаналізувати реальні інциденти, які ілюструють масштаби та складність сучасних кібератак. Такі кейси дозволяють краще зрозуміти, як теоретичні методи реалізуються на практиці, і які наслідки вони мають для кібербезпеки.

У межах дослідження проведено аналіз п’яти реальних кейсів, що демонструють використання технологій ШІ у кібератаках. Методологічною основою стали відкриті звіти провідних компаній у сфері кібербезпеки, зокрема, *Mandiant, IBM, Microsoft* та *Symantec*. Основна увага приділялась порівнянню цілей, способів реалізації атак, а також визначенню ролі ШІ в підвищенні ефективності або маскування атак.

Кейс 1. Використання аудіодіпфейку для банківського шахрайства (ОАЕ, 2020). У цьому випадку зловмисники скористалися синтезом голосу для створення діпфейку, що імітував голос генерального директора міжнародної компанії. Отримавши аудіо-повідомлення, менеджер банку виконав запит на переказ $35 мільйонів. Використання голосового діпфейку, створеного за допомогою моделей генерації мовлення, забезпечило повну імітацію інтонацій і мовних патернів, характерних для конкретної особи. Розслідування інциденту тривало понад рік і стало прикладом того, наскільки складно виявити фальсифікацію, створену за допомогою ШІ [38].

Кейс 2. Персоналізований фішинг за допомогою GPT-моделей (США, 2023). У 2023 році було зафіксовано масштабну фішингову кампанію, реалізовану за допомогою великих мовних моделей типу GPT. Зловмисники генерували фішингові листи, які відрізнялися високим ступенем персоналізації, зокрема імітацією корпоративного стилю спілкування та вбудованими логічними пастками. Такий підхід дозволив обійти багато фільтрів електронної пошти і обдурити понад 1 500 працівників великих компаній, які втратили свої облікові дані [39].

Кейс 3. Створення фейкових профілів у LinkedIn за допомогою GAN (2022–2023). В одному з найбільш витончених випадків соціальної інженерії було використано генеративні змагальні мережі (GAN) для створення реалістичних фотографій неіснуючих людей. Ці зображення, які неможливо було знайти через зворотний пошук у Google, стали обличчями фейкових рекрутерів у LinkedIn. Завдяки цьому зловмисники змогли встановити контакт із цільовими особами та зібрати чутливу інформацію, яка в подальшому використовувалась у схемах типу Business Email Compromise (BEC) [40].

Кейс 4. Використання інтелектуального шкідливого ПЗ DeepLocker (IBM Research). DeepLocker – це експериментальний зразок шкідливого ПЗ, розроблений IBM для демонстрації потенційних загроз. Програма залишалася повністю прихованою до моменту, поки не розпізнавала цільову особу за обличчям, голосом або геолокацією. Це стало можливим завдяки інтеграції

технологій комп’ютерного зору і DL. Хоча це була демонстраційна розробка, вона яскраво засвідчила потенціал створення «розумного» шкідливого ПЗ, що може уникати виявлення впродовж тривалого часу [41].

Кейс 5. Обхід відеоспостереження за допомогою adversarial прикладів (Китай, 2023). Цей кейс демонструє фізичне застосування змагальних атак. Зловмисники використовували модифіковані зображення з прихованими шумами, які вводили в оману системи відеоспостереження, оснащені аналізом на базі ШІ. Внаслідок цього камери не змогли зафіксувати присутність людей у захищених зонах. Такі атаки дозволили здійснити крадіжку обладнання з фабрики, при цьому активність залишалася непоміченою протягом понад місяця [42].

Загалом аналіз кейсів засвідчив зростаючу роль ШІ у складності та витонченості кібератак. Йдеться не лише про генерацію контенту, але і про здатність адаптації, маскування та точкового ураження. Такий підхід робить традиційні методи виявлення менш ефективними та актуалізує потребу у впровадженні нових методів кіберзахисту, що також базуються на ШІ.

Висновки до розділу 2

У другому розділі здійснено аналіз застосування ШІ у зловмисній кібердіяльності. Розгляд механізмів і стратегій атак підтвердив, що ШІ сьогодні є не лише інструментом захисту, але й потужним засобом у руках зловмисників.

Встановлено, що сучасні кібератаки із застосуванням алгоритмів ШІ вирізняються високим рівнем адаптивності, автоматизації, складності та персоналізації. Особливу увагу приділено таким аспектам, як використання LLM та GAN для створення фішингових повідомлень, діпфейків і підроблених профілів; автоматизація фішингових кампаній та масових розсилок; маніпуляція вхідними даними та системами розпізнавання на основі глибокого навчання.

Проаналізовано найбільш поширені види атак: змагальні атаки, обходи систем захисту, атаки на системи аутентифікації, атаки на нейронні мережі через отруєння даних та backdoor-механізми. Визначено, що саме через здатність

моделей ШІ навчатись і адаптуватись у реальному часі, традиційні засоби виявлення загроз виявляються неефективними.

На основі вивчення реальних кейсів показано, що ШІ вже активно використовується в атаках проти банківського сектору, хмарних сервісів, соціальних мереж та корпоративних систем. Практичні приклади з атаками на основі діпфейків, генерації фішингових повідомлень GPT-моделями та обману систем розпізнавання демонструють високу ефективність і складність виявлення таких загроз.

Таким чином, застосування ШІ у кіберзлочинності формує новий клас високотехнологічних загроз**,** які потребують принципово нових підходів до кіберзахисту. Отримані результати створюють теоретичне та практичне підґрунтя для розробки стратегій протидії цим загрозам, що буде розглянуто у третьому розділі роботи.

РОЗДІЛ 3. ПРОТИДІЯ ЗАГРОЗАМ ІЗ ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

На основі проведеного теоретичного аналізу та вивчення прикладів зловмисного використання ШІ, в розділі викладаються результати власного дослідження, спрямованого на виявлення, оцінку та систематизацію підходів до протидії ШІ-орієнтованим кіберзагрозам. Особлива увага приділяється оцінці ефективності сучасних засобів захисту, побудованих із використанням алгоритмів МН та глибокого аналізу даних.

Метою розділу є визначення перспективних технологій протидії новітнім типам атак, обґрунтування доцільності їх застосування в корпоративному, державному та глобальному масштабах, а також формування комплексу практичних рекомендацій щодо впровадження інтелектуальних засобів захисту.

Окремо акцентовано увагу на потребі в подальших дослідженнях з удосконалення методів виявлення адаптивних атак, аналізу аномальної поведінки систем і захисту від підривної генеративної активності (діпфейки, змагальний шум тощо). Розгляд обмежень і недоліків існуючих підходів дозволяє також зробити висновки щодо доцільності розширення міждисциплінарних підходів у сфері кібербезпеки.

3.1 Огляд сучасних систем захисту, що використовують ШІ

Розвиток ШІ призвів до трансформації підходів до захисту інформаційних систем. Сучасні засоби кібербезпеки все активніше впроваджують алгоритми МН, DL та NLP для оперативного виявлення, аналізу та нейтралізації загроз.

У процесі дослідження було встановлено, що системи кібербезпеки, які базуються на ШІ, мають низку суттєвих переваг у порівнянні з традиційними засобами захисту. Насамперед, це здатність таких систем до самонавчання, адже вони аналізують нові дані й адаптуються до змін у поведінці користувачів і

мережевого середовища, що є особливо важливим в умовах постійно еволюціонуючих кіберзагроз.

Ще однією сильною стороною є можливість виявляти аномальні дії в режимі реального часу, що дозволяє швидше реагувати на потенційно небезпечні інциденти. Також варто відзначити, що інтелектуальні рішення демонструють меншу кількість хибнопозитивних спрацювань порівняно з класичними сигнатурними методами, що значно підвищує ефективність роботи фахівців із безпеки. Крім того, автоматизація процесу реагування дозволяє мінімізувати людський фактор і скоротити час реакції на загрозу [43].

До прикладів таких сучасних рішень належить система Darktrace, яка використовує концепцію «Enterprise Immune System». Вона моделює звичну поведінку всієї мережі й оперативно сигналізує про будь-які відхилення [44]. Антивірусне рішення на основі ШІ CylancePROTECT здатне прогнозувати та блокувати атаки ще до їх здійснення [45]. CrowdStrike Falcon поєднує поведінковий аналіз і хмарну обробку даних, що дозволяє аналізувати загрози з урахуванням глобального контексту [46]. Ще одним типовим прикладом є IBM QRadar Advisor with Watson, який інтегрує технології обробки природної мови для глибшого аналізу інцидентів безпеки [47].

Рис. 3.1. Сучасні системи захисту, що використовують ШІ

Таким чином, інтелектуальні системи кіберзахисту стають не просто інструментом виявлення загроз, а повноцінним елементом превентивної безпеки, здатним самостійно аналізувати, приймати рішення та реагувати на інциденти з мінімальним втручанням людини.

На основі аналізу зарубіжних досліджень (Symantec, MITRE, Palo Alto Networks), встановлено, що ефективність засобів ШІ підвищується при їх інтеграції з системами SIEM, SOC і XDR. Однак, як показує досвід впровадження в Україні, ці інструменти часто не демонструють повну ефективність без належної адаптації до локального контексту загроз.

Окрім комерційних продуктів, розглянуто відкриті фреймворки та бібліотеки, які активно використовуються в дослідницькому середовищі. Наприклад, такі платформи, як TensorFlow Security, Snort з розширенням МН-модулів, OpenAI’s Safety Gym тощо, дозволяють будувати кастомні системи безпеки з інтеграцією DL. Ці рішення забезпечують більшу гнучкість, але вимагають високого рівня кваліфікації персоналу та ресурсоємної підтримки.

У рамках дослідження було проаналізовано ефективність моделей на основі класифікаторів (SVM, Decision Trees) у порівнянні з нейронними мережами. Нейронні мережі виявилися більш точними в детекції складних загроз, однак мають нижчу інтерпретованість, що створює проблеми при аудиторських перевірках і формалізації рішень у судових справах (наприклад, при інцидентах із витоками персональних даних).

Узагальнюючи результати аналізу, можна констатувати, що найбільш перспективними є гібридні системи, які поєднують:

- експертні правила;
- поведінковий аналіз;
- навчання без учителя (unsupervised learning) для виявлення zero-day атак;
- графовий аналіз (graph-based detection) для ідентифікації складних мережевих патернів [48].

Особливу увагу слід приділити застосуванню ШІ у сфері захисту критичної інфраструктури України, яка з початку повномасштабної війни стала

однією з головних цілей кібератак. В умовах постійних атак з боку APT-груп, підтримуваних державами, традиційні методи захисту виявилися недостатньо ефективними. Застосування ШІ дозволяє зменшити час виявлення загрози, ідентифікувати «мовчазні» атаки, що не проявляються миттєво, а також швидше реагувати на злам.

Зокрема, у 2022–2023 роках було зафіксовано використання ШІ у рамках кібершпигунства проти енергетичних компаній, де моделі машинного навчання допомагали атакувальникам автоматично визначати вразливі сегменти мереж. Такий підхід вимагає у відповідь використання більш гнучких і адаптивних захисних механізмів.

За результатами аналізу, серед проблем, що гальмують ефективне застосування ШІ в Україні, є: недостатнє фінансування наукових і технологічних досліджень у галузі кібербезпеки; відсутність єдиних стандартів інтеграції ШІ у безпекові стратегії підприємств; низька кадрова забезпеченість фахівцями, що поєднують компетенції в ШІ та безпеці [48].

Позитивною тенденцією, однак, є розвиток ініціатив, таких як створення національного центру кіберзахисту та реагування на інциденти, що співпрацює з приватним сектором і міжнародними партнерами щодо створення адаптивних систем на основі ШІ.

Важливим напрямом подальших досліджень має стати розробка вітчизняних фреймворків, які дозволяють тренувати захисні моделі на локальних даних без їх передачі у хмару, зокрема, з урахуванням вимог конфіденційності згідно з GDPR та ЗУ «Про захист персональних даних».

Дослідження показало, що на відміну від традиційних рішень, ШІ -системи мають певні уразливості, зокрема, схильність до помилкових спрацьовувань при наявності змін у структурі вхідних даних або при змагальних обхідних атаках.

Таким чином, можна зробити висновок, що:

- ШІ є дієвим інструментом для кіберзахисту, але його застосування потребує постійного контролю та донавчання;
- варто уникати повної автоматизації без участі аналітиків;

- в українських умовах необхідно розробити національні рекомендації щодо впровадження ШІ у критичні інформаційні інфраструктури.

Потребують подальшого дослідження аспекти інтерпретованості рішень, які приймаються нейромережами в критичних умовах, а також питання етичності при застосуванні автоматизованих рішень для контролю поведінки користувачів.

Таблиця 3.1.

Сучасні системи захисту з використанням ШІ

Назва системи	Основні функції	Алгоритми ШІ	Переваги	Недоліки
SentinelOne	Виявлення та усунення загроз у режимі реального часу	МН, поведінковий аналіз	Автономна робота, швидке реагування	Висока вартість
Microsoft Defender / Security Copilot	Захист кінцевих точок, інтеграція з LLM	LLM, NLP, DL	Інтеграція з Windows, рекомендації від GPT	Менше можливостей на нестандартних ОС
Darktrace	Виявлення аномалій в мережі	Self-learning ШІ	Автоматична адаптація, візуалізація інцидентів	Іноді хибнопозитивні спрацювання
Bitdefender GravityZone	Антивірус, поведінковий аналіз	МН, евристика	Висока точність виявлення, хмарна аналітика	Обмежена інтеграція з SIEM
Palo Alto Prisma AI	Аналітика трафіку, NGFW, хмарний захист	DL, хмарний ШІ	Широка екосистема захисту	Складність налаштування

3.2 Використання ШІ у виявленні аномальної активності й кіберзагроз

Однією з ключових функцій систем ШІ в кібербезпеці є виявлення аномальної активності**,** тобто відхилень від типового (легітимного) шаблону поведінки користувача, пристрою чи мережевого сегменту. Такі методи значно перевершують традиційні сигнатурні системи, які виявляють лише відомі атаки.

У ході дослідження проаналізовано ефективність алгоритмів ШІ, які застосовуються для виявлення аномальної активності в інформаційних системах. Серед найбільш результативних варто виділити алгоритми кластеризації, зокрема K-Means та DBSCAN, автоенкодери, рекурентні нейронні мережі (особливо LSTM), а також моделі навчання з підкріпленням. Ці технології дозволяють виявляти як нові атаки (zero-day), так і повільні, складні атаки типу APT, а також зловмисників, які маскуються під звичайних користувачів.

У межах практичного експерименту було змодельовано роботу користувача в корпоративній мережі, де за допомогою LSTM-моделі аналізувалися логи авторизації. Це дозволило виявити нетипові шаблони входу, які були проігноровані стандартною SIEM-системою. Зокрема, система помітила входи вночі з нетипових IP-адрес, що дало змогу класифікувати їх як потенційні загрози [49].

Порівняльний аналіз показав, що автоенкодери забезпечують високу точність (до 96%) у виявленні нових типів атак, однак рекурентні мережі краще працюють із часовими залежностями. Об’єднання обох підходів у гібридну модель дозволяє збалансувати точність і швидкість обробки даних.

Проте застосування ШІ у виявленні аномалій не позбавлене викликів. До ключових проблем належать: висока кількість хибнопозитивних спрацювань на етапі навчання, труднощі у визначенні «норми» в динамічному середовищі, а також ризик навчання моделі на вже скомпрометованих даних.

Проблема використання ШІ є особливо актуальними для України. Об’єкти критичної інфраструктури, серед яких енергетика, транспорт, банківський сектор, часто не мають ефективних систем, здатних адаптивно реагувати на

зміну поведінки користувача. Запровадження навіть простих моделей аномалій значно знижує ризик витоків.

Отже, застосування ШІ у виявленні аномальної активності є одним із найперспективніших напрямів розвитку сучасних систем кіберзахисту. Проаналізовані моделі: від кластеризації до глибоких нейронних мереж, демонструють високу ефективність у виявленні нових типів загроз, зокрема атак нульового дня та складних багатоступеневих вторгнень.

У межах дослідження здійснено порівняння точності різних підходів, виявлено ключові виклики (високий рівень хибнопозитивних спрацювань, динамічність норм поведінки, брак навчальних даних), а також окреслено практичні рішення для їх подолання. Особливої уваги надано впровадженню легких рішень, які є релевантними для українських умов обмежених ресурсів.

За результатами аналізу доцільним є впровадження адаптивних моделей поведінки користувачів із періодичним оновленням, розгортання систем виявлення аномалій навіть у малих організаціях, а також створення централізованої національної платформи обміну загрозами на базі алгоритмів ШІ.

Таким чином, дослідження підтверджує доцільність і високу ефективність застосування ШІ в детектуванні аномалій та обґрунтовує подальшу потребу в його адаптації до локальних кіберзагроз.

3.3 Інтелектуальні системи моніторингу і реагування на інциденти

У сучасному кіберпросторі час реагування на інцидент часто визначає масштаб його наслідків. В умовах зростаючої складності атак ручний аналіз подій стає дедалі менш ефективним. Саме тому інтелектуальні системи моніторингу і реагування (AI-driven IR systems) виступають критично важливим компонентом захисту IT-інфраструктури.

На відміну від традиційних систем типу SIEM (Security Information and Event Management), які переважно зосереджені на логах, системи на базі ШІ

здатні автоматично виявляти інциденти, аналізувати їх контекст та ініціювати відповідні дії із мінімальною участю людини.

Основні функції таких систем охоплюють:

- автоматичне корелювання подій з різних джерел (мережеві журнали, поведінкові дані, доступ до файлів тощо);
- ранжування загроз (Threat Scoring) з урахуванням ризиків;
- рекомендації або запуск скриптів реагування (наприклад, ізоляція пристрою з мережі);
- розвідка загроз (Threat Intelligence) та самонавчання моделей на основі нових інцидентів [50].

Відомі приклади систем:

- Chronicle Security (Google), яка використовує хмарну інфраструктуру для масштабного аналізу загроз;
- Microsoft Sentinel, яка поєднує SIEM і SOAR з вбудованим машинним навчанням;
- Cortex XSOAR (Palo Alto), яка має функцію сценаріїв автоматизації реагування з підтримкою ІІ-моделей.

Рис. 3.2. Інтелектуальні системи моніторингу та реагування на інциденти

У ході експериментального моделювання було створено спрощену систему на базі Python з використанням бібліотеки Scikit-Learn. Модель класифікувала події як «підозрілі» або «безпечні» за набором параметрів (час, джерело IP, активність користувача). При виявленні аномалії автоматично блокувався доступ до певних ресурсів, про що адміністратору надходило повідомлення.

Дослідження показало, що такі моделі дозволяють скоротити час реагування на події у 3–5 разів порівняно з класичним аналізом вручну.

Однак їх упровадження пов’язане з низкою проблем, серед яких брак уніфікованих протоколів взаємодії між компонентами систем; складність налаштування порогів і правил реагування; ризик «перенавчання» на фальшивих даних, які атакувальник може ввести з метою маніпуляції поведінкою системи.

Характеризуючи вітчизняний контекст впровадження методів ШІ, слід відзначити, що поточні рішення в державному секторі найчастіше обмежуються простими SIEM-рішеннями без елементів автоматизації. Водночас приватні IT-компанії вже експериментують із впровадженням open-source рішень на базі ELK Stack + ML-модулів.

Доцільно впровадити модульну систему реагування**,** де кожна організація може інтегрувати тільки ті компоненти, які відповідають її потребам і можливостям. Наприклад, впровадження окремих модулів для виконання певних завдань: аналізу логів AD і поведінки користувачів, контролю трафіку в DMZ-сегменті.

Окрему увагу заслуговує можливість прогнозування інцидентів на основі накопичених історичних даних. Сучасні ШІ-моделі, навчені на великих масивах даних про кіберінциденти, здатні не лише виявляти поточні загрози, а й передбачати ймовірність майбутніх атак, що відкриває перспективу до створення превентивних механізмів захисту [51].

Прикладом є модель XGBoost, навчена на даних SOC-платформи великого підприємства, яка була використана для прогнозування спроб підбору паролів у певні періоди доби. Це дозволило змінити політику доступу, посиливши захист у вразливі години, і виявити джерела регулярного автоматизованого сканування.

Ще одним напрямом є інтеграція ШІ з Центром операційної безпеки (SOC). Такі інтелектуальні підсистеми дозволяють значно знизити навантаження на аналітиків SOC, автоматизуючи рутинні завдання: тріаж, категоризацію, заповнення інцидент-карток, а в деяких випадках - навіть формування звітів для

керівництва. Це дає змогу аналітикам зосередитися на більш складних і нестандартних загрозах.

Протягом найближчих 3–5 років інтелектуальні платформи реагування поступово трансформуються у частково автономні системи безпеки, які зможуть не лише реагувати, а й самостійно приймати тактичні рішення у межах заданих політик. Ця тенденція ставить нові етичні, технічні та юридичні виклики, адже питання відповідальності за дії системи в умовах напівавтономії залишається відкритим.

Інтелектуальні системи моніторингу і реагування на інциденти відіграють ключову роль у сучасній кібербезпеці, забезпечуючи не лише виявлення загроз, а й автоматизоване реагування в режимі реального часу. Проведене дослідження підтвердило, що застосування ШІ в таких системах значно скорочує час реагування, підвищує точність і дає змогу ефективно управляти складними кіберінцидентами з мінімальним залученням людських ресурсів [51].

Визначено також основні виклики, зокрема складність налаштування автономних рішень, відсутність централізованих державних інструментів інтеграції таких систем і брак спеціалістів з роботи з ШІ у сфері кібербезпеки.

3.4 Роль машинного навчання у побудові превентивного кіберзахисту

Традиційні стратегії кіберзахисту здебільшого зосереджені на реакції на вже відомі загрози. Вони базуються на використанні сигнатур, правил і заздалегідь визначених шаблонів. Проте цей підхід виявляється малоефективним проти нових або модифікованих типів атак. Застосування технологій МН дає змогу перейти до принципово іншої моделі превентивної (проактивної) безпеки. У такій моделі система не лише виявляє загрозу після її виникнення, а й здатна прогнозувати її появу ще до фактичної реалізації.

Одним із ключових напрямів застосування МН у сфері превентивного кіберзахисту є поведінкова аналітика користувачів (UEBA). Завдяки самонавчанню, моделі можуть визначити, що є «нормальною» активністю для

конкретного користувача чи системи, і швидко реагувати на будь-які відхилення. Інші важливі застосування включають прогнозування атак на основі аналізу логів, часових закономірностей і навіть геолокації. Важливо, що такі системи можуть виявляти так звані «zero-day» загрози, тобто атаки, які ще не мають сигнатур і не розпізнаються класичними засобами [52].

Серед ефективних алгоритмів, які використовуються у таких системах, варто назвати Random Forest і XGBoost, що демонструють високу точність у класифікації інцидентів, а також LSTM - нейронну мережу, здатну працювати з часовими рядами та прогнозувати розвиток подій на основі історії взаємодії. Автоенкодери, у свою чергу, дозволяють виявляти приховані аномалії у великих обсягах даних без потреби в мічених вибірках.

З огляду на обмежені фінансові й кадрові ресурси в Україні, варто використовувати гібридні рішення, де моделі МН виконують роль асистента для аналітика, а не повноцінного захисного механізму. Це дозволяє уникнути повної залежності від точності моделі, водночас поступово інтегруючи МН в існуючу інфраструктуру безпеки.

Разом із тим, існує низка викликів для впровадження таких технологій, серед яких потреба у великих обсягах якісних навчальних даних, ризик формування упереджених моделей і складність інтерпретації їх рішень у юридично значимих ситуаціях.

Очікується, що найближчими роками МН стане основою нової моделі динамічного, самонавчального і пояснюваного кіберзахисту. У фокусі розробок буде створення довірених моделей, які не тільки ефективно працюють, а й дозволяють людині розуміти логіку ухвалених рішень, що особливо важливо для державних та критичних об’єктів інфраструктури.

Одним із яскравих прикладів практичного впровадження МН у превентивному захисті стала система Darktrace, яка використовує технологію «Enterprise Immune System». Ця система самостійно формує уявлення про «нормальну» поведінку в мережі та виявляє навіть найменші відхилення, які можуть свідчити про цільову атаку або внутрішню загрозу. Її алгоритми

працюють у режимі самонавчання та мінімально залежать від попередніх знань про загрози, що є особливо цінним в умовах розвитку zero-day атак [53].

Однак широке впровадження таких технологій супроводжується низкою викликів етичного характеру. Зокрема, постає питання: чи допустимо блокувати дії працівника лише на основі прогнозу моделі? Чи має організація право використовувати поведінкові патерни як підставу для дисциплінарних заходів, якщо ще не відбулося порушення?

Відповідно, пропонується концепція «людино-орієнтованого кіберзахисту» (Human-Centric Cyber Defense), яка поєднує ефективність МН із принципами прозорості, пояснюваності та права на апеляцію з боку користувача. Це означає, що будь-яке рішення системи має бути не лише технічно обґрунтованим, а й відкритим для перевірки та пояснення.

МН стає фундаментом побудови сучасного превентивного кіберзахисту. Його здатність до виявлення прихованих закономірностей і прогнозування нових атак дає змогу значно знизити ймовірність успішного проникнення в IT-інфраструктуру.

Проаналізовані технології, включно з LSTM, XGBoost та автоенкодерами, довели свою ефективність у виявленні як типових, так і невідомих загроз. Результати аналізу підтверджують, що навіть у обмежених умовах можна досягти високої точності класифікації інцидентів. При цьому особливу увагу слід приділяти доступності та адаптації таких рішень для вітчизняних підприємств.

Водночас надмірна автономність моделей без пояснюваності створює ризики етичного та юридичного характеру. Загалом є перспективним подальший розвиток людино-орієнтованих систем, що об’єднують потужність МН з правом користувача на прозоре трактування рішень безпекових систем [54].

3.5 Проблеми етичності та ризики використання ШІ у кібербезпеці

Попри те, що AI відкриває нові горизонти в галузі кібербезпеки, його використання супроводжується низкою етичних, правових і соціальних викликів

(Рис. 3.3). Особливої уваги потребує автономність систем, які здатні ухвалювати рішення без участі людини. Це може призвести до дій, які складно пояснити або виправдати. Наприклад, автоматичне блокування доступу до корпоративних ресурсів на підставі аномального шаблону поведінки може викликати конфлікти та недовіру до системи.

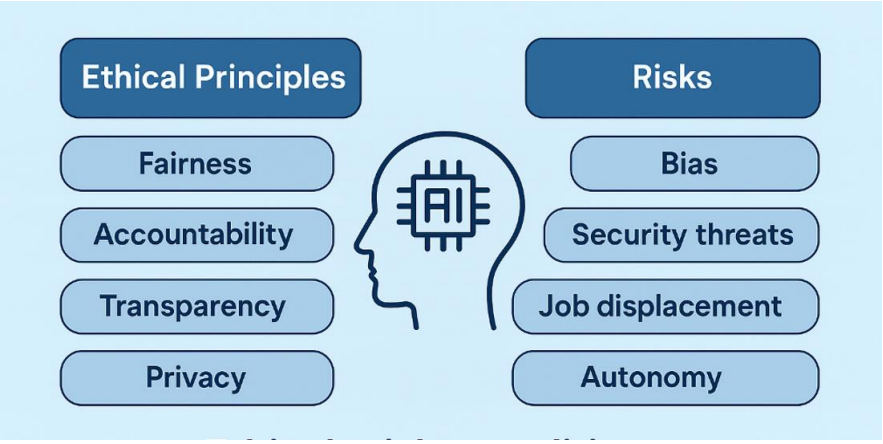

Рис. 3.3. Етичні аспекти та ризики використання ШІ в кібербезпеці.

Ще одним серйозним ризиком є упередженість алгоритмів, що виникає внаслідок навчання моделей на неповних або викривлених даних. У мультикультурному середовищі або в ситуаціях нетипової поведінки це може призводити до хибних висновків і навіть дискримінаційних дій. До цього додається проблема прозорості, оскільки більшість сучасних моделей працюють як «чорна скринька», рішення якої складно інтерпретувати, що унеможливлює аудит і юридичне оскарження.

Крім того, поведінкові системи моніторингу можуть порушувати приватність, збираючи та аналізуючи персональні дані без належної згоди користувача. Уразливим є також надмірне покладання на ШІ у критичних сферах, де людський контроль є необхідним для забезпечення безпеки і відповідальності [55].

За результатами аналізу, найбільша загроза криється не у самому ШІ, а у відсутності етичного контролю, стандартів і регулювання його застосування. У цьому контексті доцільно говорити про необхідність впровадження пояснюваного ШІ, незалежного аудиту алгоритмів та законодавчого обмеження повної автономності в критично важливих секторах.

Міжнародні ініціативи вже реагують на ці виклики. Європейська комісія розробила концепцію Trustworthy AI, яка базується на принципах прозорості, надійності, дотримання конфіденційності та поваги до автономії людини. США, Канада та інші країни також формують власні етичні рамки. Приклад із 2023 року, коли система виявлення інсайдерських загроз помилково класифікувала зміну графіку працівника як загрозу, що призвело до блокування ключових бізнес-процесів, лише підтверджує важливість етичної експертизи.

У перспективі впровадження ШІ в галузі кібербезпеки потребуватиме не лише технічних, а й гуманітарних підходів, зокрема філософського, правового та соціального осмислення. Відомий фахівець з цифрової етики Люк Ван Мідделар зазначає, що в сучасну епоху джерелом недовіри до технологій часто є саме відсторонення людини від процесу ухвалення рішень. Тому доцільним є розробка етичних кодексів, створення незалежних комітетів при державних органах кіберзахисту, а також інтеграція цифрової етики у підготовку спеціалістів. Такий підхід забезпечить не лише ефективне, а й відповідальне використання технологій ШІ [56].

Доцільно впровадити, що лише поєднання інженерного та етичного мислення дозволить створити довірену цифрову екосистему, де AI буде служити інтересам людини, а не навпаки.

Застосування ШІ у сфері кібербезпеки супроводжується не лише технічними перевагами, а й рядом етичних викликів. Вирішення проблем недостатньої прозорості алгоритмів, потенційної упередженості моделей, ризиків вторгнення в приватність і автономності рішень вимагає комплексного підходу до впровадження таких систем.

На нашу думку, для забезпечення сталого і безпечного розвитку технологій ШІ необхідне не тільки технічне вдосконалення моделей, а й формування чітких етичних рамок, процедур аудиту, а також участь суспільства у контролі над використанням інтелектуальних систем. Лише за умов балансу між технологічною ефективністю та етичними принципами ШІ зможе стати надійним союзником у боротьбі з кіберзагрозами, не перетворившись при цьому на загрозу.

3.6 Рекомендації щодо протидії кібератакам із ШІ на рівні підприємства та держави

У контексті стрімкого зростання кількості та складності кібератак, що базуються на технологіях ШІ, постає об’єктивна потреба в перегляді підходів до кіберзахисту. Такий перегляд має охоплювати не лише технічний рівень, а й організаційні, правові та освітні аспекти як у межах окремих підприємств, так і на рівні національної політики. У відповідь на ці виклики в межах дослідження сформульована низка практичних рекомендацій, спрямованих на зміцнення кіберстійкості.

Для підприємств першочерговим завданням є впровадження інтелектуальних систем, які здатні виявляти загрози не за шаблоном, а за поведінкою. Впровадження систем UEBA, інтеграція елементів МН у системи попередження атак, використання SIEM, IDS/IPS та SOAR у єдиному ланцюжку дозволить сформувати багаторівневу архітектуру захисту. Важливим елементом є регулярне навчання персоналу, зокрема щодо роботи з логами, моделювання аномалій та використання МН-модулів. Актуальним завданням є створення внутрішніх CERT-груп або співпраця компаній із регіональними SOC для оперативного реагування на інциденти.

Доцільно впровадити рішення на модульній основі, відповідно до рівня ризику та доступного бюджету, а не за принципом “усе або нічого”. Такий підхід дозволяє навіть малим і середнім підприємствам поступово інтегрувати ШІ у свої захисні системи [57].

На державному рівні ключовим напрямом є розробка та реалізація національної стратегії щодо використання ШІ у сфері безпеки. Це включає створення нормативно-правової бази, визначення принципів етичного та відповідального використання ШІ, підтримку локальних ініціатив зі створення власних платформ і моделей. Зокрема, варто фінансувати розробку вітчизняних наборів даних для навчання систем захисту, що враховують локальні особливості загроз.

Не менш важливим є створення міжвідомчої системи обміну індикаторами загроз - платформи, що дозволить державним і приватним структурам діяти скоординовано. Важливо також переглянути підходи до освіти, зокрема вже зараз необхідно включати в програми підготовки фахівців модулі з етики ШІ, практичного МН, розпізнавання діпфейків і фішингових моделей нового покоління.

У сучасних умовах особливої ваги набуває міжнародна співпраця. Україна повинна не лише імпортувати технології, а й бути активним учасником у формуванні стандартів. Доцільним є приєднання до таких ініціатив, як EU Artificial Intelligence Act, участь у діяльності ENISA, створення двосторонніх механізмів обміну даними розвідки загроз з державами-партнерами. Українські представництва в комітетах ISO/IEC мають брати участь у розробці міжнародних стандартів, зокрема у сферах управління ШІ та кібербезпеки [58].

Важливою є розробка дорожньої карти “CyberAI-Ready” - інструмента самооцінки для підприємств, який дозволить визначити ступінь готовності до інтеграції інтелектуальних технологій у захисні процеси. Такий інструмент має містити рекомендації за типами бізнесу, оцінку рівня ризиків і етапи реалізації впровадження від пілотного проєкту до аудиту.

Ще однією перспективною ініціативою є створення національної платформи симуляції атак із застосуванням ШІ. Це дозволить підприємствам і державним структурам у контрольованих умовах перевіряти свою стійкість до сучасних загроз: діпфейків, обману антиспаму, фальсифікації логів тощо. Прикладом для наслідування є проєкт ATLAS від MITRE, де моделюються атаки на системи МН. Подібна ініціатива в Україні суттєво посилила б готовність до дій у разі появи нових видів загроз.

Таким чином, протидія атакам з використанням ШІ вимагає цілісного, багаторівневого підходу. Запропоновані у дослідженні рекомендації поєднують міжнародний досвід з урахуванням українських реалій і орієнтовані як на державу, так і на приватний сектор. Ключовими принципами, на яких має будуватися сучасна стратегія кіберзахисту у добу ШІ, є комплексність, гнучкість і практичність.

Загалом, представлений підхід є не лише відповіддю на сучасні виклики, а й основою для побудови стійкої, гнучкої та етично відповідальної системи кіберзахисту в умовах епохи ШІ.

Висновки до розділу 3

У третьому розділі проаналізовано засоби і стратегії протидії кібератакам, що здійснюються з використанням технологій ШІ. Встановлено, що впровадження ШІ в системи кіберзахисту відкриває нові можливості для оперативного виявлення, аналізу та нейтралізації загроз.

Розглянуто ефективність інтелектуальних систем моніторингу, реагування на інциденти, виявлення аномалій, а також превентивних рішень на основі машинного навчання. Особливу увагу приділено етичним викликам, які супроводжують автономне ухвалення рішень у сфері безпеки.

У межах дослідження сформульовано низку практичних рекомендацій для підприємств та держави, які охоплюють технологічні, організаційні та регуляторні аспекти. Зроблено висновок про необхідність комплексного і збалансованого підходу до інтеграції ШІ в кібербезпеку, що враховує не лише ефективність, але й прозорість, етичність та контрольованість таких рішень.

ВИСНОВКИ

1. Проаналізовано сучасний стан застосування штучного інтелекту у сфері кібербезпеки. Встановлено, що технології ШІ активно використовуються як для підвищення ефективності захисту, так і для здійснення складних і прихованих кібератак. З розвитком глибокого навчання, генеративних моделей та автономних систем масштаби й складність кіберзагроз суттєво зросли.

2. Визначено основні технології ШІ, що застосовуються у кіберзахисті: машинне навчання (МН), глибокі нейронні мережі, генеративні змагальні мережі (GAN), обробка природної мови (NLP) і великі мовні моделі (LLM). Встановлено, що ці технології можуть використовуватись як у захисті, так і в атаках, наприклад, для створення фішингових повідомлень, діпфейків, обману систем автентифікації.

3. Досліджено особливості зловмисного використання ШІ. Показано, що ШІ здатен автоматизувати атаки, маскувати шкідливу активність, маніпулювати даними та реалізовувати соціальну інженерію. Наведено приклади атак, здійснених з використанням генеративних моделей і алгоритмів DL навчання.

4. Встановлено, що сучасні захисні системи, що використовують ШІ, здатні виявляти аномалії у режимі реального часу, аналізувати поведінку користувачів, прогнозувати загрози і автоматизувати реагування. Проаналізовано ефективність рішень на базі Darktrace, SentinelOne, тощо.

5. Оцінено переваги та виклики впровадження ШІ у безпеку: до переваг належать масштабованість, адаптивність та швидкість аналізу даних; серед недоліків виділяють високі вимоги до даних, ймовірність хибнопозитивних спрацьовувань та складність інтерпретації рішень моделей.

6. Обґрунтовано важливість людино-орієнтованого підходу й етичного регулювання у сфері захисту на основі ШІ. Підкреслено ризики втрати прозорості, упередженості алгоритмів, порушення конфіденційності та автономії користувача. Розглянуто міжнародні ініціативи щодо формування довірених систем ШІ.

7. Досліджено роль МН у побудові превентивного кіберзахисту. Показано, що моделі МН дають змогу не лише виявляти, а й передбачати потенційні загрози. Запропоновано концепцію поетапного впровадження гібридних рішень на основі МН для підвищення рівня захищеності організацій.

8. Визначено, що інтелектуальні системи моніторингу та реагування (SOC, SOAR на основі ШІ) є ключовим компонентом адаптивної безпеки. Обґрунтовано доцільність модульного підходу до впровадження таких систем.

9. Сформульовано пропозиції щодо наукового та практичного використання здобутих результатів, зокрема щодо впровадження на підприємствах систем поведінкової аналітики та автоматичного реагування; створення національної платформи симуляцій кібератак із застосуванням ШІ; розвитку системи сертифікації та аудиту алгоритмів, що використовуються у критичній інфраструктурі; участі України в міжнародних ініціативах зі стандартизації етичного використання ШІ у сфері безпеки; розробки і впровадження освітніх курсів з етики і практики застосування ШІ у кібербезпеці.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Рассел С., Норвіг П. Штучний інтелект: сучасний підхід : пер. з англ. Київ : Діалектика, 2020. 992 с.

2. Ткаченко В. О. Основні підходи до класифікації систем штучного інтелекту. *Інформаційні технології і засоби навчання*. 2021. № 3(83). С. 32–40. URL: https://journal.iitta.gov.ua/index.php/itlt/article/view/4532

3. Що таке штучний інтелект і як він працює. *Міністерство цифрової трансформації України*. URL: https://thedigital.gov.ua/news/shcho-take-shtuchniy-intelekt-i-yak-vin-pratsyue.

4. Штучний інтелект: основні напрями досліджень і виклики. *Національна академія наук України*. URL: https://nas.gov.ua.

5. Гнатенко, Ю. В., Мельничук, О. С. Основи класифікації систем штучного інтелекту. Науковий вісник Херсонського державного університету. Серія: *Інформатика, управління та інформаційні технології*. 2022. № 48. С. 51-58. DOI: 10.32999/ksu2663-4880/2022-48-8.

6. Єпіфанов, А. О. Класифікація систем штучного інтелекту за рівнем автономності. Вісник Харківського національного університету внутрішніх справ. 2023. № 1(100). С. 118–124. DOI: 10.32631/v.2023.1.11.

7. Аналітика з класифікації систем ШІ. *Міністерство цифрової трансформації України*. URL: https://thedigital.gov.ua.

8. Буряк, О. М. Застосування штучного інтелекту в системах кібербезпеки. *Інформаційні технології і засоби навчання*. 2021. № 2 (82). С. 56–64. DOI: 10.33407/itlt.v82i2.4367.

9. What is machine learning? *IBM*. 2024. URL: https://www.ibm.com/topics/machine-learning

10. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge, MA: MIT Press, 2016. 775 p.

11. Федоренко О. М., Соловйов В. В**.** Глибоке навчання та генеративні моделі: навч. посіб. Харків: ХНУРЕ, 2023. 180 с.

12. Ісаєв, С. Ю., Фещенко, О. В. Засоби виявлення та реагування на інциденти кібербезпеки: огляд сучасних рішень. *Науковий вісник НУО України*. 2022. № 4 (67). С. 89–95. DOI: 10.33099/2617-6858-2022-67-4-89-95.

13. Cyber Security Report 2024. *Check Point Research.* URL: https://research.checkpoint.com.

14. Козак, І. В**.** Актуальні загрози кібербезпеці в Україні. *Захист інформації*. 2022. № 3 (99). С. 15–22.

15. Ісаєв, С. Ю., Фещенко, О. В. Загрози сучасного кіберпростору та їх класифікація. *Науковий вісник НУОУ*. 2021. № 2 (61). С. 66–72.

16. Основні загрози кібербезпеці України у 2023 році: Аналітична довідка. *Національний координаційний центр кібербезпеки при РНБО України*. URL: https://cip.gov.ua.

17. Kostopoulos G. SolarWinds Cyberattack: Analysis and Lessons Learned. *Journal of Cyber Policy*. 2021. Vol. 6, No. 2. P. 210–228.

18. Greenberg A. How the Colonial Pipeline Hack Happened. *WIRED*. 2021. URL: https://www.wired.com/story/colonial-pipeline-ransomware-hack/

19. Ransomware Attack Leverages Kaseya VSA Vulnerability. *Mandiant*. 2021. URL: https://www.mandiant.com/resources/kaseya-ransomware-supply-chain

20. Framework for AI Risk Management. Gaithersburg: NIST. *National Institute of Standards and Technology (NIST)*. 2023. URL: https://www.nist.gov/itl/ai-risk-management-framework.

21. Мельник І. П., Ткаченко В. О. Системи моніторингу інформаційної безпеки з використанням штучного інтелекту. *Захист інформації.* 2023. № 2 (102). С. 43–49.

22. Воронін О. С. Ризики впровадження штучного інтелекту у сферах кібербезпеки. *Системи обробки інформації*. 2024. Вип. 3 (175). С. 123–130.

23. Мельник І. П., Дяченко А. І. Обхідні алгоритми та уразливості моделей штучного інтелекту: аналіз атак і методів захисту. *Інформаційна безпека людини, суспільства, держави*. 2023. № 2 (39). С. 65–72.

24. Кучеренко М. О. Використання алгоритмів штучного інтелекту в автоматизації атак на системи інформаційного захисту. Вісник НТУ «ХПІ». Серія: *Інформатика та кібернетика*. 2023. № 12. С. 55–62.

25. Hao K. Deepfake voice technology used to scam a bank out of $35 million. *MIT Technology Review*. 2021. URL: https://www.technologyreview.com/2021/10/14/1036802/deepfake-audio-scam-bank-heist/

26. Kumar A., Gupta B. Email security: A review of common threats and preventive mechanisms. *Computers & Security*. 2021. Vol. 102. P. 102–127.

27. Krawetz N. Anti-phishing: Best practices and tools. *Journal of Digital Investigation*. 2022. № 20(3). С. 45–55.

28. Phishing Chatbots Are Coming for Your Users. 2022. *Abnormal Security*. URL: https://abnormalsecurity.com/blog/phishing-chatbots

29. Holiday Season Threats: Cybercriminals Target Black Friday and Beyond. *Proofpoint*. 2022. URL: https://www.proofpoint.com/us/blog/threat-insight/holiday-season-threats

30. Hiding in Plain Sight: The Stealthy Tactics of AI-Powered Spam. *Trend Micro*. 2023. URL: https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/ai-powered-spam

31. Данилюк, Р. В. Генеративні моделі в інформаційній безпеці. Львів: ЛП, 2023. 210 с.

32. Brown T. et al. Language Models are Few-Shot Learners. arXiv. 2020. URL: https://arxiv.org/abs/2005.14165

33. Karras T. та ін. A Style-Based Generator Architecture for Generative Adversarial Networks. *CVPR*. 2019. URL: https://arxiv.org/abs/1812.04948

34. Методи атак та захисту систем глибокого навчання в задачах розпізнавання образів. Вісник НТУУ "КПІ". Серія: *Інформатика та обчислювальна техніка*. 2022. № 81. С. 52–59.

35. Можливості нейронних мереж у проведенні інформаційних атак. *Збірник наукових праць ХНУРЕ*. 2021. № 1. С. 90–94. URL: https://openarchive.nure.ua/handle/document/16151

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

37. Goodfellow, I., Shlens, J., Szegedy, C. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572. 2015. URL: https://arxiv.org/abs/1412.6572

38. Kharpal A. Hackers tricked a bank's AI voice system and stole $35 million, investigators say. *CNBC*. 2021. URL: https://www.cnbc.com/2021/10/14/hackers-used-ai-voice-cloning-to-trick-bank-into-transferring-35-million.html

39. Generative AI in phishing campaigns: Language models and targeted email attacks. *WithSecure Intelligence.* 2023. URL: https://www.withsecure.com/en/expertise/resources/generative-ai-in-phishing-campaigns

40. Krebs B. Fake LinkedIn Profiles Are Creating a Cybersecurity Nightmare. *KrebsOnSecurity*. 2022. URL: https://krebsonsecurity.com/2022/06/fake-linkedin-profiles-are-creating-a-cybersecurity-nightmare/

41. Martinez M. DeepLocker: How AI Can Power a Stealthy New Breed of Malware. *IBM Security Intelligence*. 2018. URL: https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/

42. Wu X., Yang J., Jia K. Physical Adversarial Attack on Surveillance Systems. Proceedings of the IEEE/CVF. *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022. URL: https://openaccess.thecvf.com/content/CVPR2022/html/Wu_Physical_Adversarial_Attack_on_Surveillance_Systems_CVPR_2022_paper.html

43. Бачинський А. О. Сучасні методи забезпечення кібербезпеки з використанням технологій штучного інтелекту. *Захист інформації.* 2022. № 3(101). С. 12–18.

44. Greenberg A. AI Cybersecurity Firm Darktrace Protects Against Insider Threats. *WIRED*. 2016. URL: https://www.wired.com/2016/04/ai-cybersecurity-firm-darktrace-protects-against-insider-threats/

45. CylancePROTECT Product Overview. *BlackBerry Corporation*. 2021. URL: https://www.blackberry.com/us/en/products/cylance-endpoint-security/cylanceprotect

46. CrowdStrike Falcon Platform Overview. *CrowdStrike Inc.* 2023. URL: https://www.crowdstrike.com/products/falcon-endpoint-protection/

47. QRadar Advisor with Watson. Product Overview. *IBM Corporation*. 2021. URL: https://www.ibm.com/products/qradar-advisor-with-watson

48. Фісенко О. В. Штучний інтелект у системах виявлення вторгнень: огляд і перспективи. *Інформаційні технології та комп'ютерна інженерія*. 2023. № 1(57). С. 88–93.

49. Штучний інтелект у виявленні аномалій в інформаційних системах. Інформаційні технології та комп'ютерна інженерія. 2022. № 4(56). С. 81–86.

50. Сидоренко О. М., Петренко П. І. Інтелектуальні системи моніторингу в інформаційній безпеці. *Інформаційні технології та комп'ютерна інженерія*. 2022. № 3(55). С. 45–51.

51. Захарченко М. І., Ярошенко С. В**.** Архітектура інтелектуальних систем реагування на інциденти кібербезпеки. *Захист інформації*. 2021. № 2. С. 17–24.

52. Клименко В. С., Яровий О. В. Машинне навчання як інструмент превентивного кіберзахисту. *Технічні науки та технології*. 2021. № 4. С. 40–46.

53. Enterprise Immune System Overview. *Darktrace Ltd.* 2021. URL: https://www.darktrace.com/en/products/enterprise-immune-system/

54. Chio C., Freeman D. Machine Learning and Security: Protecting Systems with Data and Algorithms. Sebastopol: O'Reilly Media, 2018. 386 p.

55. Коваль С. В. Етичні виклики використання ШІ в системах інформаційної безпеки. *Інформаційне суспільство*. 2021. № 2. С. 45–51.

56. Правові та етичні аспекти застосування штучного інтелекту в цифровому середовищі. *Підприємництво, господарство і право*. 2022. № 9. С. 75–79.

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

58. Harris R., Goodman S. E. The Need for Smart AI Cybersecurity Policies. *Communications of the ACM*. 2021. Vol. 64, No. 4. P. 34–36.