International Science Journal of Engineering & Agriculture
2024; 3(4): 105-114
https://isg-journal.com/isjea
doi: 10.46299/j.isjea.20240304.11
ISSN: 2720-6319

INTERNATIONAL SCIENCE GROUP

Особливості протистояння оборонного та наступального штучного інтелектів у кіберпросторі

Віталій Фесьоха
Кафедра Комп'ютерних інформаційних технологій, Військовий інститут телекомунікацій та інформатизації імені Героїв Крут, м. Київ, Україна
ORCID 0000-0001-6612-1970

Для цитування цієї статті:
Фесьоха Віталій. Особливості протистояння оборонного та наступального штучного інтелектів у кіберпросторі. International Science Journal of Engineering & Agriculture. Vol. 3, No.4, 2024, pp. 105-114. doi: 10.46299/j.isjea.20240304.11.
Надійшла до редакції: 03 липня 2024 р.; **Схвалено:** 31 липня 2024 р.;
Опубліковано: 01 серпня 2024 р.

Анотація: Застосування штучного інтелекту в кіберпросторі істотно змінює хід протистояння між оборонними та наступальними технологіями. Так, на сучасному етапі розвитку інформаційних технологій, системи та/або моделі штучного інтелекту використовуються не лише для посилення систем кіберзахисту, а й для розробки нових типів (видів) інформаційно-руйнівних впливів у вигляді адаптивних кібератак, потенційно спроможних уникати виявлення існуючими захисними системами. Кібератаки, створені з використанням штучного інтелекту характеризуються прикладною новизною, складністю, швидкістю адаптації та масштабованістю, що робить існуючі методи виявлення кібератак майже не ефективними, що у свою чергу створює серйозну загрозу для інформаційних систем як державного, так і комерційного призначення. До того ж, останнім часом спостерігається значне зростання кількості випадків створення кібератак і шкідливого програмного забезпечення з використанням моделей штучного інтелекту, як наслідок загальнодоступності та фактичної відсутності обмежень на їх використання. Проаналізовано типові підходи до підготовки та застосування як оборонного, так і наступального штучного інтелектів у кіберпросторі з метою ведення оборонних і наступальних (контрнаступальних) кібероперацій. Визначено їх спільні та відмінні особливості, а також взаємовпливаючі фактори та взаємозв'язки в ході протистояння, що дає змогу формувати підґрунтя для вирішення науково-прикладної проблеми запобігання кібератакам, які створені з використанням технологій штучного інтелекту. На основі отриманих особливостей протистояння оборонного та наступального штучного інтелектів у кіберпросторі запропоновано шляхи подальших наукових досліджень щодо забезпечення можливості нівелювання переваг використання технологій штучного інтелекту у зловмисних цілях.

Ключові слова: штучний інтелект, протистояння, кібербезпека, кібератака, інформаційні системи.

1. Вступ

Технологічні досягнення в галузі штучного інтелекту (ШІ), отримані протягом останніх років, дають змогу зловмисникам створювати ефективні деструктивні технології для застосування в кіберпросторі. Згідно зі звітом за 2024 рік міжнародної ради консультантів з електронної комерції EC-Council – найбільшого у світі органу технічної сертифікації з кібербезпеки, кібератаки стають все більш витонченими завдяки удосконаленим

методологіям, алгоритмам та експоненціальним обсягам даних, протидія яким є надскладною задачею [1]. У 2023 році зафіксовано численні факти використання зловмисниками систем генеративного ШІ, зокрема мовної моделі ChatGPT, а також випадки тестування способів обходу обмежень лабораторії досліджень технологій ШІ OpenAI для створення адаптивних кібератак [2]. При чому найбільша частка використання технологій ШІ на сьогоднішній день припадає на розробку поліморфних і метаморфних кібератак, які змінюють структуру та поведінку існуючих (класифікованих) кібератак з метою уникнення виявлення системами кіберзахисту.

Виходячи з цього, основними факторами, які обумовлюють серйозну загрозу використання технологій ШІ для створення і реалізації (супроводження) кібератак є [1,3]:

автоматизовані та адаптивні кібератаки: автоматизація різних етапів здійснення кібератаки, а також адаптація в режимі реального часу до змін у цільовому середовищі, що ускладнює їх виявлення;

цільові та індивідуальні кібератаки: аналіз великих масивів даних з метою виявлення конкретних слабких місць або вразливостей в інфраструктурі об’єкта впливу, що забезпечує здійснення більш цілеспрямованих та успішних кібератак;

автоматизоване створення шкідливого програмного забезпечення: автоматизована розробка поліморфного, метаморфного та олігоморфного шкідливого програмного забезпечення;

автоматизоване виявлення вразливостей: автоматизація процесу виявлення та використання вразливостей об’єкта впливу, що потенційно може призвести до збільшення кількості і масштабів кібератак;

змагальні кібератаки: здійснення впливу на моделі ШІ, які використовуються для захисту інформаційних систем і мереж.

Існуючі системи виявлення вторгнень (кібератак) – Intrusion Detection and Prevention System (IDS/IPS), управління інформаційною безпекою (подіями безпеки) – Security Information and Event Management (SIEM), проактивного виявлення нетипових загроз і цільових кібератак на кінцевих точках – Endpoint Detection and Response (EDR), аналітичні платформи виявлення мережевих кіберзагроз та аномалій – Network detection and response (NDR), а також системи оркестрації та автоматизації кібербезпеки – Security Orchestration and Automated Response (SOAR) використовують різноманітні інтелектуальні та/або адаптивні методи і моделі виявлення кібератак, однак не спроможні повною мірою забезпечити ефективне виявлення навіть поліморфних і метаморфних кібератак – найпоширенішого аспекту застосування ШІ в зловмисних цілях [2,4,5]. Зазначене обумовлює необхідність дослідження особливостей протистояння оборонного та наступального штучного інтелектів у кіберпросторі в рамках науково-прикладної проблеми протидії кібератакам, створених з використанням технологій штучного інтелекту.

2. Об'єкт і предмет дослідження

Об’єкт дослідження: протидія кібератакам, які створені з використанням штучного інтелекту. Предметом дослідження є особливості протистояння оборонного та наступального штучного інтелектів у кіберпросторі.

3. Мета та задачі дослідження

Метою статті є дослідження особливостей протистояння наступального та оборонного штучного інтелектів у кіберпросторі. Для досягнення поставленої мети необхідно провести аналіз типових підходів до підготовки та застосування оборонного та наступального штучного інтелектів у кіберпросторі, а також визначити їх спільні та відмінні особливості, а також взаємовпливаючі фактори в ході протистояння.

4. Аналіз літератури

Науковим дослідженням щодо застосування оборонного ШІ в кібербезпеці присвячено значну кількість робіт: від обґрунтування необхідності застосування технологій ШІ в галузі кібербезпеки [6, 7] до використання численних обчислювальних методів ШІ (еволюційних обчислень, ройового інтелекту, штучних нейронних мереж, штучних імунних систем, машинного навчання, інтелектуального аналізу даних, розпізнавання образів, нечіткої логіки, евристики тощо) [5, 8, 9]. Основна мета застосування технологій ШІ у кібербезпеці полягає в можливості автоматично приймати обґрунтовані рішення під час кіберінцидентів з такою швидкістю та масштабом, які перевищують людські можливості. Виконання завдань з кіберзахисту на основі зазначених методів ШІ (виявлення закономірностей, кластеризація, класифікація, прогнозування) зводиться до виявлення відомих кібератак – ідентифікація за ознаками у відповідності до навчальної вибірки та ідентифікація невідомих вторгнень – виявлення аномалій на основі аналізу стану або поведінки об’єкта захисту. Проте, останнім часом зустрічаються ініціативи щодо передбачення деструктивних впливів на основі емпіричних даних про попередні кібератаки шляхом побудови складних віртуальних сцен інформаційно-руйнівних впливів. [10].

Науковим дослідженням щодо застосування наступального ШІ в кіберпросторі у відкритому доступі присвячено значно менше робіт [10, 11, 12, 13], оскільки для сторони кібервпливу не вигідно нехтувати перевагою непередбачуваності комплексу її заходів у процесі протиборства в кіберпросторі. У [10, 11, 12, 13, 14, 15] зазначається, що на противагу оборонним технологіям зростає загроза з боку наступального ШІ, який застосовує ті ж самі технології, які використовуються у сфері кіберзахисту. Основна передумова використання технологій ШІ в зловмисних цілях – вільний доступ до інструментів ШІ з відкритим вихідним кодом. Метою застосування технологій ШІ у даному випадку є планування та здійснення адаптивних кібератак засобами методів (моделей) машинного навчання та інших обчислювальних технологій на значно вищому рівні, на відміну від традиційних підходів. При цьому переслідуються кілька цілей: здійснити вплив і залишитися непоміченим. Основними шляхами досягнення зазначеної цілі є [8]:

використання методів глибокого машинного навчання для створення кібератак, здатних імітувати нормальну поведінку інших застосунків;

створення інтелектуальних шкідливих програм і кібератак, що адаптуються до умов середовища;

використання змагальних кібератак: застосування моделей генеративного ШІ для отримання даних про ознаки кібератак від систем кіберзахисту, модифікація даних, на яких навчаються моделі машинного навчання.

На основі зазначеного можна зробити висновок, що незважаючи на велику кількість наукових праць, присвячених застосуванню оборонного та наступального ШІ, питання їх протистояння залишається не дослідженим.

5. Методи досліджень

Методи дослідження: аналіз літературних джерел, емпіричні дослідження, порівняльний та статистичний аналіз.

6. Результати досліджень

6.1 Підготовка та застосування оборонного ШІ в кіберпросторі

Оборонний ШІ (Defensive AI) – технології та системи, які використовують алгоритми машинного навчання, моделі нейронних мереж та інші методи обчислювального інтелекту для захисту інформаційних систем від кібератак.

Підготовка оборонного ШІ, як правило, передбачає виконання наступних етапів [8, 10, 11, 12, 13, 14, 15]:

збір та обробка даних для навчання: використання офіційних наборів даних про кібератаки та вразливостей програмного забезпечення або локальний збір інформації з мережевого трафіку, звітних файлів систем з метою опису нормальної поведінки об'єкта захисту, а також очищення та трансформація даних за необхідності;

вибір моделі ШІ (машинного навчання): визначення науково-методичного апарата для подальшого навчання (штучні нейронні мережі, штучні імунні системи, нечітка логіка, метод опорних векторів, дерева рішень тощо) на основі заздалегідь обраних критеріїв;

навчання моделей ШІ: навчання моделей машинного навчання, обраних на попередньому етапі;

розробка інтелектуальних систем кіберзахисту: до існуючих систем IDS/IPS, SIEM, EDR, NDR та SOAR інтегруються моделі ШІ для автоматичного виявлення та реагування на загрози в режимі реального часу;

тестування: моделі ШІ регулярно тестуються на нових наборах даних з метою визначення їх узагальнюючої здатності, ефективності за показниками точності та повноти, а також адаптивності до нових типів кібератак;

оновлення: постійне вдосконалення моделей ШІ через оновлення даних, алгоритмів та технічних засобів, що забезпечує їхню актуальність та ефективність у боротьбі з кібератаками.

На рисунку 1 зображено узагальнену схему підготовки та застосування оборонного ШІ в кіберпросторі.

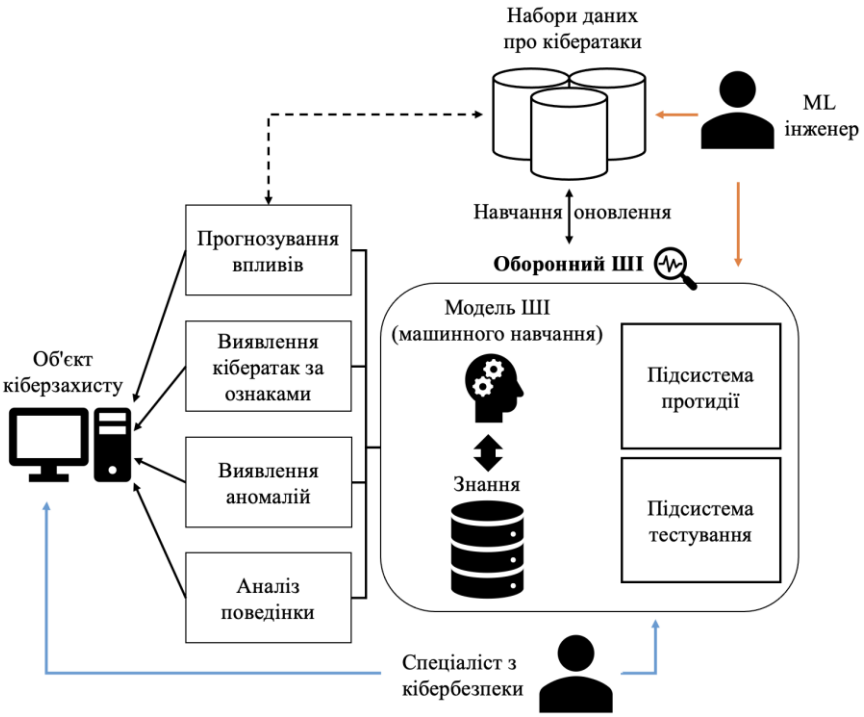

Рис. 1. Узагальнена схема підготовки та застосування оборонного ШІ в кіберпросторі.

Так, застосування оборонного ШІ в кіберпросторі охоплює різні аспекти кібербезпеки, включаючи виявлення, запобігання та протидію загрозам:

виявлення загроз у режимі реального часу: оборонні системи ШІ використовуються для моніторингу мережевого трафіка та системних логів у режимі реального часу з метою оперативного виявлення аномалій та потенційних загроз;

аналіз поведінки: технології ШІ застосовуються для аналізу поведінкових моделей користувачів та інформаційних систем з метою виявлення нетипової (аномальної) активності, яка може свідчити про спроби кібератак;

проактивне виявлення загроз: використання прогнозних моделей ШІ на основі аналізу тенденцій інформаційно-руйнівних впливів, що дозволяє певною мірою передбачати загрози до того, як вони стануть актуальними;

автоматизація протидії: оборонний ШІ використовується для автоматизації процесів протидії наступальному ШІ (ізоляція скомпрометованих підсистем, блокування шкідливого трафіку, автоматичне відновлення після кібератак);

класифікація деструктивної діяльності за ознаками: виявлення кібератак та шкідливого програмного забезпечення на основі відомих ознак;

інтеграція з існуючими засобами кіберзахисту: оборонний ШІ інтегрується з існуючими системами безпеки, такими як IDS/IPS, SIEM, EDR, NDR та SOAR, підвищуючи їх ефективність.

Таким чином, оборонний ШІ забезпечує комплексний підхід до захисту інформаційних систем від кіберзагроз, використовуючи методи машинного навчання для виявлення, та протидії загрозам в режимі реального часу.

6.2 Підготовка та застосування наступального ШІ в кіберпросторі

Наступальний ШІ (Offensive AI) – використання технологій ШІ для зловмисних цілей, включаючи кібератаки на системи ШІ (змагальне машинне навчання) та кібератаки на основі ШІ [16, 17].

Узагальнений підхід до підготовки та застосування наступального ШІ представлено на рисунку 2:

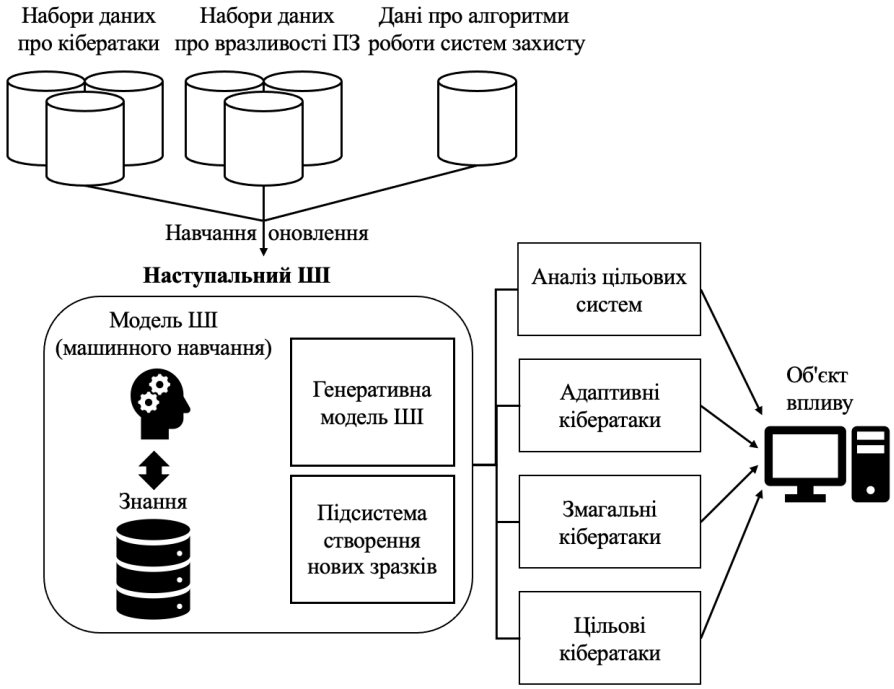

Рис. 2. Узагальнена підхід до підготовки та застосування наступального ШІ.

Основні аспекти підготовки наступального ШІ включають [8, 10, 11, 12, 13, 14, 15]:

збір даних: зловмисники збирають великі обсяги даних про цільові системи, включаючи мережеву архітектуру, програмне забезпечення, використовувані протоколи, потенційні вразливості та алгоритми захисту, зокрема на основі використання методів зворотного інжинірингу;

вибір моделі ШІ (машинного навчання): визначення науково-методичного апарата для подальшого навчання (штучні нейронні мережі, штучні імунні системи, нечітка логіка, метод опорних векторів, дерева рішень тощо) на основі заздалегідь обраних критеріїв;

навчання моделей ШІ на даних, які описують існуючі кібератаки: використання методів машинного навчання для створення поліморфних і метаморфних кібератак;

навчання моделей ШІ використанню вразливостей: використання методів машинного навчання для створення моделей, які можуть автоматично знаходити та експлуатувати вразливості;

змагальне навчання: використання моделей генеративного ШІ для створення нових зразків кібератак або шкідливого програмного забезпечення, які можуть обходити існуючі засоби кіберзахисту із використанням симуляційних моделей захисних систем;

тестування та вдосконалення: зловмисники тестують створені кібератаки та шкідливе програмне на симуляційних моделях захисних систем для визначення їхньої ефективності.

Застосування наступального ШІ в кіберпросторі охоплює різні аспекти проведення кібератак та експлуатації вразливостей:

автоматизація кібератак: використання методів машинного навчання для створення ефективних сценаріїв реалізації кібератак, у тому числі із використанням знайдених вразливостей в об'єкта впливу;

створення поліморфного та метаморфного шкідливого програмного забезпечення: ШІ використовується для створення шкідливого програмного забезпечення, яке може змінювати свою структуру та поведінку для уникнення виявлення традиційними системами захисту;

соціальна інженерія: наступальний ШІ застосовується для аналізу великих обсягів даних про потенційні жертви з метою створення персоналізованих кібератак;

змагальні кібератаки: використання моделей генеративного ШІ для створення нових кібератак, які можуть адаптуватися до алгоритмів кіберзахисту;

кібератаки на моделі ШІ: здійснення кібератак на оборонні моделі ШІ, змінюючи або маніпулюючи даними, на яких вони навчаються з метою зниження їх ефективності;

автоматизоване розгортання атак: використання платформ для автоматичного розгортання кібератак дозволяє зловмисникам масштабувати свої операції та одночасно атакувати багато цілей.

Таким чином, наступальний ШІ в кіберпросторі значно підвищує ефективність та складність кібератак, дозволяючи зловмисникам автоматизувати процеси, адаптуватися до змін в алгоритмах систем кіберзахисту та створювати нові типи загроз.

6.3 Особливості протистояння оборонного та наступального ШІ

На основі аналізу типових підходів до підготовки та використання оборонного та наступального ШІ в кіберпросторі можна зробити висновок, що їх протистояння характеризується динамічною та постійно еволюціонуючою природою кіберзагроз. До того ж, якщо головний принцип оборонного ШІ – інтелектуальне порівняння (виявлення аномалій, класифікація будь якої активності за ознаками з метою ідентифікації кіберзагроз), а наступального – повне розуміння цільової системи та алгоритмів оборонного ШІ, то їх протистояння доцільно розглядати як множину взаємозалежних та постійно змінюваних обмежень, що визначають стратегії обох сторін. Об'єктивно, сторона кібервпливу завжди має першочергову перевагу, оскільки ентропія комплексу її заходів для сторони кіберзахисту є досить високою. У зв'язку з цим, нижнім порогом можливостей оборонного ШІ є забезпечення ситуації паритету в процесі протидії наступальному ШІ в кіберпросторі.

На рисунку 3 представлено узагальнену схему протистояння оборонного на наступального штучного інтелектів у кіберпросторі.

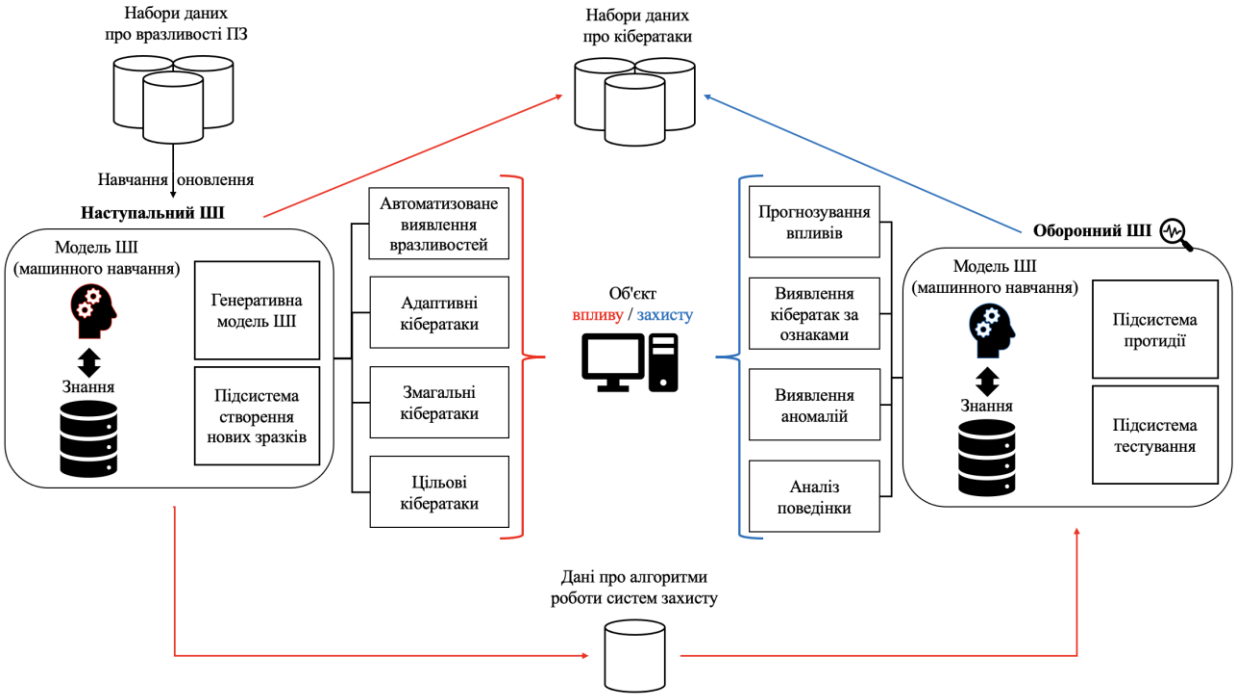

Рис. 3. Узагальнена схема протистояння оборонного на наступального штучного інтелектів у кіберпросторі.

Основні особливості зазначеного протистояння включають наступні аспекти:

1. Спільною сутністю для оборонного та наступального штучного інтелектів є інформація про застосування вже класифікованих кібератак (шкідливого програмного забезпечення). Для обох сторін протиборства такі набори даних є одним з найважливіших аспектів, оскільки використовуються для навчання моделей машинного навчання – «як використовувати» та «за якими ознаками класифікувати», тому дані мають бути коректні та актуальні. Поряд з цим для наступального ШІ в даному аспекті є перевага, суть якої полягає у можливості модифікації навчальних даних за умови наявності доступу до моделі ШІ (машинного навчання).

2. Використання одних і тих самих моделей ШІ (машинного навчання): оборонний штучний інтелект має значні обмеження на використання моделей машинного навчання з відкритим вихідним кодом з метою недопущення надання можливості зловмисникам вивчати дану модель. Окрім того, сторона кібервпливу може ініціювати кібератаки на моделі кіберзахисту за умови наявності доступу до них (фішингові кібератаки, компрометація системи, соціальна інженерія тощо).

3. Спроможності наступального ШІ в процесі створення поліморфних і метаморфних кібератак у своїй більшості обмежуються ізоморфною компонентою кібератак, на основі яких вони згенеровані, тоді як спроможності створення нових кібератак обмежуються властивостями генеративного ШІ та водночас оборонного ШІ.

4. Спроможності оборонного ШІ в процесі кіберзахисту інформаційних систем обмежуються досить високим показником невизначеності комплексу можливих заходів наступального ШІ з одного боку та складністю опису профілів нормальної поведінки усіх модулів об’єкта кіберзахисту з іншого.

5. Адаптивні кібератаки призначені для забезпечення високого рівня прихованості, гнучкості та ефективності, проте у процесі адаптації до середовища виконання та реакцій на сценарії систем кіберзахисту можуть видавати власну присутність, навіть у випадку плавних змін алгоритму функціонування.

6. Перейняття людського фактора: незважаючи на високий рівень автоматизації, як оборонний, так і наступальний ШІ у своїй більшості використовують знання людей або

навчаються на даних, отриманих та підготовлених людьми і тому їх можливості та ефективність значною мірою одинаково залежать від людського внеску.

Таким чином, протистояння оборонного та наступального штучного інтелектів у кіберпросторі вимагає від обох сторін використання передових технологій та адаптивних стратегій. Оборонний ШІ фокусується на виявленні аномалій та класифікації загроз, тоді як наступальний ШІ прагне до максимального розуміння цільових систем і захисних алгоритмів, використовуючи їх у своїх кібератаках. Взаємодія між ними визначається множиною змінюваних обмежень та залежностей, що формують стратегії обох сторін. Основними особливостями даного протистояння є спільне використання інформації про кіберзагрози, використання однакових моделей ШІ, спільна залежність від людського фактора, обмеження наступального штучного інтелекту властивостями генеративних моделей, обмеження оборонного штучного інтелекту невизначеністю комплексу можливих заходів наступального ШІ та складністю опису профілів нормальної поведінки, а також потенційною можливістю відслідковування присутності адаптивних кібератак. Це робить їхнє протистояння складним і багатогранним, вимагаючи постійного вдосконалення підходів до застосування.

7. Перспективи подальшого розвитку досліджень

Перспективним напрямком подальших наукових досліджень є розробка моделі виявлення поліморфних і метаморфних кібератак на основі визначення ізоморфної компоненти у поведінці вже класифікованих кібератак.

8. Висновки

Застосування штучного інтелекту в кіберпросторі істотно змінює хід протистояння між оборонними та наступальними технологіями. Відтак, виникає гостра необхідність пошуку та розробки ефективних рішень протидії технологіям і моделям наступального ШІ. У контексті зазначеного було проаналізовано типові підходи до підготовки та застосування як оборонного, так і наступального штучного інтелектів у кіберпросторі. Визначено їх спільні та відмінні особливості, а також взаємовпливаючі фактори та взаємозв’язки в ході протистояння, а саме: спільне використання інформації про кіберзагрози, використання однакових моделей ШІ, спільна залежність від людського фактора, обмеження наступального штучного інтелекту властивостями генеративних моделей, обмеження оборонного штучного інтелекту невизначеністю комплексу можливих заходів наступального ШІ та складністю опису профілів нормальної поведінки, а також потенційною можливістю відслідковування присутності адаптивних кібератак. Визначені особливості дають змогу краще зрозуміти хід протистояння оборонного та наступального штучного інтелектів, що у свою чергу дає змогу сформувати підґрунтя для розробки нових моделей і методів, спрямованих на підвищення ефективності кіберзахисту інформаційних систем від загроз, створених за допомогою технологій штучного інтелекту.

Список літератури:

1) EC-Council CEH Threat Report 2024: AI and Cybersecurity Report: Discover impactful stats, technical insights, and strategies from experienced cybersecurity pros—perfect for your job. Available at: https://www.eccouncil.org/cybersecurity-exchange/whitepaper/eccouncil-ceh-cybersecurity-threat-report-ai-report/.

2) Фесьоха В. В., Кисиленко Д. Ю., Фесьоха Н. О. (2024). Обґрунтування вибору підходу до визначення інваріантної компоненти у поведінці поліморфного (метаморфного) шкідливого програмного забезпечення на основі зниження розмірності простору ознак. Системи і технології зв’язку, інформатизації та кібербезпеки. Випуск №5. С 181-192. doi: https://doi.org/10.58254/viti.5.2024.16.181/.

3) AI-Powered Attacks – Future threats in cyber. Available at: https://www.linkedin.com/pulse/ai-powered-attacks-future-threats-cyber-cystel.

4) Фесьоха В. В., Кисиленко Д. Ю., Нестеров О. М. (2023). Аналіз спроможності існуючих систем антивірусного захисту та покладених у їхню основу методів до виявлення нового шкідливого програмного забезпечення у військових інформаційних системах. Системи і технології зв'язку, інформатизації та кібербезпеки. Випуск №3. С 143-151. doi: https://doi.org/10.58254/viti.3.2023.16.143.

5) Fesokha V.V., Subach I.Y., Kubrak V.O., Mykytiuk A.V., Korotaiev S.O. (2020). Zero-day polymorphic cyberattacks detection using fuzzy inference system. Austrian Journal of Technical and Natural Sciences. № 5-6. P. 8-13.

6) Гбур З. В. (2022). Використання штучного інтелекту в інформаційній безпеці України. Державне управління: удосконалення та розвиток. № 1. doi: 10.32702/2307-2156-2022.1.2.

7) Тенденції штучного інтелекту в кібербезпеці, на які варто звернути увагу в 2024 році. Режим доступу: https://www.unite.ai/uk/Тенденції-штучного-інтелекту-в-кібербезпеці%2C-на-які-слід-звернути-увагу-в-2024-році/.

8) Selma Dilek, Hüseyin Çakır and Mustafa Aydın. (2015). APPLICATIONS OF ARTIFICIAL INTELLIGENCE TECHNIQUES TO COMBATING CYBER CRIMES: A REVIEW. International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 6, No. 1.

9) Субач І.Ю., Фесьоха В.В., Фесьоха Н.О. (2017). Аналіз існуючих рішень запобігання вторгненням в інформаційно-телекомунікаційні мережі, відкритих на основі загальнодоступних ліцензій. Information Technology and Security. Том. 5, № 1. С. 29–41.

10) Chakraborty, A., Biswas, A., Khan, A.K. (2023). Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation. In: Biswas, A., Semwal, V.B., Singh, D. (eds) Artificial Intelligence for Societal Issues. Intelligent Systems Reference Library, vol 231. Springer, Cham. https://doi.org/10.1007/978-3-031-12419-8_1.

11) Truong, Thanh Cong, Quoc Bao Diep, and Ivan Zelinka. (2020). Artificial Intelligence in the Cyber Domain: Offense and Defense. Symmetry 12, no. 3: 410. https://doi.org/10.3390/sym12030410.

12) Neupane, S., Fernandez, I.A., Mittal, S., & Rahimi, S. (2023). Impacts and Risk of Generative AI Technology on Cyber Defense. *ArXiv, abs/2306.13033*.

13) Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It. Available at: https://www.belfercenter.org/publication/AttackingAI.

14) The impact of artificial intelligence on cyber offence and defence. Available at: https://www.aspistrategist.org.au/the-impact-of-artificial-intelligence-on-cyber-offence-and-defence/.

15) Alex Mathew. (2021). Artificial Intelligence for Offence and Defense - The Future of Cybersecurity. International Journal of Multidisciplinary and Current Educational Research (IJMCER). Volume 3, Issue 3, Pages 159-163.

16) Offensive AI Lab. Available at: https://offensive-ai-lab.github.io.

17) What is Defensive AI and how to use it to increase embedded cybersecurity? Available at: https://www.secure-ic.com/blog/ai/what-is-defensive-ai-and-how-to-use-it-to-increase-embedded-cybersecurity/.

Peculiarities of the confrontation between defensive and offensive artificial intelligence in cyberspace

Vitalii Fesokha
Department of Computer Information Technologies, Kruty Heroes Military Institute of Telecommunications and Information Technologies, Kyiv, Ukraine
ORCID 0000-0001-6612-1970

Abstract: The use of artificial intelligence in cyberspace significantly changes the course of the confrontation between defensive and offensive technologies. Thus, at the current stage of development of information technologies, artificial intelligence systems and/or models are used not only to strengthen cyber defence systems, but also to develop new types (kinds) of information-destructive impacts in the form of adaptive cyber attacks that can potentially avoid detection by existing defence systems. Cyberattacks created with the use of artificial intelligence are characterised by applied novelty, complexity, speed of adaptation and scalability, which makes existing methods of detecting cyberattacks almost ineffective, which in turn poses a serious threat to information systems of both state and commercial purposes. In addition, there has been a significant increase in the number of cases of cyberattacks and malware created using artificial intelligence models recently, as a result of their public availability and the virtual absence of restrictions on their use. The article analyses typical approaches to the training and use of both defensive and offensive artificial intelligence in cyberspace for the purpose of conducting defensive and offensive (counter-offensive) cyber operations. The author identifies their common and distinctive features, as well as interacting factors and interrelationships in the course of confrontation, which makes it possible to form the basis for solving the scientific and applied problem of preventing cyberattacks created using artificial intelligence technologies. Based on the obtained features of the confrontation between defensive and offensive artificial intelligences in cyberspace, the author suggests ways for further scientific research to ensure that the benefits of using artificial intelligence technologies for malicious purposes can be levelled.

Keywords: artificial intelligence, confrontation, cybersecurity, cyberattack, information systems.