УДК 519.6

ШТУЧНИЙ ІНТЕЛЕКТ У КІБЕРБЕЗПЕЦІ: ВИКЛИКИ, ЗАГРОЗИ ТА РЕГУЛЯТОРНІ ВІДПОВІДІ

Здрок Дмитро

ARTIFICIAL INTELLIGENCE IN CYBERSECURITY: CHALLENGES, THREATS AND REGULATORY RESPONSES

Zdrok Dmytro

Волинський національний університет імені Лесі Українки, просп. Волі, 13, Луцьк, 43025, Україна

Abstract. *In 2025, artificial intelligence (AI) plays a vital role in cybersecurity. Its ability to learn, analyze vast data in real-time, and adapt to new threats makes it indispensable. While AI helps detect and neutralize attacks, it also enables new threats like prompt injection, deepfakes, and automated phishing. This paper reviews the current use of AI in cybersecurity, its pros and cons, and future directions including technical and ethical considerations.*

Штучний інтелект активно використовується в кібербезпеці для таких задач, як аналіз трафіку, виявлення аномалій, захист від DDoS-атак і фішингових кампаній. Наприклад, дослідження показують, що моделі XGBoost і Random Forest можуть досягати точності понад 99% при класифікації фішингових листів, що дозволяє своєчасно виявляти та блокувати підозрілу активність [1].

Також системи на основі глибокого навчання, такі як CyberSentinel, можуть аналізувати поведінкові патерни користувачів і знаходити нові атаки, які традиційні методи захисту не помічають. Це особливо важливо, адже кіберзагрози постійно змінюються [2].

Сучасні SIEM-системи (системи управління інформацією про безпеку та події), доповнені ШІ, не тільки моніторять події безпеки, але й автоматично реагують на них — наприклад, можуть ізолювати підозрілі процеси або тимчасово заблокувати доступ до системи. Впровадження технологій обробки природної мови (NLP) дозволяє системам аналізувати лог-файли, що написані звичним для людини форматом, і виявляти приховані загрози в текстових описах.

З розвитком ШІ з’являються й нові загрози. Наприклад, prompt injection — це метод маніпулювання великими мовними моделями, коли зловмисник вставляє шкідливі інструкції в запит, змушуючи модель генерувати чи розголошувати небажану інформацію. У 2025 році цей тип атаки був визнаний одним із найбільш небезпечних за класифікацією OWASP [3].

Ще один приклад — використання технологій deepfake в фішингових атаках. Зловмисники, застосовуючи ШІ, створюють відео, на яких імітуються голос і зовнішність керівників компаній, щоб обманом змусити працівників перерахувати кошти або передати конфіденційну інформацію.

Окремо варто згадати adversarial machine learning, коли атакуючий змінює вхідні дані системи, аби змусити модель ШІ прийняти хибне рішення. Наприклад, змінена графіка CAPTCHA може обдурити систему автентифікації, засновану на комп’ютерному зорі [4].

Також ШІ використовується для автоматизації кібератак. Боти, що працюють на основі навчання з підкріпленням, здатні тестувати захист мереж у реальному часі й адаптувати свої дії залежно від реакції системи. Це робить атаки не тільки масштабнішими, але й дедалі складнішими та витонченішими [5].

Зростання використання інструментів ШІ у сфері безпеки породжує багато етичних і правових питань. Одним із них є алгоритмічна упередженість, що може призвести до хибної ідентифікації користувачів або навіть дискримінації окремих груп. Також виникає потреба у поясненні рішень, прийнятих ШІ, що означає здатність моделі роз'яснити, чому вона прийняла те чи інше рішення.

У відповідь на ці виклики з'являються міжнародні ініціативи, як-от Мережа інститутів безпеки ШІ, що займається розробкою стандартів безпечного використання ШІ. Це дозволяє узгоджувати політики різних країн і сприяє глобальній стабільності у кібербезпеці.

У грудні 2024 року Європейська комісія прийняла AI Act — перший у світі нормативний акт, що класифікує системи ШІ за рівнем ризику. Системи, які використовуються в критичній інфраструктурі та безпеці, підлягають строгим вимогам щодо прозорості, аудиту та сертифікації. Україна також працює над адаптацією цього закону до свого правового поля в контексті євроінтеграції.

Ігрова індустрія, що стрімко розвивається, стала однією з головних мішеней для кіберзлочинців. Застосування штучного інтелекту (ШІ) у сфері кібербезпеки комп'ютерних ігор дозволяє ефективно протидіяти загрозам, таким як чітінг, фішинг, DDoS-атаки та шахрайство з транзакціями.ШІ-системи здатні аналізувати поведінку гравців у реальному часі, виявляючи аномальні дії, що можуть свідчити про використання стороннього програмного забезпечення або інших форм чітінгу. Наприклад, компанія Riot Games використовує ШІ для моніторингу гравців у грі League of Legends, що дозволяє своєчасно виявляти та блокувати порушників. ШІ може ефективно виявляти та нейтралізовувати DDoS-атаки, аналізуючи мережевий трафік та ідентифікуючи підозрілу активність. Компанія Blizzard Entertainment впровадила ШІ-системи, які дозволяють забезпечити стабільну роботу серверів навіть під час спроб масових атак [3].Ігри, що використовують ШІ, також застосовуються для навчання основам кібербезпеки. Наприклад, гра CyberCIEGE, розроблена за підтримки ВМС США, дозволяє гравцям моделювати різні сценарії атак та захисту, сприяючи кращому розумінню принципів кібербезпеки [4].

Також важливо інвестувати в освітні програми для фахівців з кібербезпеки, щоб вони могли розуміти, як працюють системи ШІ. Спільна робота технічних спеціалістів та експертів з етики й права стане ключем до створення ефективного захисту.Ще одним важливим напрямком є впровадження цифрових близнюків (Digital Twins) для симуляцій атак. Це дозволяє моделювати поведінку систем в різних сценаріях і тестувати їхню стійкість без шкоди для реальної інфраструктури. Такі симуляції, керовані ШІ, дозволяють не лише прогнозувати загрози, а й навчати захисні системи в безпечному середовищі [5].

Бібліографія

1. Atuncar Flores, E. J., Chuan García, A. F., Ráez Martínez, H. T., & Pachas Quispe, G. H. (2024). *Algorithms based on artificial intelligence for the detection and prevention of social engineering attacks: Systematic review*. LACCHEI.
2. Tallam, K. (2025). *CyberSentinel: An emergent threat detection system for AI security* [Preprint]. arXiv. https://arxiv.org/abs/2502.14966
3. Prompt injection. (n.d.). *Wikipedia*. Retrieved May 5, 2025, from https://en.wikipedia.org/wiki/Prompt_injection
4. ISC2. (2024). *Artificial intelligence and cybersecurity: Navigating the complexities of a data-driven future*.
5. Kulothungan, V. (2025). *Securing the AI frontier: Urgent ethical and regulatory imperatives for AI-driven cybersecurity* [Preprint]. arXiv. https://arxiv.org/abs/2501.10467