ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА КІБЕРБЕЗПЕКУ

Проценко Н.М., к.е.н., доц.
Державний біотехнологічний університет
м. Харків, Україна, pronatanic@gmail.com

Анотація: Сучасний світ все більше йде у віртуальну реальність і проблеми захисту особистої та комерційної інформації постають особливо гостро. Кількість кібератак зростає з кожним днем, як і стабільно збільшується число компаній, які переходять працювати в інтернет. Для здійснення безпеки інтересів громадянин, суспільства та держави застосовуються та розвиваються технології штучного інтелекту та машинного навчання.
Ключові слова: штучний інтелект, кіберпростір, кібербезпека, машинне навчання, лінгвістичні нейромережі, генеративні моделі

У сучасному світі люди постійно обмінюються інформацією в Інтернеті через електронну пошту, відеоконференції, акаунти у соціальних мережах тощо. Проте, Інтернет не лише надав можливість вченим, компаніям та різним організаціям стати більш ефективними, а й зумовив безпрецедентний обмін ідеями, інформацією та культурами серед раніше роз'єднаних людей та груп, зробив глобальну революцію у веденні бізнесу, взаємодії та спілкуванні. Але, з розвитком технологій та збільшенням числа підключених пристроїв, збільшуються ризики та загрози на основні системи управління, системи інтелектуалізації, тому кібербезпека стала однією з найважливіших проблем сучасного суспільства.

Вразливість у кіберпросторі є реальною, серйозною та швидко зростаючою. Торгівля та фінансові операції, комунікації, логістика, об'єкти інфраструктури особливої важливості все повністю залежать від інформаційно-технологічних систем, що об'єднані у мережі. Відповідно, що інструменти підтримки необхідного рівня кібербезпеки специфічні і залежать від систем, що розробляються (традиційних або вбудованих), вимагаючи адаптації під конкретні цілі та завдання.

Фахівці в галузі кібербезпеки швидко зрозуміли, що надійна та ефективна безпека означає постійне розуміння загроз, а не просто зосередженість на механізмах захисту системи та даних. Традиційні заходи безпеки, такі як брандмауери та антивірусні програми, є реактивними та можуть захистити лише від відомих загроз. Тому необхідно йти в ногу з характером загроз, що постійно змінюються.

Поява принципово нових викликів у сфері кібербезпеки одночасно супроводжувалося появою та широким впровадженням технологій штучного інтелекту (ШІ). По суті, багато в чому штучний інтелект і безпека були створені один для одного.

Перелік сфер, у яких використовується штучний інтелект, дуже широкий: промисловість (роботизація виробництва), торгівля (роботизація складського бізнесу), медицина (комп'ютерна діагностика, підбір методів лікування, кардіо-браслети, глюкометри та інші гаджети), транспорт (безпілотні автомобілі та літальні апарати), системах безпеки (розпізнавання осіб за допомогою

комп'ютера) зору), космічні дослідження (робот Curiosity, розроблений NASA, переміщався по поверхні Марса без зв'язку із Землею). Також, крім Google Translate, що за останнім оновленням отримав нові можливості, які стали реальністю завдяки ШІ, ці системи застосовуються в інтернет-помічниках Google Photos, Google Assistant. Варто відмітити, обсяг глобального ринку технологій штучного інтелекту постійно зростає. Нещодавній прогноз Міжнародної корпорації даних (IDC) показує, що світовий ринок програмного забезпечення штучного інтелекту (ШІ) зросте з $64 млрд у 2022 році до майже $251 млрд у 2027 році за річних темпів зростання [1].

Системи штучного інтелекту, побудовані на основі алгоритмів машинного навчання та лінгвістичних нейромереж. Класичні системи штучного інтелекту працюють із розпізнаванням вже відомих загроз. Насамперед добре виявляють проникнення: система може зреагувати на аномальну поведінку користувачів. автоматично розпізнати сигнатуру зловмисного коду або інтернет-трафіку, впоратися з фільтруванням спаму.

Однак, нові види загроз з'являються майже кожного дня, тому у коло першочергових завдань виходять питання щодо вміння передбачати зловмисні дії та реагувати на вразливість ще до того, як вони стануть активними. Впоратися з таким завданням здатні інструменти на основі генеративних моделей штучного інтелекту, які можуть створювати нові дані, зображення або текст на основі шаблонів та даних, на яких вони були навчені.

Якщо розглядати можливості генеративних моделей ШІ в площині кіберзахисту, відмітимо наступне: при виявленні нетипової поведінки система не тільки може відстежити критерії, з якими до того ще не стикалися, але й згенерувати моделі, які якнайшвидше відреагують на виявлену загрозу та імплементують систему захисту.

Великою перевагою генеративного штучного інтелекту є його здатність моделювати середовище, в якому імітація реальних сценаріїв дозволяє тестувати та оцінювати засоби контролю безпеки та застосовувати відповідні заходи. Це може допомогти виявити слабкі місця та підвищити загальну готовність системи захисту.

Також варто відмітити, що важливою складовою будь-яких «військових дій» виступає розвідка загроз. Можливість швидко аналізувати великі обсяги даних дозволяють генеративному ШІ виявляти закономірності та індикатори компрометації, які можна використовувати для виявлення загроз та реагування на них у режимі реального часу. Це створює певні умови для команд кібербезпеки бути на крок попереду загроз і швидко реагувати на атаки.

Однак, полегшуючи вирішення багатьох завдань, системи штучного інтелекту одночасно породжують нові загрози у сфері безпеки. Незважаючи на те, що штучний інтелект впроваджується повсюдно, проте з технологічної точки зору залишається слабким та недостатньо захищеним. Кіберзлочинці також використовують ШІ, щоб здійснювати більш складні та цілеспрямовані атаки.

Наприклад, поява автоматизованого алгоритму WormGPT, який був розроблений на основі мовної моделі JPT-J з відкритим вихідним кодом

спеціально для незаконної діяльності, допомагає шахраям генерувати переконливі спам-листи, які оминають систему спам-фільтрів [2, с. 134]. Для цього система використовує датасет бізнес-листів зі зламаних корпоративних поштових скриньок. Як наслідок, за останній рік збільшилася кількість інцидентів з фішингом, вірусами-здирниками тощо. При цьому кіберзлочинців не побільшало просто в них розширився діапазон охоплення своєї діяльності, можна надсилати не 100 тис. листів-спамів щомісяця, а 3 млрд. Генеративний інтелект дозволяє підробляти й голос іншої людини. У США щорічні втрати від таких атак сягають 20–27 млн доларів [3].

У зв’язку з тим, що в основі сучасного штучного інтелекту лежать моделі машинного навчання саме вони стають слабким місцем з погляду кібербезпеки.

Одним із поширених методів атаки є отруєння даних і моделей. У невелику кількість прикладів, що навчають, додається тригер – спеціально підготовлений фрагмент зображення. В результаті навчання на такому наборі даних модель стає отруєною. Передбачені отруєні моделі можуть поширюватися через Інтернет і нести загрозу при переносі знань [4]. Можливі крадіжки даних та моделей з хмарних середовищ [5].

Підводячи підсумки відмітимо наступне: штучний інтелект і, зокрема його різновид генеративний ШІ, мають величезний потенціал для перетворення сфери кібербезпеки, включаючи хмари, пристрої та навіть домашні системи безпеки. Створюючи прогностичні моделі, генеруючи симульовані середовища та аналізуючи великі обсяги даних, ці системи можуть допомогти виявляти загрози та реагувати на них до того, як вони завдадуть шкоди. Однак до використання штучного інтелекту слід підходити з обережністю з погляду правил, положень та моральних суджень.

Список літератури

1. IDC Forecasts Revenue for Artificial Intelligence. URL: https://www.idc.com/getdoc.jsp?containerId=prUS51345023#:~:text=NEEDHAM%2C%20Mass.%2C%20December%2020,(CAGR)%20of%2031.4%25.

2. Герус В. А. Хакерські угрупування. *Комп’ютерні технології: інновації, проблеми, рішення*: тези VI Всеукраїнської науково-технічної конференції (Житомир, 29-30 листопада 2023 р.). С. 134-138. URL: https://conf.ztu.edu.ua/wp-content/uploads/2024/01/povnyj-tekst-2.pdf.

3. Кольцов М. Штучний інтелект не захистить, якщо не використовувати інтелект природний»: як розвиток ШІ впливає на кібербезпеку. R_d media. URL: https://robotdreams.cc/uk/blog/352-shtuchniy-intelekt-ne-zahistit-yakshcho-ne-vikoristovuvati-intelekt-prirodniy-yak-rozvitok-shi-vplivaye-na-kiberbezpeku/.

4. Gu T., Dolan-Gavitt B., Garg S. BadNets: Identifying vulnerabilities in the machine learning model supply chain. 2017. arXiv preprint arXiv:1708.06733. URL: https://www.researchgate.net/publication/319235406_BadNets_Identifying_Vulnerabilities_in_the_Machine_Learning_Model_Supply_Chain.

5. Tramèr F., Zhang F., Juels A. et al. Stealing machine learning models via prediction APIs // Proceedings of the 25th USENIX Conference on Security Symposium (SEC'16). 2016. USENIX Association, USA. P. 601– 618. URL: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/tramer.