УДК 004.89

ЗАГРОЗИ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ДЕРЖАВНОЇ БЕЗПЕКИ

Скіцько О.І., кандидат технічних наук, старший науковий співробітник, oiskitsko@gmail.com, Національна академія Служби безпеки України

З розвитком штучного інтелекту (ШІ) з'являється новий клас загроз, які повинні бути враховані державою. Ці загрози включають:

1. Кібератаки: використання ШІ у сфері інформаційних технологій підсилює вплив кібератак та робить їх більш результативними. Штучний інтелект в автоматичному режимі виявляє вразливості в системах безпеки та програмному забезпеченні та може здійснювати експлуатацію виявлених вразливостей з наступним маскуванням слідів своєї роботи. Прикладом я реалізації кібератак з застосуванням ШІ є його використання під час здійснення широкомасштабної кібератаки на електроенергетичну інфраструктуру України, що призвело до збоїв у роботі мереж розподілу та передачі електроенергії споживачам різних рівнів.
2. Системи озброєння: ШІ вже зараз використовується в автономних збройних системах. ШІ здійснює координацію взаємодії таких систем у заданому секторі, автоматичне визначення, ведення та враження цілі. Також елементи штучного інтелекту включаються до навігаційних систем автономних збройних систем для вибору оптимального маршруту в межах отриманого завдання [1].
3. Інформаційні війни: ШІ широко використовується в інформаційних війнах та під час проведення інформаційно-психологічних операцій. За допомогою ШІ створюються переконливі фейкові матеріали різних форматів (створення неправдивих відео- та аудіо записів, цільових текстових новин, орієнтованих на різні фокус-групи). та автоматизується поширення таких матеріалів з використанням різних каналів розповсюдження [2]. За допомогою систем з використанням ШІ існує можливість відслідковувати хід інформаційної (дезінформаційної) кампанії та керувати їй в режимі реального часу.
4. Отруєння даних: атака data poisoning (також відома як атака отруєння даних) – атака на системи ШІ, яка полягає в цільовій зміні або спотворенні масивів даних, що використовуються під час навчання моделей машинного навчання. Один з механізмів навчання ШІ полягає у використанні великої кількості даних, на яких відбувається тренування моделі. Дані збираються з різних джерел і з високим ступенем імовірності містять помилки та/або неточності. Атака data poisoning полягає у введенні хибних даних у навчальний масив, завдяки чому рішення, що будуть прийматися з використанням отруєних даних будуть некоректними.

5. Неправильне (незаконне) використання даних: ШІ використовується для збору, нормалізації та аналізу великих масивів структурованих та неструктурованих даних. Результатом оброблення цих масивів може бути порушення приватності та виявлення за непрямими ознаками інформації, що є власністю держави. У 2017 році компанія Strava, яка є розробником популярного фітнес-додатку опублікувала всесвітню карту використання фітнес-гаджетів і програм, і як наслідок було оприлюднено інформацію, що призвела до розкриття місцезнаходження військових баз деяких країн [3].

Реагування на ці загрози вимагає розуміння архітектури, механізмів роботи та технологій штучного інтелекту для запровадження механізмів протидії їх реалізації. Першим кроком запобігання таким загрозам є створення та впровадження механізмів управління ризиками ШІ.

Управління ризиками ШІ

Одним з найважливіших заходів має стати система оцінки та управління ризиками ШІ, на якій базуватиметься політика держави щодо використання систем ШІ.

Вимоги до системи оцінки та управління ризиками наведено нижче [4]:

1. Виявлення ризиків: Це початковий етап, на якому визначаються можливі загрози, асоційовані з використанням ШІ. Він включає в себе ідентифікацію ризиків протягом життєвого циклу систем з ШІ у галузях можливого їх застосування.
2. Оцінювання ризиків: На цьому етапі проводиться аналіз ідентифікованих ризиків за ступенем їх впливу та ймовірності виникнення.
3. Рішення по ризиках: Після аналізу ризиків необхідно визначити, механізми, за допомогою яких ризики будуть оброблятися, а саме: прийняття ризику, обробка ризику, уникнення ризику або страхування ризику використання системи з ШІ. Ці механізми можуть використовуватися окремо або в необхідній комбінації.
4. Контроль ризиків: Охоплює заходи щодо контролю за ризиками, які були ідентифіковані та оброблені раніше.
5. Моніторинг і переоцінка ризиків: Ризики потрібно постійно моніторити і переоцінювати для визначення ефективності заходів обробки. Це може включати проведення аудитів, оцінка впливу, збір відгуків від користувачів та інші методи моніторингу

Висновки

Вищезазначені етапи необхідно циклічно повторювати (наприклад, з використанням моделі Plan-Do-Check-Act) для того, щоб врахувати ризики, які змінюються або з’являються з часом. Також, оцінка ризиків необхідна на всіх етапах життєвого цикли систем ШІ, від формування вимог для створення до виводу з експлуатації. Оцінка та управління ризиками ШІ може бути основою

системи заходів для протидії ризикам та загрозам, які виникають внаслідок використання систем ШІ.

Список використаних джерел

1. Will Knight. The AI-Powered, Totally Autonomous Future of War Is Here: веб-сайт. URL: https://www.wired.com/story/ai-powered-totally-autonomous-future-of-war-is-here/ (дата звернення: 29.02.2024).
2. Raphael Satter. Exclusive: AI being used for hacking and misinformation, top Canadian cyber official says : веб-сайт. URL:https://www.reuters.com/technology/ai-being-used-hacking-misinfo-top-canadian-cyber-official-says-2023-07-20 (дата звернення: 29.02.2024).
3. Liz Sly. U.S. soldiers are revealing sensitive and dangerous information by jogging: веб-сайт. URL: https://www.washingtonpost.com/world/a-map-showing-the-users-of-fitness-devices-lets-the-world-see-where-us-soldiers-are-and-what-they-are-doing/2018/01/28/86915662-0441-11e8-aa61-f3391373867e_story.html (дата звернення: 29.02.2024).
4. Скіцько О., Складний П., Ширшов Р., Гуменюк М., Ворохоб М. Загрози та ризики використання штучного інтелекту. Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка» - Том 2 (22). 2023, С. 6-18. DOI: https://doi.org/10.28925/2663-4023.2023.22.618

УДК 004.89

СИНТЕЗ ПОЯСНЕННИХ ВЕРБАЛЬНИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ

Фастовський Е.Г., аспірант, eduard.fastovskyi@khpi.edu.ua, НТУ «ХПІ»

Інформаційна технологія аналізу і синтезу пояснювальних моделей штучного інтелекту (ШІ) на основі вербальних методів передбачає використання обробки природної мови і машинного навчання для створення розмовних систем ШІ, які можуть розуміти людську мову і реагувати на неї. Ці системи можна використовувати в різних додатках, таких як чат-боти, віртуальні асистенти та голосові помічники, для покращення розуміння рекомендації, що надаються користувачеві [1-3].

Методи вербального аналізу використовуються для розв'язування складних неструктурованих проблем. Дослідження в цієї галузі зосереджені на розробці методів підтримки прийняття рішень, які включають як числові, так і вербальні аспекти. Вони звертають увагу на важливість включення вербальних елементів у процеси прийняття рішень, підкреслюючи цінність лінгвістичної