Секція 1

ДОСЛІДЖЕННЯ ВАРІАНТІВ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В КОНТЕКСТІ КІБЕРБЕЗПЕКИ: СИСТЕМАТИЗАЦІЯ БАЗОВИХ СЦЕНАРІЇВ І КОНТРЗАХОДІВ

Веприцька О. Ю.
Національний аерокосмічний університет ім. М. Є. Жуковського «ХАІ»
Науковий керівник: Харченко В. С.

Актуальність. Незважаючи на те що, розвиток штучного інтелекту (ШІ) вплинув на зростання рівня автоматизації та інновацій, він також відкрив нові можливості для зловмисників. Використання ШІ як зброї через «вепонізацію» технологій дозволяє здійснювати більш ефективні атаки. Окрім того, виникають нові загрози – «атаки на штучний інтелект», які дозволяють зловмисникам маніпулювати системами штучного інтелекту (СШІ) та змінювати їхню поведінку [1].

Метою даної роботи є класифікація та аналіз сценаріїв, де ШІ розглядається як: об'єкт (система або актив), що має бути захищеним; засоби (технологія) для здійснення атаки, засоби (технологія) захисту від кібератак, а також систематизація контрзаходів для кібератак для визначення їх впливу на цілісність, конфіденційність та доступність. Дослідження базується на принципах і моделях, описаних в [2, 3], та розвиває їх задля зменшення ризиків успішних атак.

Основні положення. В дослідженні використовуються визначення, пов'язані з кібератаками та ШІ, зокрема:

- атака на ШІ – цілеспрямована маніпуляція СШІ з кінцевою метою спричинення її непрацездатності;
- атака, підсилена ШІ (AI powered attack) – атака з використанням технологій ШІ для підвищення дієвості поточних кібератак та створення нових сервісів з використанням ШІ;
- засоби захисту підсилені ШІ (AI powered protection) – програмно-апаратні засоби, які побудовані з використанням технологій ШІ та забезпечують проактивний захист від кібератак.

В рамках проведеної роботи досліджено загрози/атаки на:

- традиційні системи: DDoS атака, генерація DGA, атаки вторгнення, генерація фейкових даних (текстових, аудіо, зображень, відео), генерація фішингових посилань, автоматизоване CAPTCHA-проходження;
- системи ШІ: змагальні атаки (фізичні, цифрові), атаки отруєння (цільові, невибіркові, backdoor), спонж-атака, атаки на моделі ШІ (кража та інверсія моделі, визначення належності до тренувальних даних тощо).

Розглянуто можливі заходи безпеки для попередження і толерування кожної з атак, їхні переваги, недоліки та виклики, пов'язані з обмеженнями

для впровадження. Для оцінювання запропоновано використовувати ризик орієнтований метод, що базується на розширеній техніці IMECA.

Висновки. В доповіді представлено таксономію кібератак з урахуванням аспекту ШІ, надано рекомендації щодо впровадження контрзаходів, базуючись на результатах IMECA-оцінювання ризиків.

Основним науковим результатом є розширена множина сценаріїв, що описується декартовим добутком множин систем ШІ, засобів їх захисту з використанням ШІ та атак, підсилених ШІ. Це надає змогу підвищити повноту оцінювання загроз і атак, а також перейти до глибшого кількісного аналізу з використанням, наприклад, методів теорії ігор і марковських випадкових процесів. Крім того, важливими напрямами подальших досліджень є: розроблення та аналіз сценаріїв з послідовностями атак, підсилених ШІ, з різними просторово-часовими моделями реалізації, а також з огляду на природну резильєнтність засобів штучного інтелекту [4].

Список літератури

1. Kaloudi N., Li J. The AI-Based Cyber Threat Landscape. ACM Computing Surveys. 2020. Vol. 53, no. 1. P. 1–34. URL: https://doi.org/10.1145/3372823;
2. Security-Informed Safety Analysis of Autonomous Transport Systems Considering AI-Powered Cyberattacks and Protection / O. Illiashenko et al. Entropy. 2023. Vol. 25, no. 8. P. 1123. URL: https://doi.org/10.3390/e25081123;
3. Veprytska O., Kharchenko V. AI powered attacks against AI powered protection: classification, scenarios and risk analysis. 2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 9–11 December 2022. 2022. URL: https://doi.org/10.1109/dessert58054.2022.10018770;
4. Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods / V. Moskalenko et al. Algorithms. 2023. Vol. 16, no. 3. P. 165. URL: https://doi.org/10.3390/a16030165.

Відомості про авторів

Веприцька Олена Юріївна, аспірантка кафедри комп'ютерних систем, мереж і кібербезпеки, НАУ «ХАІ», o.veprytska@csn.khai.edu
Харченко Вячеслав Сергійович, завідувач кафедри комп'ютерних систем, мереж і кібербезпеки, НАУ «ХАІ», д.т.н., професор, v.kharchenko@csn.khai.edu