COMPUTER SYSTEMS AND NETWORKS

Vol. 7, No. 1, 2025

DOI: https://doi.org/10.23939/csn2025.01.068

УДК 004.056.53

ДОСЛІДЖЕННЯ КІБЕРАТАК З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ НА СИСТЕМИ УПРАВЛІННЯ ІНФОРМАЦІЙНОЮ БЕЗПЕКОЮ

А. В. Габрильчук, В. А. Сусукайло, Є. О. Курій, С. І. Василишин

Національний університет "Львівська політехніка",
кафедра захисту інформації
E-mail: vitalii.a.susukailo@lpnu.ua, yevhenii.o.kurii@lpnu.ua, andrii.habrylchuk.kb.2021@lpnu.ua, sviatoslav.i.vasylyshyn@lpnu.ua

Проаналізовано, як сучасні алгоритми машинного навчання інтегруються в кіберзагрози, змінюючи традиційні підходи до проведення кібератак. Штучний інтелект дає змогу зловмисникам автоматизувати компрометацію систем і адаптувати свої дії до захисних механізмів у реальному часі. Виявлення таких атак є одним із найбільших викликів, оскільки традиційні засоби кіберзахисту не завжди можуть адекватно реагувати на швидкість і динамічність загроз, створених за допомогою штучного інтелекту. Розглянуто ризики, пов'язані із застосуванням ШІ-загроз, зокрема компрометація конфіденційності, порушення репутації організацій та фінансові втрати.

Для протидії цим викликам у статті пропонуються заходи захисту, які базуються на міжнародних стандартах, таких як ISO 27001. Зокрема, наголошується на важливості впровадження контролю доступу, моніторингу загроз, забезпечення цілісності даних, використання криптографії та проведення регулярних аудиторських перевірок безпеки. Також акцентується на необхідності розробки новітніх інструментів для виявлення загроз і запобігання маніпуляціям, які здійснюються за допомогою ШІ.

Ключові слова: EvilProxy, PassGAN, DeepLocker, FaceSwap, Respeecher ISO 27001: 2022.

Вступ

Штучний інтелект (ШІ) став інструментом, який використовується не лише для поліпшення різних сфер життя, а й для створення нових загроз у кіберпросторі. Зловмисники активно інтегрують інструменти ШІ у свої стратегії для підвищення ефективності атак, масштабування їх обсягів та ускладнення процесів їх виявлення. Роль штучного інтелекту в мінливому ландшафті загроз має серйозні наслідки для інформаційної безпеки, що відображає ширший вплив штучного інтелекту через ботів і відповідних систем в інформаційній ері [1]. Наприклад, технології на основі ШІ, такі як PassGAN, використовуються для автоматичного зламу паролів через перебір або аналіз слабких комбінацій, а EvilProxy допомагає обходити механізми двофакторної аутентифікації, що значно ускладнює забезпечення безпеки доступу. Інструменти типу DeepLocker застосовують адаптивні алгоритми для маскування шкідливого програмного забезпечення, яке активується лише за певних умов, наприклад, при взаємодії із заданою ціллю. Використання Deepfake-інструментів, таких як FaceSwap і Respeecher, дає змогу зловмисникам створювати реалістичні фальшиві відео та

аудіо, які вводять в оману як людей, так і автоматизовані системи розпізнавання. Такі технології використовувалися для шахрайства, маніпуляцій громадською думкою та компрометації організацій. Крім того, чат-боти з ШІ на кшталт AI for Hackers дають змогу автоматизувати створення персоналізованих фішингових атак або навіть допомагають непрофесійним зловмисникам генерувати шкідливі скрипти. Атаки на основі ШІ стають дедалі небезпечнішими через здатність адаптуватися до захисних систем у реальному часі, роблячи традиційні методи захисту недостатньо ефективними. Це потребує активного використання сучасних систем захисту, включно з рішеннями на основі машинного навчання, стандартами безпеки, такими як ISO 27001, та інструментами для виявлення маніпуляцій (наприклад, Deepfake).

Огляд літературних джерел

Сучасний розвиток штучного інтелекту (ШІ) та кібербезпеки спричиняє значний вплив на міжнародну безпеку, бізнес та інформаційні системи. Дослідження у цій сфері охоплюють широкий спектр питань – від використання ШІ у військових операціях до його ролі у виявленні кіберзагроз і формуванні нових стандартів безпеки. Зокрема, у праці Горовіца та ін. “Штучний інтелект і міжнародна безпека” акцент зроблено на геополітичних наслідках розвитку ШІ [1]. Автори досліджують питання етичного регулювання та міжнародного контролю над такими технологіями, наголошуючи на потенційних ризиках ескалації конфліктів внаслідок помилок у роботі автоматизованих систем.

У сфері кібербезпеки особливе місце має дослідження методів машинного навчання, які активно впроваджуються для виявлення загроз. У праці “Машинне навчання як ключовий інструмент у захисних кібероперація: ефективність виявлення фішингових загроз” проаналізовано продуктивність алгоритмів глибокого навчання у контексті виявлення фішингових атак [2]. Дослідники порівнюють різні підходи, оцінюючи точність, швидкість реагування та здатність до адаптації таких систем у реальному часі, що дає змогу суттєво знижувати ризики для корпоративних інформаційних систем.

Національний центр кібербезпеки Великої Британії (NCSC) у своєму звіті за 2025 рік також підкреслює стрімке поширення ШІ-інструментів серед кіберзлочинців. Особливу увагу приділено автоматизованим фішинговим кампаніям, генерації персоналізованих атак за допомогою генеративних моделей, а також використанню ШІ для створення шкідливого коду, що ускладнює його виявлення традиційними методами [4]. Цей звіт актуалізує потребу у переосмисленні підходів до кіберзахисту та впровадженні гнучких, адаптивних систем реагування.

Із зростанням популярності чат-ботів на базі штучного інтелекту виникає необхідність аналізу їхнього впливу на інформаційну безпеку, особливо в контексті Систем Управління Інформаційною Безпекою (СУІБ). У праці “Дослідження можливостей використання чатботів зі штучним інтелектом для дослідження журналів подій” розглядається потенціал використання таких ботів для автоматизованого аналізу логів подій [3]. Виявлено, що ШІ-боти можуть прискорити процес виявлення аномалій та реагування на інциденти, що значно покращує ефективність системи моніторингу безпеки в організаціях.

Узагальнюючи, можна сказати, що сучасні дослідження чітко демонструють: впровадження ШІ та новітніх підходів у сфері кіберзахисту є визначальними чинниками для забезпечення стійкості та безпеки інформаційних систем як на рівні окремих компаній, так і на рівні глобальної безпеки. Водночас це потребує комплексного регулювання, етичних стандартів і міжнародної координації зусиль для мінімізації ризиків, пов’язаних із розвитком таких технологій.

Постановка задачі

З огляду на масштабність ризиків і стрімкий розвиток технологій ШІ питання протидії таким загрозам має ключове значення для безпеки компаній, державних установ і суспільства загалом.

Використання рішень на основі штучного інтелекту зловмисниками стає серйозною загрозою у сфері кібербезпеки, оскільки такі технології дають змогу автоматизувати атаки, аналізувати великі обсяги даних для пошуку вразливостей, обходити захисні системи та створювати складні, адаптивні загрози, які важко виявити.

Метою статті є дослідження впливу технологій машинного навчання на зміну та адаптацію кіберзагроз, зокрема на способи використання штучного інтелекту зловмисниками для підвищення ефективності, масштабу та складності атак.

Завдання:

- Дослідити вплив рішень на основі машинного навчання на зміну ландшафту кіберзагроз.
- Аналізувати нові загрози, які виникають через рішення з алгоритмами машинного навчання.
- Дослідити методи адаптації зловмисників з використанням технологій машинного навчання.
- Запропонувати засоби захисту від визначених загроз.

Огляд наявних інструментів

1. Соціальна інженерія з EvilProxy

Соціальна інженерія з використанням рішень машинного навчання, згідно з дослідженнями компанії Microsoft [1], вважається однією з поширених атак у 2024 році. Зокрема, такі інструменти, як EvilProxy, використовують ШІ для створення персоналізованих електронних листів. EvilProxy автоматизує процес вилучення даних із соціальних мереж, даючи змогу зловмисникам створювати фішингові електронні листи, пристосовані до конкретних цілей на основі їхньої поведінки в інтернеті. Від традиційних методів фішинг з використанням штучного інтелекту відрізняє його здатність адаптуватися. ШІ може аналізувати звички та інтереси користувача, автоматично змінюючи вміст листа, щоб підвищити його правдоподібність. Наприклад, в електронному листі, створеному штучним інтелектом, можуть міститися посилання на нещодавні публікації в соціальних мережах, поточні проєкти або навіть особисті інтереси, що робить його майже неможливим виявити як фішинговий на перший погляд. Розробка і впровадження систем для автоматизованого виявлення фішингових атак є важливою складовою сучасних оборонних кібероперацій. Використання таких алгоритмів, як випадкові ліси, логістична регресія та метод опорних векторів, дає змогу комплексно оцінювати ознаки фішингу, зокрема наявність IP-адрес у посиланні, валідність SSL-сертифікатів та кількість субдоменів [2].

2. Атаки грубої сили за допомогою PassGAN

PassGAN використовує машинне навчання для прогнозування ймовірних паролів на основі шаблонів, знайдених у витоку наборів даних паролів. Традиційні атаки грубої сили просто перебирають всі можливі комбінації, але підхід PassGAN з використанням машинного навчання робить цей процес набагато ефективнішим, надаючи пріоритет найбільш вірогідним комбінаціям. На відміну від традиційних інструментів для атак грубої сили, які були обмежені необхідністю послідовно перебирати паролі, PassGAN здатний навчатись на попередньо отриманих позитивних результатах. Інструмент генерує потенційні паролі на основі реальних даних. Таке рішення може передбачити типи паролів, які користувачі, ймовірно, створять, наприклад, на основі особистої інформації, загальних фраз або передбачуваних шаблонів.

3. Шкідливе програмне забезпечення, кероване штучним інтелектом: DeepLocker

DeepLocker представляє нове покоління атак на основі штучного інтелекту, які можуть уникати традиційних методів виявлення. Компанія IBM його розробила як тестовий зразок, DeepLocker перебуває в сплячому режимі доти, доки не визначить свою конкретну ціль, використовуючи машинне начання для розпізнавання рис обличчя, геолокації або навіть голосових даних. Як тільки ціль підтверджено, шкідливе програмне забезпечення активується, доставляючи своє корисне

навантаження у такий спосіб, що його майже неможливо виявити, поки не стане надто пізно. Традиційне шкідливе програмне забезпечення програмується з єдиною метою, але машинне навчання дає змогу шкідливим застосункам вчитися у свого оточення і коригувати свою поведінку. Наприклад, якщо виявлено антивірусне програмне забезпечення, зловмисник може змінити метод атаки, щоб уникнути виявлення. Наприклад, як у випадку з DeepLocker – застосунок може маскуватися під легітимні програми.

4. FaceSwap і Respeecher

Технологія Deepfake дала змогу зловмисникам створювати реалістичні відео та аудіозаписи, які неможливо відрізнити від справжніх. Такі інструменти, як FaceSwap і Respeecher, зазвичай використовуються для того, щоб видавати себе за керівників, знаменитостей або навіть членів сім'ї, що призводить до шахрайства, шантажу або інших зловмисних дій. Ці інструменти штучного інтелекту можуть генерувати переконливий контент, що імітує голос або зовнішність практично будь-якої людини, що полегшує шахраям завдання обманом змусити жертв передати конфіденційну інформацію або великі суми грошей.

Діпфейки особливо небезпечні, оскільки підривають довіру до цифрової комунікації. Оскільки зловмисники вдосконалюють ці інструменти, навіть відеодзвінки або голосова автентифікація можуть стати ненадійними.

5. Чат-боти зі штучним інтелектом

Чат-боти на основі штучного інтелекту зробили обслуговування клієнтів більш ефективним, але зловмисники також використовують цю технологію в зловмисних цілях. SNAP_R - це чат-бот зі штучним інтелектом, розроблений спеціально для фішингових кампаній, що дає змогу зловмисникам автоматизувати персоналізовані фішингові атаки через електронну пошту та соціальні мережі. На відміну від традиційних фішингових листів, які часто є шаблонними і їх легко розпізнати, SNAP_R може вступати в реалістичні розмови зі своїми цілями, коригуючи свої відповіді на основі даних жертви, щоб зробити атаку більш правдоподібною. Окрім уже згаданих інструментів, сучасні зловмисники мають у своєму розпорядженні ще більше можливостей завдяки спеціалізованим чат-ботам та програмам, які базуються на великих мовних моделях (LLM). Одним із таких інструментів є Worm GPT. Цей чат-бот, створений на основі GPT-J з відкритим вихідним кодом, призначений для підтримки зловмисників у програмуванні та кіберзлочинній діяльності. Відсутність обмежень і фільтрів безпеки дає змогу використовувати Worm GPT для створення шкідливого коду, аналізу вразливостей та автоматизації атак. Ще одним прикладом використання машинного навчання зловмисниками є Auto GPT - експериментальна програма, що працює на основі GPT-4. Вона здатна самостійно виконувати завдання, генеруючи необхідні підказки та дії з мінімальним втручанням людини. Auto GPT може стати інструментом у руках кіберзлочинців для автоматизації складних процесів, таких як сканування систем на вразливості чи розробка зловмисного ПЗ.

DAN Prompt ("Do Anything Now") дає змогу знімати обмеження з Chat GPT і розкривати його повний потенціал. Цей інструмент дає змогу створювати контент на заборонені теми, зокрема наркотики, злочини чи насильство, що робить його особливо привабливим для зловмисників. Інший інструмент, Fraud GPT, спеціалізується на створенні текстів для фішингових атак, шахрайських схем та написання шкідливого коду. Його функції дають змогу створювати переконливі повідомлення, які обманом змушують жертв розголошувати конфіденційну інформацію або виконувати небажані дії. Нарешті, Poison GPT демонструє, як можна використати великі мовні моделі для поширення неправдивої інформації та створення шкідливих програм. Його потенціал відкриває нові горизонти для атак із використанням соціальної інженерії.

Всі ці інструменти, від Worm GPT до Poison GPT, є прикладом того, як штучний інтелект може бути використаний не лише для інновацій, а й для створення значних кіберзагроз. Це підкреслює важливість розробки захисних механізмів, які зможуть протистояти цим новим викликам.

Проте чат-боти на основі штучного інтелекту можуть також революціонізувати спосіб розслідування кіберзлочинів, забезпечуючи швидке, масштабоване та економічно ефективне рішення для аналізу великих обсягів даних і виявлення потенційних загроз [3].

Дослідження впливу інструментів з використанням машинного навчання на методи кібератаки

Щороку машинне навчання покращує процес моніторингу, допомагаючи аналітикам безпеки сортувати загрози і швидше закривати вразливості. Згідно з опитуванням, проведеним компанією Sapio Research на замовлення Vanta, 62 % організацій планують збільшити інвестиції в безпеку, використовуючи ШІ протягом наступних 12 місяців.

Але різні моделі штучного інтелекту та методи машинного навчання також допомагають зловмисникам здійснювати більш масштабні та складні атаки, дають змогу обходити засоби контролю безпеки та можливість пошуку нових вразливостей з безпрецедентною швидкістю та руйнівними наслідками для IT-інфраструктур різних компаній. За даними щорічного опитування зловмисників Bugcrowd, опублікованого в жовтні, 77 % зловмисників використовують ШІ для злому, а 86 % стверджують, що він докорінно змінив їхній підхід до злому. Сьогодні 71 % вважають, що технології штучного інтелекту корисні для злому, тоді як у 2023 році таких було лише 21 %.

Експерти попереджають про потенційні проблеми поширення штучного інтелекту. Зокрема, нещодавній звіт Національного центру кібербезпеки Великої Британії (NCSC) наголошує, що протягом найближчих двох років технології ШІ майже напевно збільшать динаміку кібератак та посилять їхній вплив. За оцінками Центру, машинне навчання відкриває широкі можливості у цьому напрямі навіть для тих злочинців, що не мають відповідних технічних навичок. Інакше кажучи, хакерство ще ніколи не було таким доступним [4]. Це занепокоєння поділяють і в бізнесі. Опитування SecurityMagazine 2023 року показало, що 75 % фахівців з безпеки відзначили збільшення кількості кібератак за минулі 12 місяців. 85 % опитаних пов’язують це зростання з використанням зловмисниками генеративного ШІ.

Також згідно з жовтневим звітом компанії Keeper Security, 51 % керівників IT-відділів і служб безпеки вважають, що атаки з використанням ШІ є найсерйознішою загрозою, з якою стикаються їхні організації. Зловмисники можуть використовувати штучний інтелект для найрізноманітніших злочинів: від автоматизованих DDoS-атак до складного фішингу і соціальної інженерії. Наприклад, у 2021 році дослідники McAfee викрили кампанію кібершпигунства під назвою Operation Diànxùn. Злочинці використовували ШІ для створення фішингових електронних листів, націлених на телекомунікаційні компанії у всьому світі. Зловмисники застосували генерацію природної мови, аби створювати переконливі повідомлення, схожі на типові листи рекрутерів або галузевих експертів. Однак ці листи містили вкладення та посилання із шкідливим ПЗ.

У 2023 році зловмисники застосували ШІ, щоб обійти систему біометричної автентифікації криптобіржі Bitfinex, яка вимагала від користувачів підтверджувати свою особу за допомогою обличчя та голосу. Злочинці застосували Deepfake, аби генерувати реалістичні зображення облич, а також голосу і поведінки жертв. У результаті зловмисники привласнили цифрові активи на $150 млн.У лютому 2024 року фінансовий фахівець міжнародної інженерної компанії Arup був уведений в оману злочинцями та перевів їм понад $25 млн. Аби подолати сумніви менеджера щодо дивної транзакції, шахраї у режимі реального часу згенерували за допомогою технологій Deepfake відеодзвінок з фінансовим директором та деякими іншими співробітниками Arup.

Як описано у вищенаведеному прикладі, можливість генерувати адаптивні фішингові листи зловмисники використовують для автоматизації вчинення злочинів у кіберпросторі. Експерти з кібербезпеки з SoSafe, станом на 23 квітня 2023 року, провідного європейського постачальника обізнаності та навчання з питань безпеки давно попереджають про можливість того, що генеративний штучний інтелект може писати фішингові електронні листи краще, ніж люди. Початкові

дослідження SoSafe показують, що це попередження про загрозу виправдане: дані показали, що фішингові електронні листи, написані штучним інтелектом, відкривали 78 % людей, а 21 % продовжували натискати на шкідливий контент всередині наприклад, посилання або вкладення. У проведеному дослідженні фішингові електронні листи, створені людиною, отримали трохи більше кліків - 27 %, тоді як показники відкриття були однаковими як для штучного інтелекту, так і для фішингових листів, створених людиною. Дані демонструють, що люди не можуть відрізнити фішингові електронні листи, створені штучним інтелектом, від фішингових атак, написаних вручну.

Це дослідження також показує, що інструменти генеративного штучного інтелекту можуть допомогти хакерським групам складати фішингові електронні листи принаймні на 40 % швидше, а це означає, що навіть за допомогою простих фішингових атак, створених штучним інтелектом, кіберзлочинці можуть значно підвищити свої показники успіху. Водночас бар'єр входу також знижується для проведення цільового фішингу у великих масштабах, використовуючи вподобання та звички конкретних цілей для створення індивідуальних атак завдяки тому, що штучний інтелект забезпечує можливість проводити масштабовані, персоналізовані фішингові атаки. Інструменти штучного інтелекту можуть отримувати персоналізовану інформацію, яка допомагає підтримувати якість цільових фішингових атак – навіть якщо кількість цілей атаки велика.

Ще одним небезпечним інструментом, який завдяки ШІ отримав безпрецедентний розвиток, є deepfake-технології. Зловмисники можуть генерувати відео або аудіо, яке важко відрізнити від справжньої людини. За останні кілька років було оприлюднено кілька гучних справ, в яких підроблене аудіо обходиться компаніям у сотні тисяч або мільйонів доларів. "Люди отримували телефонні дзвінки від свого боса - це була фальшивка", - каже Мурат Кантарчіоглу, колишній професор комп'ютерних наук у Техаському університеті. Найчастіше шахраї використовують штучний інтелект для створення реалістичних фотографій, профілів користувачів, електронних листів, навіть аудіо та відео, щоб їхні повідомлення виглядали більш правдоподібними. Це великий бізнес. За даними ФБР, шахрайство з компрометацією електронної пошти бізнесу призвело до збитків на суму понад 55 мільярдів доларів за останні десять років [5]. Згідно з опитуванням страхової компанії Nationwide, опублікованим наприкінці вересня, 52 % власників малого бізнесу визнають, що їх обдурило зображення або відео з діпфейком, а 9 із 10 кажуть, що шахрайство з генеративним штучним інтелектом стає дедалі витонченішим [6].

Крім того, одним із головних пріоритетів для кіберзлочинців залишаються паролі, які є ключем до особистої інформації та доступу до систем. Перебір паролів грубою силою - брутфорс, ще один напрям, в якому ШІ-моделі можуть допомогти злочинцю. Завдяки алгоритмам машинного навчання зловмисники більше не обмежені заздалегідь визначеними списками або ручними коригуваннями. Ці складні алгоритми вчаться з кожної спроби, адаптуючи свої стратегії в режимі реального часу, щоб підвищити шанси на успішне вгадування. Вони також можуть інтегрувати дані з різних джерел, зокрема даркнет і витоки даних, щоб оптимізувати свої шаблони вгадування. Здатність штучного інтелекту самонавчатися та адаптуватися перетворює підбір пароля з простої гри "вдар або промах" на високостратегічну та ефективну операцію, що знаменує значну ескалацію в ландшафті кібербезпеки. Традиційні методи вгадування паролів працюють за статичними правилами або попередньо складеними списками, але підходи, керовані штучним інтелектом, можуть негайно адаптуватися до нової інформації. Це робить їх високоефективними навіть проти складних паролів і заходів безпеки, даючи змогу зазирнути в майбутнє загроз кібербезпеці та проблеми, які вони створюють. Традиційні атаки грубої сили обмежені обчислювальною потужністю, доступною зловмиснику. Навпаки, алгоритми, керовані штучним інтелектом, оптимізують кожне припущення, отримуючи максимальну віддачу від кожного обчислювального циклу.

Значну автоматизацію процесів для хакера за допомогою ШІ можна спостерігати на етапі пошуку інформації про жертву. Штучний інтелект і машинне навчання можуть використовувати

для досліджень і розвідки, щоб зловмисники могли переглядати загальнодоступну інформацію та схеми трафіку своєї цілі, захист і потенційні вразливості. Багато організацій не усвідомлюють обсяг наявних даних. І це не просто списки зламаних паролів, які поширюють у даркнеті та пости в соціальних мережах співробітниками. Наприклад, коли компанії розміщують оголошення про вакансії або конкурс пропозицій, вони можуть розкрити типи технологій, які вони використовують, хоч раніше збір усіх цих даних та їх аналіз був трудомістким, але зараз багато з цього можна автоматизувати.

Застосування штучного інтелекту може стати потужним інструментом для зловмисників у створенні шкідливого програмного забезпечення, що використовує технології ШІ. Таке ПЗ здатне здійснювати різні атаки, зокрема автоматичне генерування шкідливого коду, що дає змогу зловмисникам створювати його без необхідності глибоких знань чи часу на написання вручну, виявлення вразливих цілей, наприклад, через аналіз мережевого трафіку для знаходження відкритих портів або для еволюції шкідливого ПЗ, що дає змогу адаптувати код або змінювати його поведінку, обминаючи засоби безпеки. Кіберзлочинці отримали можливість автоматизувати процес вдосконалення атак за допомогою шкідливого ПЗ - від розвідки до ухилення від виявлення. Такі атаки стають дедалі точнішим та підступнішими, здатними обходити традиційні засоби безпеки. Антивіруси і фаєрволи втрачають ефективність проти ПЗ на основі ШІ через здатність його коду динамічно змінюватися. Наприклад, ШІ-боти можуть “мутувати” в реальному часі, залишаючись непоміченими і завдаючи значної шкоди протягом тривалого часу.

Не менш небезпечними стають і атаки типу DDoS, які набули нових масштабів завдяки можливостям штучного інтелекту. Впровадження технологій ШІ відкриває новий рівень автоматизації і для здійснення атак типу відмова в обслуговуванні - DDoS. Тепер навіть користувачі без глибоких технічних знань можуть легко долучитися до DDoS-атак і, за певних умов, завдати суттєвої шкоди цільовій системі. Зловмисники можуть використовувати потужність ШІ для обробки великих обсягів телеметрії безпеки. Цей безперервний аналіз дає змогу їм оптимізувати стратегії атак в реальному часі, ефективно розподіляючи ресурси для досягнення максимального руйнівного ефекту. ШІ також дає змогу чудово маскувати DDoS-атаки під звичайний трафік, що значно ускладнює виявлення та протидію їм. Автоматизовані DDoS-атаки на основі ШІ мають значні переваги порівняно з традиційними, зокрема простоту масштабування, гнучкість і високий рівень маскування. Завдяки низькому порогу входу та високій ефективності ШІ кількість DDoS-атак може значно зрости, створюючи додаткове навантаження на мережі та фахівців, навіть якщо самі атаки будуть технічно не дуже складними.

Підсумовуючи роль штучного інтелекту у кібератаках, важливо зрозуміти, як саме ШІ може впливати на кожен етап злочинної діяльності. Нижче наведено табл. 1, згідно з матрицею MITRE ATT&CK, яка ілюструє, як штучний інтелект посилює ефективність атак та їх етапів.

Таблиця 1

Приклади використання ШІ відповідно до MITRE ATT&CK

Категорія MITRE ATT&CK	Приклади використання ШІ
1	2
Reconnaissance (Розвідка)	Автоматичний збір даних із відкритих джерел - OSINT, аналіз трафіку для виявлення вразливих точок, використання ШІ для обробки великих обсягів інформації про жертву
Resource Development (Розвиток ресурсів)	Генерація переконливих фішингових листів, створення deepfake-контенту: аудіо, відео, зображення, автоматизація створення підроблених профілів користувачів

Продовження табл. 1

1	2
Initial Access (Початковий доступ)	Використання фішингових листів, створених ШІ, персоналізованих для жертв, обхід біометричної автентифікації за допомогою deepfake-технологій
Execution (Виконання)	Розгортання шкідливого програмного забезпечення, створеного ШІ, з функціями самонавчання та адаптації для обходу засобів безпеки
Persistence (Збереження доступу)	Створення динамічних бекдорів за допомогою ШІ, які змінюють поведінку залежно від середовища
Privilege Escalation (Підвищення привілеїв)	Використання ШІ для аналізу та експлуатації відомих вразливостей у системах, адаптація атак у режимі реального часу.
Defense Evasion (Уникнення виявлення)	Мутація шкідливого коду під час виконання, імітація легітимного трафіку або файлів за допомогою генеративних моделей
Credential Access (Отримання облікових даних)	Атаки на паролі з використанням алгоритмів машинного навчання, які адаптуються до кожної невдалої спроби, автоматизоване добування даних із витоків
Discovery (Дослідження)	Аналіз мережевого трафіку для виявлення доступних пристроїв, вивчення системних конфігурацій і політик безпеки з використанням алгоритмів ШІ
Lateral Movement (Бічний рух)	Автоматизація пошуку слабких місць у внутрішніх мережах і переміщення між системами за допомогою розподілених атак
Collection (Збір даних)	Швидкий аналіз даних, зібраних із компрометованих систем, оптимізація їх ексфільтрації на основі аналізу ризиків виявлення
Command and Control (Командування та контроль)	Створення стелс-каналів зв’язку, які імітують звичайний трафік, або використання генеративних моделей для адаптивного управління мережею ботів
Exfiltration (Виведення даних)	Застосування ШІ для маскування виведених даних як звичайних файлів, оптимізація шляхів ексфільтрації залежно від структури мережі жертви
Impact (Вплив)	Оптимізація атак типу DDoS, створення шкідливого коду, здатного адаптуватися для максимального руйнування, атаки на системи за допомогою deepfake-контенту для маніпуляцій

Результати дослідження

Для протидії кібератакам, що використовують елементи машинного навчання (ML), важливо застосувати комплексний підхід із технічними, організаційними та адміністративними заходами захисту. Нижче наведено визначені приклади основних заходів для протидії кібератакам з використанням ML (табл. 2).

Таблиця 2

Приклади заходів протидії атакам з використанням ML

Тип атаки	Опис	Заходи протидії
Введення шкідливих даних	Введення шкідливих даних у модель під час тренування	Аудит і перевірка даних, використання фільтрації
Маніпуляція вхідними даними	Маніпуляція вхідними даними для отримання неправильного результату	Впровадження стійкості до атак
Крадіжка моделі	Крадіжка моделі шляхом запитів до API	Захист API, ліміти запитів, обфускація моделі
Відновлення даних	Відновлення початкових даних із моделі (наприклад, зображень чи тексту)	Шифрування даних, контроль доступу до моделей
Deepfake	Використання підроблених зображень, відео чи голосу для обману	Використання систем розпізнавання Deepfake, навчання персоналу
Обхід моделі	Обхід ML-моделі шляхом створення специфічних даних для уникнення детекції	Покращення алгоритмів детекції, регулярне тестування моделей
Соціальна інженерія з використанням штучного інтелекту	Генерація більш переконливих фішингових повідомлень за допомогою штучного інтелекту	Моніторинг листів, навчання співробітників розпізнавати загрози
Компрометація API	Використання шкідливих запитів до API для маніпуляції поведінкою системи	Захист API через аутентифікацію, моніторинг запитів
Заміна чи підробка даних	Заміна чи підробка даних для тренування моделей, щоб змінити її поведінку	Контроль версій даних, перевірка джерел даних, резервування
Помилки конфігурації рішень штучного інтелекту	Помилки конфігурації рішень штучного інтелекту можуть призвести до витоку конфіденційної інформації або тренуванні публічної моделі на конфіденційних даних	Загалом організації мають розробити та реалізувати політики керування конфігурацією як для нових систем, так і для будь-яких, які вже використовуються [7]. Внутрішній контроль має передбачати критично важливі для бізнесу елементи, такі як конфігурації безпеки рішень штучного інтелекту

Також визначено контролі ISO 27001:2022, які можуть пом’якшити та протидіяти атакам з використанням ML. Стандарт ISO 27001 є визнаним міжнародним стандартом, що визначає вимоги до систем управління інформаційною безпекою (СУІБ) для організацій будь-якого типу та розміру. Цей стандарт надає рамки для розроблення, впровадження, функціонування, моніторингу, оцінки та вдосконалення СУІБ. Міжнародні стандарти ISO / IEC 27001/02 [9, 10] допомагають організаціям різних секторів забезпечувати конфіденційність, цілісність та доступність інформації завдяки застосуванню процесу управління ризиками та надає впевненості зацікавленим сторонам у тому, що ризики адекватно оцінюються та управляються.

Таблиця 3

Контролі ISO 27001:2022 для протидії атакам з використанням ML

Інструмент	Опис загрози	Відповідні контролі ISO 27001:2022	Опис контролю
EvilProxy	Фішингові атаки за допомогою передових методів зворотного проксі-сервера для викрадення облікових даних і файлів cookie	6.3, 8.5, 8.23	Програми поінформованості (6.3), елементи керування автентифікацією користувачів (8.5), вебфільтрація (8.23)
PassGAN	Вгадування та злом паролів на основі ШІ	5.17, 8.24, 8.9	Політики керування паролями (5.17), використання надійного шифрування (8.24) і політики контролю доступу (8.3)
DeepLocker	Зловмисне програмне забезпечення на основі штучного інтелекту приховує зловмисне корисне навантаження в законних програмах	8.7, 8.8, 8.16	Запобігання зловмисному програмному забезпеченню (8.7), керування вразливістю (8.8) і моніторинг аномальної поведінки (8.16)
FaceSwap	Підробка ідентифікаційної інформації та загрози видавання себе за іншу особу на основі Deepfake	6.3, 5.35	Програми підвищення обізнаності (6.3)
Respeecher	Уособлення голосу, згенероване ШІ, яке використовується для шахрайства чи обману	8.5, 8.28, 5.31	Елементи керування автентифікацією (8.5), виявлення аномалій у системах зв’язку (8.16) і відповідність законодавству (5.31)
Чат-боти з ШІ для зловмисників	Використання чат-ботів на основі ШІ для автоматизації соціальної інженерії та розвідки	6.3, 8.23	Програми підвищення обізнаності (6.3), вебфільтрація (8.23)

Такі контролі можуть бути застосовані для СУІБ різних типів організацій. Вони передбачають контроль доступу, криптографію, безпеку зв’язку та керування інцидентами, а також інші аспекти, які допомагають захистити дані та системи, що використовуються.

Висновки

Дослідження підкреслює нагальну необхідність підвищення обізнаності про загрози, пов’язані з використанням технологій штучного інтелекту, зокрема для шахрайства та соціальної інженерії. У сучасних умовах технології ШІ стають інструментами, які можуть бути використані як для конструктивних, так і для деструктивних цілей. Однією з таких технологій є Respeecher, яка дає змогу створювати голосові уособлення, згенеровані ШІ. Попри потенційні позитивні застосування, ці технології можуть бути використані для шахрайства чи обману. Для запобігання таким зловмисним діям необхідно вдосконалювати елементи керування автентифікацією голосу, виявлення аномалій у системах зв’язку та забезпечення відповідності законодавству. Чат-боти на основі ШІ також стають потужним інструментом для зловмисників, які використовують їх для автоматизації соціальної інженерії та розвідки. Впровадження програм підвищення обізнаності про безпеку, вебфільтрації зловмисних дій та етичних практик управління ШІ є критичними для мінімізації ризиків, пов’язаних із використанням таких технологій.

Отже, важливо продовжувати навчальні програми, спрямовані на підвищення обізнаності про безпеку та етичне управління ШІ, щоб забезпечити відповідність законодавству та захистити користувачів від новітніх загроз у сфері кібербезпеки. Ці заходи допоможуть створити більш безпечне інформаційне середовище та запобігти зловживанням новітніми технологіями.

Список літератури

1. *Horowitz M. C. et al. Artificial intelligence and international security. Center for a New American Security. 2022.*

2. *Burova N., Oprysk R., Kurii Y., Lakh Y., Susukailo V. Machine learning as a key tool in defensive cyber operations: the effectiveness of phishing threat detection. Social Development and Security. 2024. 14(5). 113- 123. Doi: https://doi.org/10.33445/sds.2024.14.5.11.*

3. *Opirskyy I., Susukailo V., Vasylyshyn S. Дослідження можливостей використання чатботів зі штучним інтелектом для дослідження журналів подій. Ukrainian Information Security Research Journal. 2023. 24(4). 177- 183. Doi: https://doi.org/10.18372/2410-7840.24.17380.*

4. *The near-term impact of AI on the cyber threat. NCSC. URL: https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat (дата звернення: 19.01.2025)*

5. *Business Email Compromise: The $55 Billion Scam. URL: https://www.ic3.gov/PSA/2024/PSA240911 (дата звернення: 19.01.2025).*

6. *The AI Guardian: A New Era of Cyber Defense. Nationwide. URL: https://news.nationwide.com/the-ai-guardian-a-new-era-of-cyber-defense/ (дата звернення: 19.01.2025).*

7. *Vakhula O., Kurii Y., Opirskyy I., Susukailo V. Security as Code Concept for Fulfilling ISO/IEC 27001: 2022 Requirements. In CPITS, 2024. Pp. 59- 72.*

8. *Kurii Y., Susukailo V., Opirskyy I. Розробка методології оцінки відповідності стандарту ISO 27001. Ukrainian Information Security Research Journal. 25(3). 132- 139. Doi: https://doi.org/10.18372/2410-7840.25.17938.*

9. *ISO/IEC 27001:2022. URL: https://www.iso.org/standard/27001 (дата звернення: 19.01.2025).*

10. *ISO/IEC 27002:2022. URL: https://www.iso.org/standard/75652.html (дата звернення: 19.01.2025).*

ANALYSIS OF CYBER ATTACKS USING MACHINE LEARNING ON THE INFORMATION SECURITY MANAGEMENT SYSTEMS

A. V. Habrylchuk, V. A. Susukailo, Y. O. Kurii, S. I. Vasylyshyn

Lviv Polytechnic National University,
Department of Information Protection

The article analyzes how modern machine learning algorithms are integrated into cyber threats, changing traditional cyberattack approaches. Artificial intelligence allows attackers to automate systems compromise and adapt their actions to real-time defense mechanisms. Detecting such attacks is one of the biggest challenges, as traditional cyber defense tools cannot always adequately respond to the speed and dynamism of threats created with the help of artificial intelligence. The article also examines the risks associated with using AI threats, including privacy compromise, damage to the reputation of organizations, and financial losses. The article proposes protection measures based on international standards, such as ISO 27001, to counter these challenges. In particular, it emphasizes the importance of implementing access controls, threat monitoring, ensuring data integrity, using cryptography, and conducting regular security audits. It also emphasizes the need to develop new tools to detect threats and prevent manipulations carried out using AI.

Keywords: EvilProxy, PassGAN, DeepLocker, FaceSwap, Respeecher, ISO 27001:2022.