© Vavryk Yu., Opirskyy I. Artificial Intelligence: Cybersecurity of the New Generation // Ukrainian Scientific Journal of Information Security, 2024, vol. 30, issue 2, pp. 244-255. ©

DOI: 10.18372/2225-5036.30.19235

ШТУЧНИЙ ІНТЕЛЕКТ: КІБЕРБЕЗПЕКА НОВОГО ПОКОЛІННЯ

Юліан Ваврик, Іван Опірський

Національний університет «Львівська політехніка»

ВАВРИК Юліан Любомирович, студент
Рік та місце народження: 2005 рік, м. Львів, Львівська обл., Україна.
Освіта: Національний університет «Львівська Політехніка», 2022 рік.
Наукові інтереси: застосування штучного інтелекту в кібербезпеці, розвідка на основі відкритих джерел, дослідження deep dark web, соціальна інженерія.
E-mail: yulian.vavryk.kb.2022@lpnu.ua.
Orcid ID: 0009-0006-4863-2914.

ОПІРСЬКИЙ Іван Романович, д.т.н., проф.
Рік та місце народження: 1987 рік, м. Сімферополь, АР Крим, Україна.
Освіта: Національний університет «Львівська Політехніка», 2008 рік.
Посада: завідувач кафедри захисту інформації з 2023 рок.
Наукові інтереси: методи і засоби технічного захисту інформації, охорона державної таємниці, проектування комплексних систем захисту інформації, лазерні системи акустичної розвідки, математичні методи та моделі захисту інформації, технічні канали витоку інформації, спецвимірювання.
Публікації: більше 120 наукових публікацій, серед яких наукові статті, монографії, навчальні посібники, тези та матеріали доповідей на конференціях.
E-mail: ivan.r.opirskyi@lpnu.ua.
Orcid ID: 0000-0002-8461-8996

Анотація. *У статті розглядається роль штучного інтелекту у формуванні кібербезпеки нового покоління. Зі зростанням кіберзагроз та вдосконаленням методів зловмисників, традиційні методи захисту стають недостатньо ефективними. ШІ пропонує інноваційні рішення для протидії цим викликам завдяки здатності аналізувати великі обсяги даних, виявляти аномалії та прогнозувати поведінку зловмисників. Розглядаються різні способи застосування ШІ для захисту даних, включаючи: виявлення кіберзагроз на ранніх стадіях; аналіз даних з різних джерел для ідентифікації потенційних загроз; прогнозування ймовірності кібератак; розробка інтелектуальних систем контролю доступу; автоматизація реагування на кіберінциденти; та вдосконалення алгоритмів шифрування. Також розкриваються нові методи кібератак, які використовують зловмисники, такі як отруєння даних, генерація шкідливого програмного забезпечення за допомогою генеративно-змагальних мереж (GAN), створення фейкового контенту та дипфейків, атаки на IoT-пристрої та використання ШІ для підвищення ефективності соціальної інженерії. Крім того, розглядаються методи аналізу програмного коду на вразливості за допомогою ШІ, зокрема статичний аналіз коду (SAST) та динамічний аналіз коду (DAST), а також використання пісочниць для безпечного тестування коду. Особлива увага приділяється можливостям та ризикам, пов'язаним з використанням "темного боку" ШІ, таким як FraudGPT, WormGPT та Evil-GPT, які використовуються зловмисниками для здійснення кібератак. Наголошується на необхідності комплексного підходу до забезпечення кібербезпеки, що поєднує потужність ШІ-алгоритмів з досвідом фахівців, навчанням користувачів та міжнародним співробітництвом. Подальші дослідження та розробки в цій сфері є критично важливими для забезпечення безпеки цифрового світу в умовах постійно зростаючої кіберзагрози. Для забезпечення ґрунтовної основи дослідження було здійснено комплексне вивчення наукової літератури та актуальних публікацій, присвячених ролі штучного інтелекту в сучасній кібербезпеці.*

Ключові слова. *Штучний інтелект (ШІ), кібербезпека, кіберзагрози, машинне навчання, глибоке навчання, системи виявлення вторгнень (IDS), системи запобігання вторгненням (IPS), Deception технології, SIEM системи, UEBA системи, аналіз програмного коду, пісочниці, зловмисне використання ШІ, цифрова трансформація, інтернет речей (IoT)*

Постановка проблеми

Сучасний світ стрімко еволюціонує під впливом штучного інтелекту (ШІ), який пронизує всі аспекти життя, змінюючи наше повсякдення, роботу та взаємодію з технологіями. Сфера кібербезпеки не стала винятком. Сьогодні кіберзлочинність перетворилась на добре організовану індустрію зі значними ресурсами, що традиційні методи захисту, такі як мережеві екрани та антивіруси, не завжди ефективні, та потребує пошуку нових, більш дієвих підходів до захисту цифрового простору. Саме ШІ стає потужним інструментом нового покоління для забезпечення кібербезпеки. Його здатність аналізувати великі обсяги даних, виявляти аномалії та прогнозувати поведінку

зловмисників відкриває нові можливості для захисту цифрового простору. Завдяки ШІ ми можемо прогнозувати кібератаки та вживати заходів для їхнього запобігання, замість активного реагування на інциденти.

Мета та постановка завдання

Ця стаття розглядає роль штучного інтелекту у формуванні кібербезпеки майбутнього. Ми опишемо широкий спектр методів автоматизації та покращення захисту даних за допомогою ШІ, включаючи ресурси та інструменти для виявлення загроз, аналізу даних та прогнозування кібератак. Крім того, розкриємо нові методи кібератак, які використовують зловмисники, та пояснимо, як ШІ застосовується для аналізу програмного коду на наявність вразливостей та використання пісочниць. Нарешті, охарактеризуємо можливості та ризики використання модифікованих версій нейромереж.

Аналіз останніх досліджень і публікацій

В останній час, частота та складність кібератак продовжує зростати, а злочинці використовують прогресивні методи, щоб обійти звичайні засоби захисту. У результаті організації, незалежно від їх розміру, стикаються з величезним завданням вирішення постійно зростаючої проблеми кібербезпеки. Ця тема набуває особливої актуальності через стрімку цифрову трансформацію суспільства, що призводить до збільшення обсягів та цінності цифрових даних, привертаючи увагу кіберзлочинців. Розвиток хмарних технологій, хоча й має багато переваг, також створює нові виклики для кібербезпеки, ускладнюючи захист та виявлення кібератак в хмарних інфраструктурах [1]. Зростання кількості підключених пристроїв інтернету речей (IoT) створює нові точки входу для кіберзлочинців та збільшує поверхню атаки, оскільки ці пристрої часто мають слабкий захист [2].

Дослідники та експерти активно вивчають можливості застосування ШІ для покращення кібербезпеки та створення систем захисту нового покоління [3, 4].

Штучний інтелект здатен автоматизувати рутинні процеси, такі як аналіз журналів безпеки та моніторинг мережевого трафіку, звільняючи фахівців з кібербезпеки для вирішення складніших завдань. ШІ-алгоритми можуть аналізувати великі обсяги даних, виявляючи складні закономірності та допомагаючи ідентифікувати потенційні кіберзагрози на ранніх стадіях, прогнозуючи майбутні атаки. Застосування ШІ дозволяє автоматизувати процес реагування на кіберінциденти, скорочуючи час реакції та мінімізуючи збитки. ШІ-системи постійно навчаються на нових даних та адаптуються до нових загроз, що робить їх більш ефективними в порівнянні з традиційними методами захисту [4]. Використання ШІ в кібербезпеці – це не просто тренд, а необхідність, зумовлена еволюцією кіберзагроз та зростаючою складністю цифрового світу. Подальші дослідження та розробки в цій галузі є критично важливими для забезпечення безпеки цифрового світу в умовах постійно зростаючої кіберзагрози.

Виклад основного матеріалу дослідження

Необхідність еволюції кіберзахисту в епоху ШІ

Традиційні методи кіберзахисту, такі як мережеві екрани, антивіруси та системи виявлення вторгнень, часто не встигають за еволюцією загроз. Вони зазвичай базуються на сигнатурах та правилах, які можуть бути неефективними проти нових та невідомих типів атак.

Окрім того, організації стикаються з проблемою нестачі кваліфікованих фахівців з кібербезпеки, здатних ефективно протистояти сучасним загрозам.

Також, організації збирають та зберігають величезні обсяги даних, що ускладнює їх захист та виявлення кібератак.

Враховуючи ці виклики, метою даного дослідження є аналіз ролі та визначення потенціалу штучного інтелекту у трансформації кібербезпеки з урахуванням сучасних викликів, а також визначення можливостей та ризиків, пов'язаних з його застосуванням.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

- визначити обмеження традиційних методів кіберзахисту;
- дослідити застосування ШІ для захисту даних;
- описати ресурси та інструменти для виявлення загроз за допомогою ШІ;
- виявити нові методи кібератак з використанням ШІ;
- проаналізувати застосування ШІ для аналізу програмного коду на наявність вразливостей;
- охарактеризувати можливості та ризики використання ШІ зловмисниками;
- визначити перспективи розвитку ШІ в сфері кібербезпеки.

За даними різних досліджень, розглянемо кілька прикладів та статистику:

- згідно зі звітом Cybersecurity Ventures, збитки від кіберзлочинності у світі до 2025 року можуть сягнути 10,5 трильйонів доларів США на рік;

- за даними IBM, середній час виявлення та усунення порушення даних становить 287 днів;

- згідно зі звітом Verizon, 82% порушень даних пов'язані з людським фактором;

- дослідження Microsoft показало, що 99,9% атак на облікові записи можна було б запобігти за допомогою багатофакторної аутентифікації.

Враховуючи ці фактори, стає очевидним, що традиційні підходи до кібербезпеки потребують переосмислення та оновлення. Штучний інтелект має потенціал стати ключовим інструментом для вирішення цієї проблеми та створення більш ефективних та адаптивних систем захисту.

Застосування ШІ для захисту даних

У сучасному світі, де інформація є одним з найцінніших ресурсів, питання кібербезпеки набуває особливої актуальності. З розвитком технологій зростає і винахідливість кіберзлочинців, які постійно шукають нові шляхи для крадіжки, пошкодження або блокування доступу до даних.

На щастя, на допомогу приходить штучний інтелект, який здатен не лише автоматизувати рутинні завдання, але й ефективно протидіяти кіберзагрозам. Завдяки своїй здатності до машинного навчання, ШІ може аналізувати величезні обсяги даних, виявляти аномалії та підозрілу активність, а також прогнозувати потенційні кібератаки.

Розглянемо детальніше, як саме ШІ використовується для захисту даних:

1. Виявлення кіберзагроз: ШІ-алгоритми здатні аналізувати величезні масиви даних з мережевих журналів, системних подій та інших джерел, виявляючи підозрілі патерни, які можуть свідчити про кібератаку. Це дозволяє виявляти загрози на ранніх стадіях, ще до того, як вони завдадуть шкоди [4, 5];

2. Аналіз даних: ШІ може аналізувати дані з різних джерел, таких як електронні листи, соціальні мережі та веб-сайти, щоб виявити потенційні загрози. Наприклад, ШІ може ідентифікувати фішингові листи, виявляти ботів у соціальних мережах та блокувати шкідливі веб-сайти [1];

3. Прогнозування кібератак: ШІ може використовуватись для прогнозування ймовірності кібератак на основі аналізу історичних даних та поточних тенденцій. Це дозволяє організаціям вживати превентивних заходів для захисту своїх даних [3];

4. Контроль доступу: ШІ може використовуватися для розробки більш інтелектуальних систем контролю доступу, які можуть ідентифікувати користувачів на основі їх поведінки та надавати доступ лише до тих ресурсів, які їм потрібні;

5. Реагування на кіберінциденти: ШІ може допомогти організаціям швидко та ефективно реагувати на кіберінциденти, автоматизуючи деякі завдання, такі як ізоляція заражених систем та відновлення даних;

6. Шифрування даних: ШІ може бути використаний для розробки та вдосконалення алгоритмів шифрування, що забезпечує більш надійний захист даних від несанкціонованого доступу;

Напрямки застосування ШІ для захисту даних та їх взаємозв'язок можна візуально представити за допомогою діаграми Венна (рис. 1).

Таблиця детальніше описує кожен напрямок застосування ШІ, включаючи конкретні приклади, переваги та обмеження (табл. 1).

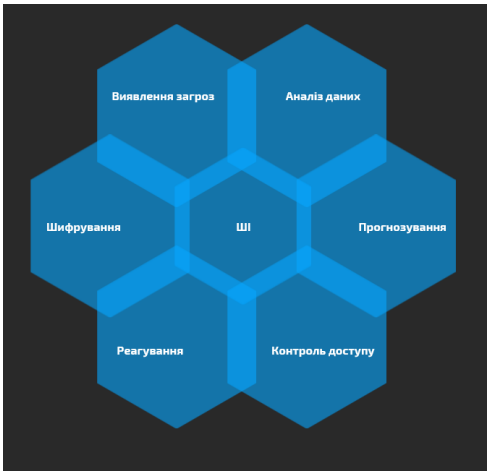

Рис. 1. Діаграма Венна напрямків застосування ШІ в кібербезпеці.

Звичайно, ШІ не є панацеєю від усіх кіберзагроз. Кіберзлочинці також використовують ШІ для своїх цілей, тому боротьба з кіберзлочинністю перетворюється на своєрідне змагання між ШІ-системами захисту та ШІ-системами атаки.

Проте, використання ШІ для захисту даних дає змогу значно підвищити рівень кібербезпеки, автоматизувати рутинні завдання та звільнити фахівців з кібербезпеки для вирішення більш складних проблем.

Таблиця 1

Напрямки застосування ШІ у кібербезпеці: приклади, переваги та обмеження

Напрямок	**Приклади**	**Переваги**	**Обмеження**
Виявлення загроз	Аналіз журналів, виявлення аномалій	Швидкість та точність виявлення нових загроз	Потреба в великих обсягах даних, ризик помилкових спрацьовувань
Аналіз даних	Фішинг, боти, шкідливі сайти	Розуміння загроз, виявлення прихованих зв'язків	Складність інтерпретації результатів, потреба в експертному аналізі
Прогнозування	Аналіз тенденцій, моделювання сценаріїв	Пріоритизація загроз, підготовка до атак, оптимізація ресурсів безпеки	Залежність від якості даних
Контроль доступу	Ідентифікація користувачів на основі поведінки	Покращена безпека, запобігання несанкціонованому доступу, зменшення впливу людського фактору	Потреба в великих обсягах даних для навчання, ризик помилкової ідентифікації
Реагування	Автоматизація дій, ізоляція систем	Швидкість реакції, мінімізація збитків, ефективність	Обмежена гнучкість, потреба в контролі з боку людини
Шифрування	Розробка алгоритмів, управління ключами	Захист даних від несанкціонованого доступу, конфіденційність, забезпечення цілісності даних, відповідність нормативним вимогам	Складність управління ключами, потреба в обчислювальних ресурсах

Ресурси, інструменти для виявлення загроз за допомогою ШІ

Штучний інтелект стрімко революціонізує сферу кібербезпеки, пропонуючи нові можливості для захисту від кіберзагроз. Різноманітні інструменти та ресурси, що ґрунтуються на ШІ-технологіях, дають змогу ефективніше аналізувати дані, прогнозувати ризики, автоматизувати рутинні завдання та реагувати на інциденти.

Наступна таблиця надає огляд основних інструментів, що використовують ШІ для виявлення кіберзагроз, включаючи системи виявлення та запобігання вторгненням (IDS/IPS), системи Deception, системи управління інформацією та подіями безпеки (SIEM) та системи аналізу поведінки користувачів та сутностей (UEBA).

У таблиці описані функції, переваги та приклади кожного інструмент (табл. 2).

Таблиця 2

Інструменти для виявлення загроз з використанням ШІ: функції, переваги та приклади

Інструмент	Функції	Переваги	Приклади
IDS/IPS	Виявлення/блокування вторгнень	Швидкість, автоматизація, захист від відомих та нових загроз	FireEye Helix, Palo Alto Networks Cortex XDR, IBM QRadar
Deception	Створення приманок для зловмисників	Виявлення атак, збір інформації про зловмисників	Attivo Networks, Illusive Networks, Cymmetria
SIEM	Аналіз журналів, виявлення інцидентів	Централізований моніторинг, кореляція подій	Splunk Enterprise Security, LogRhythm SIEM, IBM QRadar
UEBA	Аналіз поведінки користувачів та пристроїв	Виявлення аномалій, інсайдерських загроз	Exabeam, Forcepoint, Gurucul

Системи виявлення вторгнень (IDS) та системи запобігання вторгненням (IPS) є важливими компонентами кібербезпеки, які використовуються для захисту мереж та систем від несанкціонованого доступу, атак та зловживань.

IDS постійно моніторять мережевий трафік та шукають підозрілу активність, яка може свідчити про кібератаку. Це може включати сканування портів, атаки на відмову в обслуговуванні (DoS), спроби проникнення.

IPS роблять те ж саме, що й IDS, але з однією ключовою відмінністю: вони можуть блокувати підозрілий трафік, щоб запобігти йому досягнення цілі.

ШІ революціонізує IDS/IPS, роблячи їх більш ефективними та точними.

Використання штучного інтелекту в IDS/IPS веде до трансформаційної революції, результатом якої є підвищення ефективності та точності. Давайте розглянемо кілька основних застосувань ШІ в IDS/IPS: Здатність ідентифікувати нові та незнайомі кібератаки стала можливою завдяки використанню моделей штучного інтелекту, які проходять навчання з використанням великих наборів даних, що містять інформацію про різні кібератаки. Це дозволяє цим моделям виявляти раніше невидимі та некласифіковані форми атак, які не можуть бути ідентифіковані звичайними методами на основі сигнатур. Використовуючи алгоритми штучного інтелекту, аналіз поведінки можна застосовувати для ретельного вивчення дій користувачів і мережевих пристроїв, щоб виявити будь-які порушення, які можуть свідчити про наявність кібератаки. Прикладом цього може бути виявлення раптового сплеску мережевого трафіку, що надходить із певної IP-адреси, що може вказувати на виникнення DDoS-атаки. Системи штучного інтелекту мають можливість автоматизувати реакцію на кібератаки, ефективно блокуючи трафік, ізолюючи зламані пристрої та надсилаючи сповіщення. Аналізуючи дані про кіберзагрози, моделі штучного інтелекту також можуть прогнозувати потенційні атаки, дозволяючи організаціям проактивно впроваджувати превентивні заходи для свого захисту.

Зараз, системи IDS/IPS активно впроваджують ШІ. Ось кілька прикладів: FireEye Helix — це передова платформа, яка використовує потужність машинного навчання для виявлення та запобігання новим атакам, яких раніше не було. З іншого боку, Palo Alto Networks Cortex XDR використовує передову технологію штучного інтелекту для ретельного аналізу дій користувачів і пристроїв, що дозволяє виявляти та запобігати атакам, які інакше було б неможливо ідентифікувати. Платформа IBM Security QRadar використовує штучний інтелект, щоб передбачати кібератаки та спрощувати процес реагування на них.

Існує кілька переваг використання штучного інтелекту в системах виявлення та запобігання вторгненням. Моделі штучного інтелекту продемонстрували значне підвищення точності виявлення атак, перевершивши ефективність традиційних підходів на основі сигнатур. Крім того, системи ШІ демонструють надзвичайну здатність швидко реагувати на атаки, значно зменшуючи потенційну шкоду, яка може бути завдана. Моделі штучного інтелекту працюють над мінімізацією виникнення помилкових спрацьовувань, що призводить до значної економії часу та ресурсів. Крім того, системи штучного інтелекту відіграють вирішальну роль в автоматизації різноманітних завдань кібербезпеки, що призводить до значної економії витрат і часу для організацій.[4]

Системи Deception розгортають у мережі фальшиві сервери, веб-сайти, програми та інші ресурси, які імітують реальні системи, що можуть зацікавити зловмисників. Ці приманки можуть бути налаштовані для імітації різних типів систем, від банківських до управління контентом.

Як приклад, дослідник з кібербезпеки, може використовувати цю IPS систему для захисту банківської системи.

Може бути розгорнутий фальшивий банківський портал, який буде візуально схожим на реальний, з схожими URL-адресами та функціональними можливостями, такими як вхід, перегляд балансу та переказ коштів. Зловмисники, які шукають вразливості, ймовірно, атакують цей фальшивий портал. У результат – система зафіксує їх дії (введені логіни та паролі, використані інструменти та час атаки) та допоможе розробити кращі методи захисту реальної банківської системи [4].

Системи SIEM (Security Information and Event Management) збагачені ШІ, революціонізують аналіз даних безпеки, пропонуючи нові можливості для зменшення кількості хибних спрацювань, прогнозування майбутніх атак та пріоритизації попереджень.

Автоматизуючи повсякденні завдання, системи ШІ значно підвищують ефективність, дозволяючи експертам з кібербезпеки приділяти увагу більш складним обов'язкам. Крім того, моделі ШІ підвищують точність аналізу даних безпеки, мінімізуючи помилкові спрацьовування та ефективно виявляючи складні загрози.

Автоматизуючи різні завдання з розслідування інцидентів, системи ШІ дозволяють швидше реагувати на кібератаки, ефективно збільшуючи швидкість реагування. Крім того, моделі штучного інтелекту аналізують дані з багатьох джерел, надаючи експертам з кібербезпеки покращену видимість і повне розуміння ландшафту безпеки.

Існує кілька систем SIEM, які включають можливості штучного інтелекту. Splunk Enterprise Security використовує потужність машинного навчання для виявлення порушень і оцінки дій користувачів, а LogRhythm SIEM використовує штучний інтелект для оптимізації запитів про інциденти та прогнозування кібератак.

SIEM системи використовуються для збирання, аналізу та оцінки журналів безпеки з різних джерел, зокрема:

- пристрої, які є частиною мережі;
- сервери;
- брандмауери;
- IPS системи.

Ці дані призначені для виявлення будь-якої ненормальної поведінки, яка потенційно може означати кібернапад [4].

Системи UEBA (User and Entity Behavior Analytics) аналізують поведінку користувачів та пристроїв у мережі. Збираючи дані з журналів подій, мережевого трафіку та даних кінцевих точок, система створює базові профілі "нормальної" активності. Постійний моніторинг дозволяє виявляти відхилення від цих профілів, що можуть свідчити про потенційні загрози: незвичний час входу, доступ до нетипових файлів, підозрілий об'єм переданих даних тощо.

Переваги UEBA систем численні та вагомі: вони виявляють недобросовісних співробітників чи скомпрометовані облікові записи, що використовуються для крадіжки даних або саботажу; розпізнають нові кібератаки, виявляючи аномалії в поведінці, які вказують на раніше невідомі типи загроз; забезпечують швидке реагування, дозволяючи вжити заходів до того, як зловмисники завдадуть значної шкоди; пріоритезують інциденти, допомагаючи зосередити увагу на найважливіших загрозах та оптимізуючи роботу фахівців з безпеки [4].

Хоча інструменти на основі ШІ є потужними засобами для виявлення кіберзагроз, організаціям важливо мати доступ до додаткових ресурсів та рекомендацій для ефективного впровадження та використання цих інструментів. Існує ряд організацій та ініціатив, які надають цінну підтримку.

NIST Cybersecurity Framework (CSF) розроблений Національним інститутом стандартів і технологій США, пропонує цінні рекомендації щодо використання штучного інтелекту для посилення кібербезпеки. Цей фреймворк не лише допомагає організаціям зрозуміти та оцінити свої поточні ризики, але й надає практичні рекомендації щодо впровадження ШІ-технологій для покращення захисту.

CSF рекомендує використовувати ШІ для автоматизації аналізу даних про кібербезпеку, що дозволяє виявляти нові кіберзагрози на ранніх стадіях. Завдяки ШІ-алгоритмам можна ефективніше виявляти аномалії у поведінці користувачів та пристроїв, а також виявляти підозрілу активність у мережі.

Крім того, CSF підкреслює важливість використання ШІ для покращення реагування на кібератаки. ШІ-системи здатні автоматизувати певні дії з реагування на інциденти, що дозволяє скоротити час реакції та мінімізувати потенційні збитки.

Нарешті, CSF рекомендує використовувати ШІ для підвищення стійкості кіберінфраструктури, ШІ-технології допомагають прогнозувати потенційні кібератаки та розробляти ефективні стратегії захисту [9].

ITU AI for Good Global Summit – це щорічна подія, що збирає провідних експертів з усього світу, які прагнуть використати потенціал ШІ для вирішення нагальних глобальних проблем. Організований Міжнародним союзом електрозв'язку (ITU), саміт зосереджується на застосуванні ШІ у різних сферах, включаючи кібербезпеку. На саміті обговорюються захист критичної інфраструктури від кібератак, створення безпечного цифрового середовища для всіх та боротьба з кіберзлочинністю. Учасники діляться досвідом, передовими практиками та інноваційними ідеями щодо використання ШІ для підвищення кіберстійкості, виявлення та протидії кіберзагрозам, а також забезпечення безпеки в інтернеті [10].

OpenAI – провідна дослідницька лабораторія штучного інтелекту, яка робить значний внесок у сферу кібербезпеки. Вона розробляє та публікує передові дослідження та інструменти, які дозволяють виявляти аномалії в мережевому трафіку, аналізувати поведінку користувачів та прогнозувати кібератаки. Завдяки доступу до відкритого коду та API, OpenAI сприяє розвитку інновацій у сфері кібербезпеки та створенню нових інструментів для боротьби з кіберзагрозами. Ці інструменти та ресурси, демонструють значний прорив у сфері, дозволяючи прогнозувати загрози – підвищують ефективність захисту та оптимізують роботу фахівців з кібербезпеки [11].

NIST та ITU, визнаючи потенціал та обмеження ШІ, пропонують ряд ключових рекомендацій для ефективного його використання у кібербезпеці. По-перше, вони наголошують, що ШІ має доповнювати, а не замінювати людський досвід та експертизу. По-

друге, рекомендується зосереджуватися на конкретних завданнях, які ШІ може автоматизувати найбільш ефективно, таких як аналіз даних чи виявлення аномалій. І по-третє, підкреслюється важливість якості та неупередженості даних, які використовуються для навчання ШІ-моделей, оскільки це впливає на точність та ефективність їх роботи.[9, 10].

Ці інструменти та ресурси, демонструють значний прорив у сфері, дозволяючи прогнозувати загрози – підвищують ефективність захисту та оптимізують роботу фахівців з кібербезпеки.

Розробка нових методів кібератак, що використовують ШІ

Зловмисники активно використовують ШІ для розробки нових, більш складних та витончених методів кібератак, які можуть обходити традиційні системи кіберзахисту. Системи машинного навчання, що використовуються для виявлення шкідливого ПЗ та спаму, самі можуть стати ціллю атак. Зловмисники можуть отруювати дані – втручатися в процес навчання моделі додаючи до них спеціально створені приклади, які призводять до неправильної класифікації даних у майбутньому.

Також, можуть використовуватися дані, які спеціально модифіковані з метою обману моделі машинного навчання.

Ці приклади можуть бути майже невідмінними від справжніх даних для людини, але модель класифікує їх неправильно. Ще однією поширеною загрозою є отримання зловмисником доступу до моделі машинного навчання та копіювання її або вивчання поведінки, щоб потім використовувати цю інформацію для проведення атак [13].

Одним з яскравих прикладів атаки на системи машинного навчання є використання генеративно-змагальних мереж (GAN) для створення шкідливого ПЗ, що обходить системи виявлення.

Метод "MalGAN", використовує GAN для генерації прикладів шкідливого ПЗ, які важко відрізнити від доброякісного ПЗ для систем виявлення на основі машинного навчання.[12] Процес роботи MalGAN детально представлений (рис. 2).

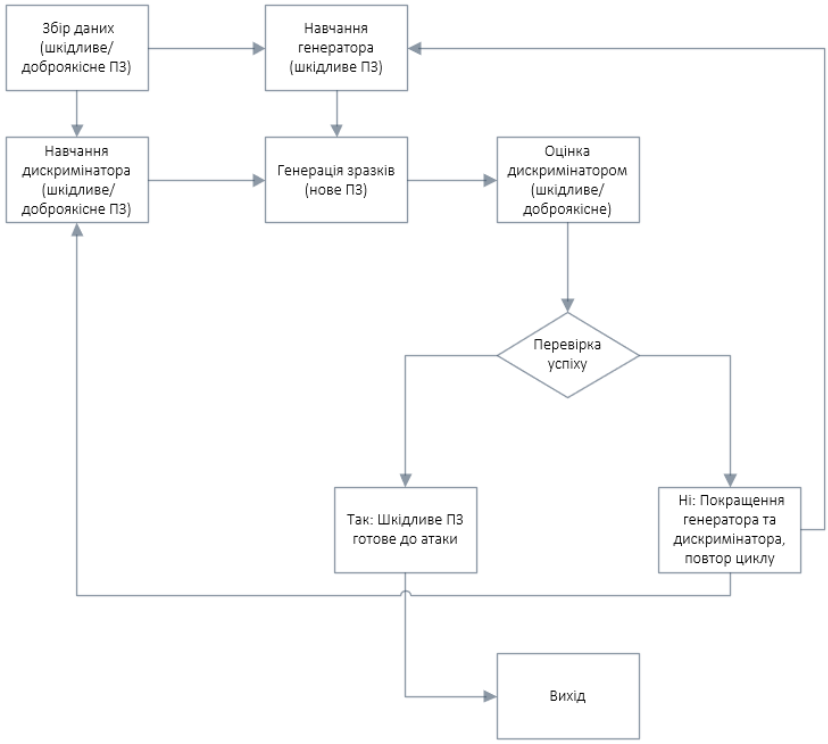

Рис. 2. Блок-схема роботи MalGAN.

Генеративні моделі, такі як ChatGPT, можуть бути використані зловмисниками для створення реалістичного фішингового контенту, який важко відрізнити від справжнього. Генеративні моделі, також, можуть бути використаними для створення фейкових новинних статей, оскільки, контент створений ними, наближений до того, що створений людиною, публікацій у соціальних мережах, щоб маніпулювати громадською думкою або сіяти хаос. Ще, можуть бути створені дипфейки: реалістичні фото-, відео- та аудіоматеріали, на яких люди говорять чи роблять те, чого вони ніколи не робили.Вони можуть бути використаними для шантажу, дискретизації або інших зловмисних цілей [6, 7].

IoT-пристрої, такі як камери відеоспостереження, розумні лампочки або навіть холодильники, часто мають слабкий захист, що робить їх вразливими до використання в ботнетах. ШІ може автоматизувати пошук та експлуатацію таких вразливостей, перетворюючи пристрої на інструмент для атак, зокрема на хмарні сервіси, де вони обслуговуються. Через заражені пристрої ботнет може отримати доступ до хмарних сервісів для здійснення DDoS-атак, крадіжки даних та ін. Завдяки ШІ, ботнети здатні адаптувати свої атаки до мінливих умов та обходити захисні механізми, що робить їх особливо небезпечними для хмарної інфраструктури [2, 8]. Нижче зображена структура такого ботнету (рис. 3) [29].

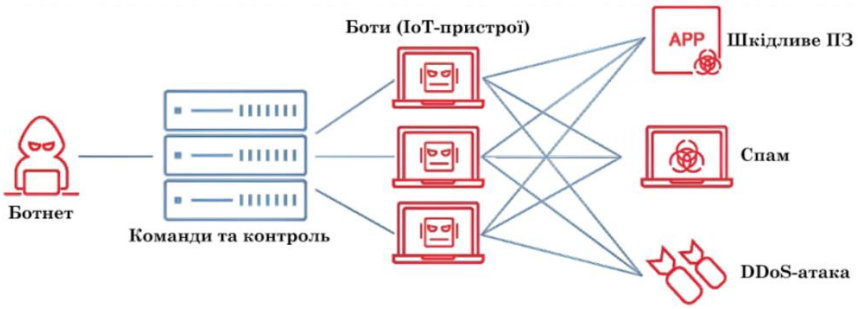

Рис. 3. Схема ботнету з IoT-пристроїв

ШІ може бути використаний для підвищення атак з використанням соціальної інженерії, для різних видів фішингу, зокрема і спискового – метод фішингу, який націлений на конкретних осіб або групи всередині організації. Персоналізований фішинг, за допомогою ШІ та автоматичною розвідкою з відкритих джерел (OSINT), використовується для створення персоналізованих повідомлень, які відповідають інтересам та звичкам жертви, та більшою ймовірністю обдурять її. Також, ШІ використовується для створення ботів, які імітують поведінку реальних користувачів у соціальних мережах та на форумах, щоб маніпулювати громадською думкою та поширенням дезінформації, або збирає конфіденційну інформацію [14].

Шкідливе ПЗ з поліморфними властивостями, яке постійно змінює свій код, щоб уникнути виявлення антивірусними програмами, стає ще небезпечнішим завдяки використанню штучного інтелекту. ШІ може автоматизувати процес модифікації коду, роблячи його більш динамічним і складним для розпізнавання.

При автоматичній генерації, алгоритми глибокого навчання можуть генерувати код, який імітує поведінку легальних програм та приховує шкідливі наміри. Можуть вноситися дрібні зміни в код шкідливого ПЗ, роблячи його неідентичним для виявлення методами, що ґрунтуються на сигнатурах. ШІ може генерувати складні алгоритми шифрування та змінювати динамічні ключі, що робить розшифрування без ключа практично неможливим.

ШІ може бути використаний для просунутої автоматизації атак методом грубої сили, які використовуються для злому паролів та інших облікових даних.

Окрім, ШІ може бути використаний для автоматизації багатьох етапів кібератаки, таких як сканування мереж на наявність вразливостей, розробка експлойтів та розповсюдження шкідливого програмного забезпечення. Це дозволяє кіберзлочинцям проводити атаки швидше та ефективніше.

Як приклад ШІ-керованих кібератак, можна розглянути, вимагач DeepLocker, який використовує ШІ для вибору цілей атаки, аналізуючи дані про жертву, щоб вибрати її цінність та найкращий час для атаки; максимально ефективно шифрує файли жертви, щоб мінімізувати шанси на відновлення даних. DeepLocker володіє унікальними властивостями, які допомагають йому приховувати свої наміри та ціль (рис. 4). Він адаптує свою поведінку, щоб обійти захисти пристрою, що атакується [15].

Як видно з наведених прикладів, штучний інтелект стрімко перетворюється на потужний інструмент у руках зловмисників. Спільнота кібербезпеки повинна адаптуватися до цієї нової реальності, розробляючи інноваційні методи захисту, які враховують постійний розвиток та адаптивність ШІ-керованих кібератак.

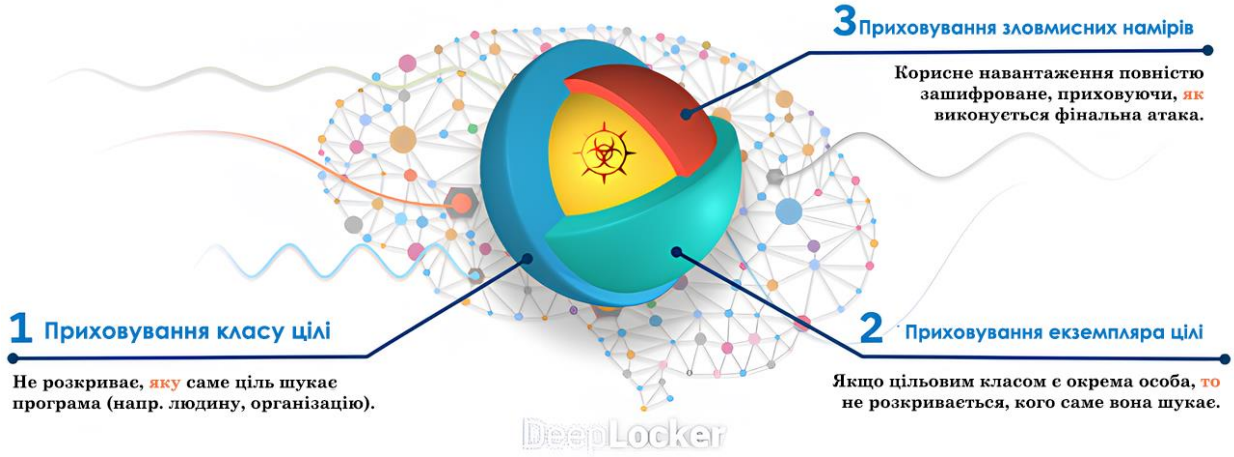

Рис. 4. Методи приховування DeepLocker

Аналіз програмного коду на вразливості та використання пісочниці

У світі веб-розробки, Cross-Site Scripting (XSS) вразливості залишаються однією з найпоширеніших та небезпечних загроз. XSS дозволяє зловмисникам впроваджувати шкідливий код на веб-сторінки, що може призвести до крадіжки даних, або й отримання повного доступу над ресурсом.

Традиційні методи статичного аналізу коду (SAST) часто не здатні виявити складні XSS вра-

зливості, що виникають через поєднання різних факторів та контекстів. Для вирішення цієї проблеми дослідники звертаються до глибокого навчання, яке дозволяє навчати моделі на великих обсягах даних та виявляти складні патерни в коді.

У дослідженні [16] представлено метод статичного аналізу коду на XSS з використанням глибокого навчання. Модель навчається на синтетичному наборі даних PHP та Node.js коду, що містить як безпечні, так і вразливі приклади. Використовуючи різні методи представлення коду, такі як токенізація та AST-дерева, дослідники досягли високої точності виявлення XSS вразливостей, перевершуючи результати традиційних SAST інструментів.

SAST зі ШІ – це потужний інструмент, який поєднує швидкість та масштабованість глибокого навчання з гнучкістю та точністю традиційних SAST інструментів. Завдяки цьому він здатен виявляти складні XSS вразливості, які могли б залишитися непоміченими, а також автоматизувати процес аналізу коду, економлячи час та ресурси фахівців з безпеки. Проте, важливо пам'ятати про ризик помилкових спрацювань та складність інтерпретації результатів. Тому SAST з ШІ рекомендується використовувати в поєднанні з іншими інструментами аналізу коду, ретельно перевіряючи результати та застосовуючи його для раннього виявлення XSS вразливостей [16, 18, 20].

Хоча статичний аналіз коду є корисним інструментом для виявлення потенційних вразливостей, він не може охопити всі можливі сценарії виконання програми. Динамічний аналіз коду (DAST) доповнює SAST, дозволяючи тестувати програму в реальному часі та виявляти вразливості, які проявляються лише під час виконання.

DeepSign – це інноваційний метод для автоматичної генерації сигнатур шкідливого ПЗ з використанням глибокого навчання. DeepSign використовує глибоку мережу переконань (DBN) для створення інваріантного представлення поведінки шкідливого ПЗ, що дозволяє ефективно виявляти нові варіанти відомих загроз [19].

Процес починається з запуску шкідливого ПЗ в пісочниці та збору даних про його поведінку, які потім перетворюються у бінарний вектор для навчання DBN. Навчена DBN може як генерувати сигнатури для виявлення відомих загроз, так і безпосередньо аналізувати нове ПЗ для виявлення потенційно шкідливої активності. Детально зображений процес генерації сигнатур з DeepSign (рис. 5).

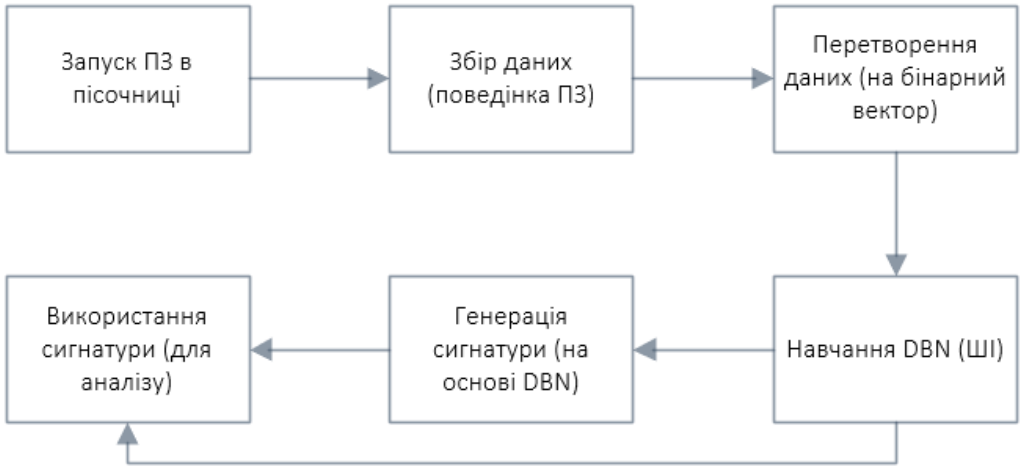

Рис. 5. Процес генерації сигнатур шкідливого ПЗ з DeepSign

DeepSign, як приклад DAST дає можливість виявляти вразливості, які неможливо знайти за допомогою статичного аналізу, наприклад, ті, що залежать від умов виконання або вводу користувача. Цей метод має високу точність завдяки здатності DBN виявляти складні патерни в поведінці шкідливого ПЗ, роблячи його універсальним інструментом для аналізу широкого спектру загроз [18, 19, 20].

Машинне навчання (ML) відкриває нові можливості для аналізу коду, пропонуючи ефективні та масштабовані методи для виявлення вразливостей, прогнозування ризиків та автоматизації завдань розробки.

Однак, для успішного застосування ML в аналізі коду, важливо обрати правильний метод представлення коду, який дозволить моделі "зрозуміти" його структуру та семантику.

Наявні три ключові методи представлення коду, які використовуються в задачах машинного навчання: токенізація, AST-дерева та графи. Кожен з цих методів має свої особливості та переваги, що робить їх корисними для різних завдань аналізу коду.

Токенізація розбиває код на послідовність токенів, таких як ключові слова, ідентифікатори та оператори. Це робить код доступним для аналізу на рівні окремих елементів, що може бути корисним для таких завдань, як класифікація коду, прогнозування ризиків та виявлення вразливостей.

AST-дерева (абстрактні синтаксичні дерева) відображають синтаксичну структуру коду. Це дозволяє моделі аналізувати ієрархію коду та взаємозв'язки між його елементами. AST-дерева можуть використовуватися для розуміння коду, автоматичного виправлення помилок та генерування коду.

Графи ж дозволяють моделювати різноманітні відносини між елементами коду: потоками даних та контролю, викликами функцій та залежності. Це робить графове представлення коду корисним для таких завдань, як аналіз статичного коду, трасування коду та виявлення помилок.

Вибір методу представлення коду залежить від конкретного завдання машинного навчання. Токенізація може бути кращим вибором для задач, які потребують аналізу на рівні окремих елементів, а AST-дерева можуть бути більш корисними для задач, що пов'язані з аналізом структури коду. Графи ж можуть бути оптимальним вибором для задач, які потребують моделювання складних відносин між елементами коду. Важливо зазначити, що жоден з цих методів не є універсальним. Різні методи можуть бути більш або менш ефективними для різних завдань. Тому важливо ретельно обирати метод представлення коду, враховуючи специфіку конкретного завдання.

ML-моделі, навчені на наборах даних з вразливим та безпечним кодом, можуть автоматично класифікувати нові фрагменти коду, значно економлячи час розробників.

ML також може допомогти оцінити ризик виникнення вразливостей у коді, враховуючи його складність, історію змін та наявність подібних проблем в інших проєктах. Це дозволяє зосередити зусилля на аналізі та виправленні ділянок коду з найбільшим ризиком.

Ще однією перевагою ML є автоматичне виправлення помилок. Моделі, навчені на наборах даних з виправленими помилками, можуть знаходити та виправляти помилки у коді, звільняючи час розробників для більш складних завдань.

Машинне навчання має величезний потенціал для автоматизації аналізу коду та покращення безпеки програмного забезпечення. Різноманітні методи представлення коду дозволяють моделювати його структуру та семантику, відкриваючи шлях до нових інструментів та методів аналізу [18, 20].

Пісочниці давно стали невід'ємною частиною процесу забезпечення безпеки програмного забезпечення, надаючи ізольоване середовище для запуску та аналізу підозрілого коду, не ризикуючи безпекою основної системи. Пісочниці моніторять поведінку коду, виявляючи шкідливу активність: спроби доступу до файлів, мережі чи реєстру.

Однак, з розвитком штучного інтелекту, можливості пісочниць значно розширилися, відкриваючи нові горизонти для виявлення та аналізу загроз.

ШІ-алгоритми можуть автоматизувати аналіз даних, зібраних у пісочниці. Це значно економить час та ресурси дослідників, дозволяючи їм зосередитися на більш складних завданнях. ШІ може виявляти підозрілу активність, класифікувати загрози та генерувати попередження про потенційні вразливості.

Алгоритми глибокого навчання, такі як DeepSign, можуть використовуватися для створення сигнатур шкідливого ПЗ та його класифікації. Це дає можливість пісочницям швидше та точніше ідентифікувати зловмисний код, а також краще розуміти його функціональність та цілі [19].

Розподілені системи, що використовують ШІ та пісочниці, можуть обмінюватися інформацією про виявлені загрози. Це дає можливість спільно покращувати ефективність захисту, адже кожна система отримує доступ до знань та досвіду інших.

Серед кращих пісочниць, які використовують ШІ, можна виділити CrowdStrike Falcon, ThreatGrid, MalwareBytes та VirusTotal. Ці платформи демонструють, як штучний інтелект трансформує традиційні методи аналізу на свою користь [21].

Наголошуючи, важливо зазначити, що ШІ та ML не можуть замінити фахівців з кібербезпеки, а лише доповнюють їх, стаючи необхідними інструментами для захисту сучасних складних інформаційних систем.

Можливості та ризики використання темного боку штучного інтелекту

Штучний інтелект стрімко розвивається, пропонуючи безпрецедентні можливості у різних сферах життя. Від автоматизації рутинних завдань до проривів у медицині та науці, ШІ має потенціал революціонізувати світ. Однак, як і з будь-якою потужною технологією, існує і темна сторона ШІ, що несе значні ризики [25, 26]. Зловмисники все частіше використовують ШІ-інструменти для здійснення кібератак або поширення дезінформації. Далі розглянемо, як ШІ перетворюється на інструмент для хакерів та шахраїв, зосереджуючись на таких платформах, як FraudGPT, WormGPT, Evil-GPT, DarkBERT та інших. Дослідимо методи обходу обмежень AI, зокрема DAN (Do Anything Now), та їх потенціал для генерації шкідливого програмного забезпечення, фейків та неточного контенту.

Одним з таких інструментів є FraudGPT. Він став справжнім "швейцарським ножем" для кіберзлочинців, пропонуючи можливості для створення переконливих фішингових повідомлень, розробки шкідливого коду, а також виявлення вразливостей у системах та пошуку потенційних жертв. Доступність FraudGPT за підпискою робить його ще більш загрозливим, розширюючи коло потенційних користувачів [22, 25].

WormGPT – це ще один приклад зловмисного ШІ, спеціалізованого на BEC-атаках (Business Email Compromise). Цей інструмент майстерно імітує стиль ділового листування, створюючи фальшиві повідомлення, які важко відрізнити від справжніх. Зловмисники використовують WormGPT для маніпулювання співробітниками компаній, виманюючи у них гроші або конфіденційну інформацію [25, 27].

Evil-GPT позиціонується як більш доступна альтернатива WormGPT. Обидва інструменти написані на Python, але Evil-GPT пропонується за значно нижчою ціною – всього $10, що рекламується як "неперевершена" пропозиція. Використовується слоган "Ласкаво просимо на Evil-GPT, ворога ChatGPT!", щоб привернути увагу хакерської спільноти. Його поява свідчить про зростаючу конкуренцію на ринку шкідливих ШІ-інструментів та зниження порогу входження для потенційних кіберзлочинців [28].

DarkBERT – інструмент, що використовує ШІ для аналізу величезних обсягів даних з даркнету. Він допомагає зловмисникам знаходити інформацію про вразливості у програмному забезпеченні, потенційні цілі для атак та інші дані, корисні для проведення кібератак [25].

Окрім перерахованих вище, існує безліч інших інструментів, таких як XXXGPT та Wolf GPT, які використовуються зловмисниками для різних цілей, від генерації шкідливого коду до проведення DDoS-атак.

Для кращого розуміння різноманітності та функціональності інструментів для зловмисників, розглянемо порівняльну таблицю (табл. 3).

Таблиця 3

Порівняння зловмисних чат-ботів

Чат-бот	Функціональність	Доступність	Актуальність
FraudGPT	Створення фішингових повідомлень, розробка шкідливого коду, пошук вразливостей та жертв	Підписка	Висока
WormGPT	Імітація ділового листування, створення фальшивих повідомлень	Підписка	Середня
Evil-GPT	Створення фальшивих повідомлень, фішинг	Відкритий код	Низька
DarkBERT	Пошук вразливостей, потенційних жертв, інформації для атак; пошук даркнетом	Відкритий код	Середня

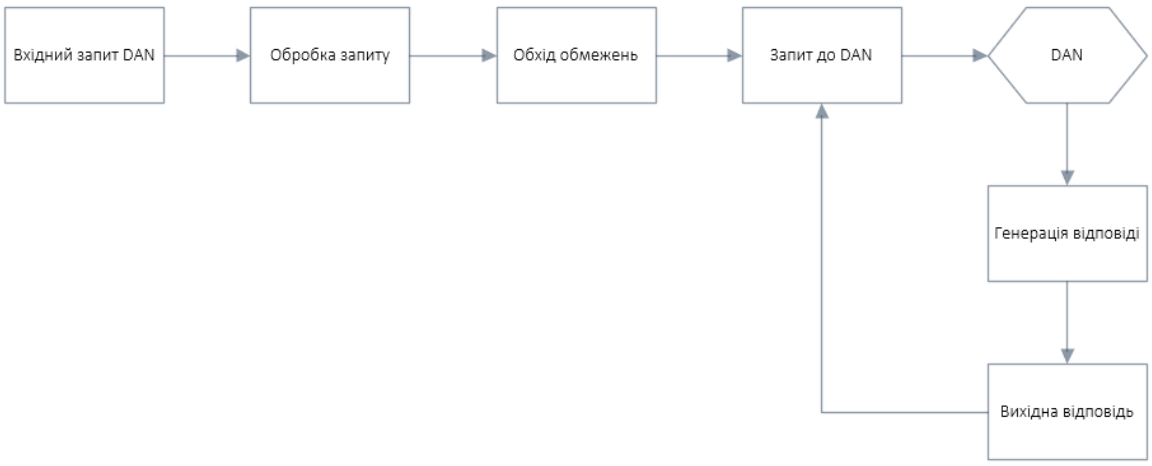

Рис. 6. Процес обходу обмежень генеративної моделі

Хоча розробники генеративних моделей, таких як ChatGPT, впроваджують обмеження для запобігання зловмисному використанню, хакери знаходять способи обійти ці захисні механізми.

Одним з найвідоміших методів є DAN (Do Anything Now). Цей підхід використовує спеціальні запити та інструкції, щоб "зламати" обмеження ChatGPT та змусити його генерувати шкідливий контент, який зазвичай блокується.[23, 24] Наступний рисунок зображає приклад такого запиту-інструкції (рис. 6).

Окрім DAN, існують й інші методи "джейлбрейку", такі як STAN, DUDE та Mongo Tom. Кожен з цих методів використовує різні підходи для маніпулювання моделлю та обходу її обмежень. Наприклад, STAN фокусується на уникненні етичних норм та біасів, а DUDE ігнорує моральні принципи та політику OpenAI [24].

Одним з найбільш тривожних аспектів джейлбрейку AI є можливість генерації шкідливого програмного забезпечення. Зловмисники можуть використовувати DAN та інші методи для створення коду для вірусів, троянів та інших шкідливих програм, що становлять загрозу для комп'ютерних систем [23].

Ще одним ризиком є можливість створення фейків та неточного контенту. ШІ-моделі, звільнені від обмежень, можуть бути використані для генерації фальшивих новин, дипфейків та іншого маніпулятивного контенту, що може дестабілізувати суспільство та підірвати довіру до інформації [23, 26].

Зростаюча загроза зловмисного використання ШІ вимагає комплексного підходу до захисту. Одним з ключових аспектів є навчання користувачів. Необхідно підвищувати обізнаність про ризики, пов'язані з ШІ, та способи захисту від них. Користувачі повинні бути обережними з інформацією, яку вони споживають та діляться в Інтернеті, а також вміти розпізнавати фейки та маніпулятивний контент.

Міжнародне співробітництво також необхідне для ефективної боротьби зі зловмисним ШІ. Уряди, компанії та дослідницькі установи повинні об'єднати зусилля для розробки та впровадження спільних стратегій та обміну інформацією про загрози.

Комплексний підхід, що включає навчання користувачів, технічні рішення, етичні принципи та міжнародне співробітництво, допоможе мінімізувати ризики, пов'язані зі зловмисним ШІ, та забезпечити безпечне та корисне використання цієї технології [26].

Висновки. Дане дослідження було спрямоване на аналіз ролі штучного інтелекту у формуванні кібербезпеки майбутнього. Ключовим питанням було визначення, чи може ШІ стати ключовим інструментом для вирішення проблеми кібербезпеки та створення більш ефективних та адаптивних систем захисту. Для досягнення цієї мети, дослідження зосереджувалося на вирішенні наступних завдань:

Визначення обмежень традиційних методів кіберзахисту: дослідження підтвердило, що традиційні методи кіберзахисту часто неефективні проти нових та невідомих типів атак. Їх залежність від сигнатур та правил робить їх вразливими до сучасних методів зловмисників.

Дослідження застосування ШІ для захисту даних: аналіз показав, що ШІ має значний потенціал для

покращення кібербезпеки. Він може бути використаний для виявлення кіберзагроз на ранніх стадіях, аналізу даних з різних джерел, прогнозування кібератак, розробки інтелектуальних систем контролю доступу, автоматизації реагування на кіберінциденти та вдосконалення алгоритмів шифрування.

Опис ресурсів та інструментів для виявлення загроз за допомогою ШІ: дослідження виявило ряд інструментів та ресурсів, які використовують ШІ для виявлення кіберзагроз. Це включає системи виявлення та запобігання вторгненням (IDS/IPS), системи Deception, системи управління інформацією та подіями безпеки (SIEM) та системи аналізу поведінки користувачів та сутностей (UEBA).

Виявлення нових методів кібератак з використанням ШІ: було виявлено, що зловмисники активно використовують ШІ для розробки нових методів кібератак, таких як отруєння даних, генерація шкідливого програмного забезпечення за допомогою GAN, створення фейкового контенту та дипфейків, атаки на IoT-пристрої та використання ШІ для підвищення ефективності соціальної інженерії.

Аналіз застосування ШІ для аналізу програмного коду на наявність вразливостей: дослідження продемонструвало, як ШІ може бути використаний для аналізу програмного коду на наявність вразливостей. Зокрема, було розглянуто методи статичного аналізу коду (SAST) та динамічного аналізу коду (DAST), а також використання пісочниць для безпечного тестування коду.

Характеристика можливостей та ризиків використання ШІ зловмисниками: було виявлено, що існують значні ризики, пов'язані з використанням "темного боку" ШІ зловмисниками. Такі інструменти, як FraudGPT, WormGPT та Evil-GPT, можуть бути використані для здійснення кібератак та поширення дезінформації. Визначення перспектив розвитку ШІ в сфері кібербезпеки: ШІ має величезний потенціал для подальшого розвитку та вдосконалення кібербезпеки. Подальші дослідження та розробки в цій сфері є критично важливими для забезпечення безпеки цифрового світу в умовах постійно зростаючої кіберзагрози. Результати дослідження підтверджують, що ШІ є потужним інструментом, який може бути використаний як для захисту, так і для атаки. Здатність ШІ аналізувати великі обсяги даних, виявляти аномалії та навчатися робить його цінним активом для розробки адаптивних та ефективних систем кібербезпеки. Однак, важливо розробляти та впроваджувати ШІ відповідально, враховуючи етичні аспекти та потенційні ризики.

Комплексний підхід, що поєднує ШІ-інструменти з традиційними методами захисту, навчанням користувачів та міжнародним співробітництвом, є ключовим для забезпечення безпеки цифрового світу в умовах постійно зростаючої кіберзагрози. Розвиток ШІ в сфері кібербезпеки – це безперервний процес, який вимагає подальших досліджень та розробок для того, щоб випереджати зловмисників та забезпечувати безпеку цифрового світу.

Список літератури

[1]. Shutenko V. (2023, September 13). AI in Cybersecurity: Exploring the Top 6 Use Cases. https:// www.techmagic.co/blog/ai-in-cybersecurity/.

[2]. Jansasoy J. (2023, July 24). 10 IoT Security Challenges and Solutions To Protect Your Devices. https://www.linkedin.com/pulse/10-iot-security-challenges-solutions-protect-your-2023-jansasoy-/.

[3]. Roytman M. (2024, March 5). The Future Of AI And ML In Cybersecurity. https://www.forbes.com/sites/forbestechcouncil/2024/03/05/the-future-of-ai-and-ml-in-cybersecurity/.

[4]. Rjoub G., Bentahar J., Abdel Wahab O., Mizouni R., Song A., Cohen R., Otrok H., Mourad A. (2023). A Survey on Explainable Artificial Intelligence for Cybersecurity https://arxiv.org/pdf/2303.12942.pdf.

[5]. Thwaini Mohammed. (2022). Anomaly Detection in Network Traffic using Machine Learning for Early Threat Detection, https: // dm.saludcyt.ar/index.php/dm/article/view/72/.

[6]. Bisceglia N. (2023, October 22). ChatGPT and Cybersecurity: Risks, Potential Benefits & More, https://teampassword.com/blog/chatgpt-and-cybersecurity/.

[7]. Brundage M., Avin S., Clark J., Toner H., Eckersley P., Garfinkel B., ... & Amodei D. (2018) The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, https: // arxiv.org/ftp /arxiv / papers/1802/1802.07228.pdf.

[8]. Binns R. (2024, March 29). AI Names 2024's Biggest Cybersecurity Threats – and AI is One of Them, https://www.techopedia.com/ai-names-biggest-cybersecurity-threats.

[9]. National Institute of Standards and Technology (NIST). (2024, February 26). The NIST Cybersecurity Framework (CSF) 2.0 [NIST Cybersecurity White Paper (CSWP) NIST CSWP 29], https://doi.org/10.6028/NIST.CSWP.29.

[10]. International Telecommunication Union (ITU). AI for Good Global Summit. https://aiforgood.itu.int/.

[11]. OpenAI Blog, https://openai.com/blog.

[12]. Hu W., Tan Y. A. (2020). Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN, https://arxiv.org/pdf/1702.05983.pdf.

[13]. Rosenberg I., Shabtai A., Rokach L., Elovici Y. (2018). Generic Black-Box End-to-End Attack Against State of the Art API Call Based Malware Classifiers, https://arxiv.org/pdf/1707.05970.pdf.

[14]. Cloudflare. (2023, October 26). The phishing implications of AI chatbots, https://www.cloudflare.com/the-net/chatgpt-phishing/.

[15]. Stoecklin M. P. (2018, August 8). DeepLocker: How AI Can Power a Stealthy New Breed of Malware, https: // securityintelligence.com / deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/.

[16]. Maurel H., Vidal S., Rezk, T. (2021). Statically identifying XSS using deep learning, https: // inria. hal. science/hal-03273564/document.

[17]. Lodeiro-Santiago M., Caballero-Gil C., Caballero-Gil P. (2022). Collaborative SQL-injections detection system with machine learning, https://arxiv.org/pdf/2209.06553.pdf.

[18]. Allamanis M., Barr E. T., Devanbu P., Sutton C. (2018). A survey of machine learning for big code and naturalness, https://arxiv.org/pdf/1709.06182.pdf.

[19]. David O. E., Netanyahu N. S. (2015). DeepSign: Deep learning for automatic malware signature generation and classification, https: // arxiv.org / pdf/1711. 08336.pdf.

[20]. Zhang J. M., Harman M., Ma L., Liu Y. (2019, May). Machine learning testing: Survey, landscapes and horizons, https: // arxiv.org/pdf/1906.10742.pdf.

[21]. Hassan A. (2023, May 16). Top Sandbox Environments Using AI, https://www.linkedin.com/pulse /top-sandbox-environments-using-ai-ahmed-hassan.

[22]. Amos Z. (2023, August 11). What Is FraudGPT? https://hackernoon.com/what-is-fraudgpt.

[23]. Berezovskyi D. (2023, March 15). DAN у GPT-4: як обійти модерацію вмісту. https://chatgpt.com.ua/post/dan-gpt-4-hacking-and-jailbreaking.

[24]. coolaj86. Chat GPT "DAN" (and other "Jailbreaks"), https: // gist.github.com/coolaj86/6f4f7b301-29b0251f61fa7baaa881516.

[25]. Lukatsky A. (2023, September 7). FraudGPT, DarkGPT, WormGPT, Evil-GPT и другие джепеты, https://www.securitylab.ru/blog/personal/Business_ without_danger/353102.php.

[26]. Buckley O., Nurse J. R. C. (2024, February 8). Cybercriminals are creating their own AI chatbots to support hacking and scam users, https://theconversation.com/cybercriminals-are-creating-their-own-ai-chatbots-to-support-hacking-and-scam-users-222643.

[27]. WormGPT, https://wormgpt.com.co/.

[28]. Dutta T. S. (2023, August 10). Hackers Released New Black Hat AI Tool Evil-GPT as a Replacement for Worm GPT, https://cybersecuritynews.com/hackers-released-evil-gpt/.

[29]. A10 Networks. How a bot herder attacks [Image], https: // www.a10networks.com/wp-content/uploads/how-a-bot-herder-attacks.png.

УДК 654.071

Vavryk Yu., Opirskyy I. Artificial Intelligence: Cybersecurity of the New Generation

Abstract. *The article discusses the role of artificial intelligence in shaping the cybersecurity of the new generation. With the rise of cyber threats and the sophistication of attackers' methods, traditional defense mechanisms become insufficiently effective. AI offers innovative solutions to counter these challenges by its ability to analyze large volumes of data, detect anomalies, and predict the behavior of attackers. Various applications of AI for data protection are considered, including: early detection of cyber threats; data analysis from various sources to identify potential threats; predicting the likelihood of cyber attacks; developing intelligent access control systems; automating responses to cyber incidents; and improving encryption algorithms. New cyber attack methods used by hackers are also revealed, such as data poisoning, generating malicious software using generative adversarial networks (GANs), creating fake content and deepfakes, IoT device attacks, and using AI to enhance social engineering effectiveness. Additionally, methods for analyzing code vulnerabilities using AI, including static code analysis (SAST) and dynamic code analysis (DAST), as well as the use of sandboxes for secure code testing, are discussed. Special attention is paid to the capabilities and risks associated with the "dark side" of AI, such as FraudGPT, WormGPT, and Evil-GPT, which are used by cybercriminals to carry out cyber-attacks. Emphasis is placed on the need for a comprehensive approach to cybersecurity, combining the power of AI algorithms with expert knowledge, user education, and international cooperation. Further research and development in this field are critically important for ensuring the security of the digital world in the face of constantly growing cyber threats. To provide a solid research foundation, a comprehensive study of scientific literature and relevant publications dedicated to the role of artificial intelligence in modern cybersecurity was conducted.*

Keywords: *Artificial Intelligence (AI), cybersecurity, cyber threats, machine learning, deep learning, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), Deception technologies, SIEM systems, UEBA systems, code analysis, sandboxes, malicious use of AI, digital transformation, Internet of Things (IoT).*

Ваврик Юліан Любомирович, студент, кафедра захисту інформації, Національного університету «Львівська політехніка».

Yylian Vavryk, student, Department of Information Security, Lviv Polytechnic National University.

Опірський Іван Романович, доктор технічних наук, професор, кафедра захисту інформації, Національного університету «Львівська політехніка».

Ivan Opirskyy, doctor of Technical Sciences, professor, Department of Information Security, Lviv Polytechnic National University.

Отримано *26 травня 2024 року,* затверджено редколегією *26 червня 2024 року*