УДК 174.4:004.8

В.О. Тишківська

ЕТИЧНІ РИЗИКИ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В КІБЕРБЕЗПЕЦІ

Вінницький національний технічний університет

Анотація

У тезах розглянуто етичні ризики використання штучного інтелекту (ШІ) у сфері кібербезпеки. Зокрема, проаналізовано як позитивні аспекти впровадження ШІ — автоматизацію виявлення загроз, реагування на інциденти та імітацію атак для підвищення обізнаності працівників, так і потенційні етичні виклики. Основну увагу приділено проблемам упередженості даних, які можуть призвести до дискримінації, загрозам конфіденційності внаслідок обробки персональних даних, а також складнощам у визначенні відповідальності за автономні рішення ШІ. Наголошено на необхідності прозорості, постійного моніторингу, перевірки моделей та впровадження надійних механізмів контролю. А також наголошується, що розвиток технологій ШІ повинен супроводжуватись етичними стандартами, спрямованими на захист прав людини та забезпечення довіри до цифрових систем.

Ключові слова: етичні ризики, штучний інтелект, кіберзагрози, спам-фільтри.

Abstract

The thesis discusses the ethical risks of using artificial intelligence (AI) in the field of cybersecurity. In particular, both the positive aspects of AI implementation - automation of threat detection, incident response, and simulation of attacks to raise employee awareness - and potential ethical challenges are analysed. The article focuses on the problems of data bias that can lead to discrimination, privacy threats resulting from the processing of personal data, and difficulties in determining responsibility for autonomous AI solutions. The need for transparency, continuous monitoring, model validation, and implementation of reliable control mechanisms is emphasised. It is also emphasised that the development of AI technologies should be accompanied by ethical standards aimed at protecting human rights and ensuring trust in digital systems.

Key words: ethical risks, artificial intelligence, cyber threats, spam filters.

Вступ

У сучасному цифровому світі, де обсяги даних постійно зростають, а кіберзагрози стають дедалі витонченішими, штучний інтелект відіграє все більш важливу роль у забезпеченні кібербезпеки. Завдяки здатності швидко обробляти великі масиви інформації, виявляти аномалії та адаптуватися до нових типів атак, ШІ стає потужним інструментом як для захисту інформаційних систем, так і для зловмисників. Водночас поява та впровадження технологій ШІ викликають низку етичних питань, зокрема щодо конфіденційності даних, прозорості алгоритмів, автономності прийняття рішень і потенційної дискримінації. У сфері кібербезпеки, де навіть незначна помилка може мати серйозні наслідки, етичні ризики використання ШІ набувають особливого значення. У цьому контексті постає необхідність критичного осмислення не лише технічних, а й морально-правових аспектів використання штучного інтелекту.

Результати дослідження

Штучний інтелект швидко змінює світ, відкриваючи нові горизонти для бізнесу, науки та суспільства[1]. Етичні питання, пов'язані зі штучним інтелектом та конфіденційністю даних, стають дедалі важливішими, оскільки він продовжує ставати все більш актуальним у різних аспектах нашого повсякденного життя (наприклад спам-фільтри на базі ШІ, щоб захистити себе від шкідливих електронних листів). В основі цих етичних питань лежить баланс між використанням потенціалу штучного інтелекту для покращення кібербезпеки та захистом прав людей на конфіденційність. Збір, зберігання та використання даних повинні диктуватися прозорістю та принципом інформованої згоди, гарантуючи, що люди знають, як і з якою метою використовуються їхні дані.

Проте, попри безліч переваг, широке впровадження штучного інтелекту супроводжується низкою етичних та технічних проблем. Вкрай важливо переконатися, що алгоритми ШІ ненавмисно не створюють нових вразливостей або упереджень, які можуть посилити існуючі ризики або створити нові. Передусім, алгоритми ШІ навчаються на наявних даних, які можуть містити упередження — зокрема, за географічними, національними чи поведінковими ознаками [1]. Такі упередження можуть призвести до помилкової ідентифікації підозрілої активності або навіть дискримінаційних дій. Крім того, у разі, якщо ШІ навчається на даних із вже скомпрометованої мережі, він може помилково вважати шкідливу активність нормальною, що спотворює його здатність ефективно захищати системи [2]. Системи штучного інтелекту також потребують доступу до великих обсягів даних, серед яких можуть бути конфіденційні або персональні. Без належного захисту це створює ризик порушення конфіденційності [3].

Застосування ШІ у кібербезпеці також несе ризики, пов'язані з автономним ухваленням рішень. Наприклад, автоматичне блокування IP-адрес або ізоляція файлів може відбуватися без участі людини. У разі помилки постає питання відповідальності: хто винен — фахівець з кібербезпеки, розробник, чи вся організація [4]? Додатково, неетичне або непрозоре використання ШІ, зокрема без чіткої інформованої згоди користувачів на обробку їхніх даних, може підірвати довіру до таких технологій [5].

Щоб уникнути цих ризиків, необхідно впроваджувати низку рішень. Передусім, важливо забезпечити прозорість збору та використання даних, дотримуючись принципу інформованої згоди. Організації повинні ретельно перевіряти якість навчальних даних, гарантувати їхню чистоту, точність і репрезентативність. Необхідно впроваджувати постійний процес оновлення та перенавчання систем ШІ, який враховуватиме нові типи загроз, атаки та поведінкові зміни в мережах. Важливо також здійснювати регулярний моніторинг моделей, усуваючи упередження й вразливості, які можуть виникнути внаслідок зміни середовища або дій зловмисників [2]. Окрім технічних заходів, потрібне також етичне регулювання — застосування ШІ має бути не лише ефективним, а й відповідальним [5].

Лише після усунення етичних та технічних ризиків – зокрема упередженості даних, порушення конфіденційності, автономних помилкових рішень та забезпечення цілісності навчання – можна буде стверджувати, що використання ШІ в кібербезпеці є справді безпечним, ефективним і відповідальним.

Висновки

Штучний інтелект має величезний потенціал для зміцнення кібербезпеки, забезпечуючи ефективні інструменти для виявлення загроз, автоматизації реагування та навчання персоналу. Водночас його використання породжує низку етичних викликів, пов'язаних із конфіденційністю, упередженістю алгоритмів, якістю даних та розподілом відповідальності. Особливу небезпеку становить використання ШІ зловмисниками, що вимагає постійного вдосконалення захисних систем. Для безпечного й етичного впровадження ШІ в кібербезпеку необхідно дотримуватися принципів прозорості, відповідального управління даними та безперервного моніторингу систем. Лише за таких умов можливо зберегти довіру до технологій та забезпечити реальний захист цифрового середовища.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Етика штучного інтелекту: виклики та можливості [Електронний ресурс] // Agiliway - Custom Software Development Company. – Режим доступу: https://www.agiliway.com/uk-ua/etyka-shtuchnoho-intelektu-vyklyky-ta-mozhlyvosti/ (дата звернення: 30.05.2025). – Назва з екрана.
2. Ethical implementation of AI in cybersecurity | evolve security [Електронний ресурс] // Offensive Security Solutions | Evolve Security. – Режим доступу: https://www.evolvesecurity.com/blog-posts/ethical-implementation-of-ai-in-cybersecurity (дата звернення: 30.05.2025). – Назва з екрана.
3. Ethical issues in AI cybersecurity [Електронний ресурс] // Redress Compliance - Just another WordPress site. – Режим доступу: https://redresscompliance.com/ethical-issues-ai-cybersecurity/ (дата звернення: 30.05.2025). – Назва з екрана.
4. The ethical dilemmas of AI in cybersecurity [Електронний ресурс] // Cybersecurity Certifications and Continuing Education | ISC2. – Режим доступу: https://www.isc2.org/Insights/2024/01/The-Ethical-Dilemmas-of-AI-in-Cybersecurity (дата звернення: 30.05.2025). – Назва з екрана.
5. Owen-Jackson C. Navigating the ethics of AI in cybersecurity | IBM [Електронний ресурс] / Charles Owen-Jackson // IBM - United States. – Режим доступу: https://www.ibm.com/think/insights/navigating-ethics-ai-cybersecurity (дата звернення: 30.05.2025). – Назва з екрана.

Тишківська Вероніка Олександрівна – студентка факультету менеджменту та інформаційної безпеки, Вінницький національний університет, м. Вінниця, електронна пошта: tiskivskaveronika@gmail.com

Науковий керівник: ***Зоря Ірина Сергіївна*** – асистент, магістр з кібербезпеки, Вінницький національний технічний університет, м. Вінниця, електронна пошта: ira.zoria@vntu.edu.ua

Tyshkivska Veronika O. – Department of Information Systems Management and Security, Vinnytsia National Technical University, Vinnytsia, e-mail : tiskivskaveronika@gmail.com

Supervisor: ***Zorya Iryna S.*** – assistant, master in Cybersecurity, Vinnytsia National Technical University, Vinnytsia, e-mail : ira.zoria@vntu.edu.ua