Секція 1

АНАЛІЗ АТАК НА СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ

Стацишина І. П.
Національний аерокосмічний університет ім. М. Є. Жуковського «ХАІ»
Науковий керівник: Пєвнєв В. Я.

Актуальність. В сучасному світі аналіз атак на системи штучного інтелекту це доволі актуальне питання. Можливо це стверджувати через те, що у наші дні штучний інтелект стає більш популярним і його використання зустрічається у все більшої кількості програмних продуктів. Тому передчасне виявлення вразливостей систем захисту, аналіз та використання засобів протидії атакам є дуже важливим крок, від якого іноді може залежати робота компанії, безпека майна або навіть власне життя.

Метою даної роботи є аналіз існуючих загроз щодо систем штучного інтелекту та методів їх реалізації.

Основні положення. Штучний інтелект — це здатність машин симулювати розум та імітувати людські когнітивні здібності. Тобто збирати й адаптувати зовнішні дані, а на їх основі навчатися ухвалювати рішення та робити висновки, як могла би людина.

Технології штучного інтелекту міцно увійшли у життя людей на всіх рівнях — від голосових помічників до керованого алгоритмами синтезу стовбурових клітин. І це далеко не межа того, як вони можуть змінити розвиток людської цивілізації [1].

В рамках проведеної роботи було проаналізовано такі загрози:

- змагальні атаки;
- системні маніпуляції;
- пошкодження та отруєння даних;
- передача атак навчання;
- онлайн маніпуляції системою;
- конфіденційності даних.

В доповіді представлено наступні результати: механізми проведення атак, можливі варіанти відбиття цих атак, наслідки втручання в систему штучного інтелекту.

Висновки. В ході аналізу було виявлено, що кожен випадок і набір даних можуть вимагати розгортання іншої стратегії захисту, щоб зберегти базову модель. Різні змагальні підходи можна вирішити за допомогою різних засобів захисту, але немає чітких ознак того, що існує стратегія

захисту, яка може належним чином охопити широкий спектр методів нападу. А високі показники успіху змагальних атак проти справжнього випадку з використанням фактичних даних, отриманих із реального виробничого середовища, вказують на те, що сучасна виробнича система з підтримкою штучного інтелекту може стати плодом розумних зловмисників. Тобто дослідження та інновації необхідно заохочувати, щоб розробити надійні архітектури для очищення конвеєрів даних у виробничих середовищах, фільтрації зловмисних екземплярів і виявлення ін'єкції ворожих прикладів у процесі [2].

Щоб організації могли захистити свої програми штучного інтелекту та моделі машинного навчання, вони повинні використовувати рішення безпеки, які забезпечують дуже безпечне конфіденційне обчислювальне середовище. Коли конфіденційне обчислення поєднується з правильними рішеннями кібербезпеки, такими як стійкий апаратний модуль безпеки (HSM), це може забезпечити надійний наскрізний захист даних у хмарі для додатків штучного інтелекту – великих і малих [3].

Список літератури

1. Даниленко Ю. Від Ш до І: що таке штучний інтелект та як він трансформує світ. *Speka*. URL – https://speka.media/ai/vid-s-do-i-shho-take-stucnii-intelekt-ta-yak-vin-transformuje-svit-xv7039#shho-take-stucnii-intelekt (дата звернення: 01.11.2022);

2. Security threats for AI and machine learning. *Hubsecurity*. URL – https://hubsecurity.com/blog/cyber-security/security-threats-for-ШІ-and-machine-learning/ (дата звернення: 03.11.2022);

3. Towards Robustifying Image Classifiers against the Perils of Adversarial Attacks on Artificial Intelligence Systems. *MDPI*. URL – https://www.mdpi.com/1424-8220/22/18/6905/htm (дата звернення: 06.11.2022).

Відомості про авторів

Стацишина Ірина Павлівна, студент кафедри комп'ютерних систем, мереж і кібербезпеки, м.т. 067-649-59-21, i.statcyshyna@student.csn.khai.edu

Пєвнєв Володимир Яковлевич, професор кафедри комп'ютерних систем, мереж і кібербезпеки, д.т.н., доцент, v.pevnev@csn.khai.edu