Секція 1

МЕТОДИ ЗАХИСТУ ВІД АТАК НА СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ

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

Актуальність. Кібербезпека ШІ стає одним із пріоритетних напрямів роботи урядів і міжнародних організацій, оскільки все більше країн усвідомлюють потенційні ризики, пов'язані з використанням і розвитком штучного інтелекту [2].

Метою даної роботи є аналіз існуючих методів захисту систем штучного інтелекту від кіберзагроз.

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 методи:

- вороже навчання;
- робастна оптимізація;
- регуляризація;
- ансамблеві методи;
- нормалізація даних.

Вони дозволяють системам ШІ краще адаптуватися до атак і знижувати ймовірність маніпуляцій. Утім, ці підходи потребують значних обчислювальних ресурсів, і їхня ефективність в реальних умовах залишається предметом подальших досліджень. Загрози на кшталт ворожих атак, отруєння даних та маніпуляцій алгоритмами є надзвичайно небезпечними для ШІ-систем, особливо тих, які використовуються у критичних застосуваннях (медицина, транспорт, фінанси). Виявлено, що навіть мінімальні зміни в даних можуть спричинити серйозні помилки, що підкреслює важливість розвитку нових стратегій для запобігання таким атакам. Загалом, хоча розроблено декілька підходів для захисту ШІ,

дослідження в цій сфері лише на початковому етапі. Подальший розвиток технологій потребує створення нових методів, які враховують реальні умови застосування, забезпечуючи безпеку ШІ на всіх рівнях – від алгоритмів до даних.

Висновки. Таким чином, розвиток ШІ відкриває безпрецедентні можливості, але супроводжується значними ризиками. Важливо, щоб технології захисту ШІ не відставали від темпів розвитку самих ШІ-систем. Використання таких методів як вороже навчання, робастна оптимізація, регуляризація та нормалізація даних допомагає забезпечити певний рівень захисту, проте ці підходи ще не досконалі.

У майбутньому необхідно зосередити увагу на:

- адаптивних системах захисту, здатних працювати в режимі реального часу;
- розробці нових етичних і правових норм для захисту ШІ і його користувачів;
- комплексному підході до тестування систем ШІ в умовах реальних загроз.

Безпека штучного інтелекту є однією з найбільш критичних проблем сучасної науки та технологій, і її вирішення вимагатиме зусиль як інженерів, так і законодавців та дослідників у сфері етики.

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-takestucnii-intelekt-ta-yak-vin-transformuje-svit-xv7039#shho-take-stucnii-intelekt (дата звернення: 02.11.2024).

2. Харченко В., Неретін О. Забезпечення кібербезпеки систем штучного інтелекту: аналіз вразливостей, атак і контрзаходів. Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі. 2022. №12. С. 7-22. URL: https://science.lpnu.ua/sisn/all-volumes-and-issues/volume-12-2022/ensurance-artificial-intelligence-systems-cyber-security (дата звернення: 16.10.2024).

Відомості про авторів

Стацишина Ірина Павлівна, магістрантка кафедри комп'ютерних систем, мереж і кібербезпеки, НАУ «ХАІ», i.statcyshyna@student.csn.khai.edu

Клюшніков Ігор Миколайович, доцент кафедри комп'ютерних систем, мереж і кібербезпеки, НАУ «ХАІ», к.т.н., с.н.с., i.kliushnikov@csh.khai.edu