УПРАВЛІННЯ ІНФОРМАЦІЙНИМИ РИЗИКАМИ З ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ

Золотарьов В.А., Кабаченко В.О.

Харківський національний університет радіоелектроніки,
кафедра інформаційно-мережної інженерії,
Україна.

E-mail: vadym.zolotarov@nure.ua,
viacheslav.kabachenko@nure.ua

Abstract

The management of information risks in the context of artificial intelligence (AI) has become increasingly critical in today's digital landscape. This paper explores the integration of AI technologies in information risk management, focusing on their potential to enhance risk identification, assessment, and mitigation processes. By leveraging machine learning algorithms and predictive analytics, organizations can proactively identify vulnerabilities and minimize the impact of potential threats. Additionally, the study examines the ethical implications and challenges associated with AI in this realm, including data privacy concerns and algorithmic biases. Through a comprehensive analysis of case studies and current practices, this research aims to provide insights into effective strategies for utilizing AI in managing information risks, thereby fostering a more resilient information security environment. The findings underscore the importance of developing robust frameworks that not only embrace technological advancements but also prioritize ethical considerations in information risk management.

Використання штучного інтелекту (ШІ) не тільки суттєво покращує всі сфери життя людства, але і породжує нові, невідомі раніше виклики його інформа-ційній безпеці.

Вчені Массачусетського технологічного інституту (MIT) у серпні 2024 р. завершили розробку «Репозиторію ризику ШІ». Документ являє собою систематизований каталог існуючих ризиків використання ШІ, в якому докладно описуються умови і причини інформаційних загроз. Передбачається, що база даних буде постійно оновлюватися відразу після виявлення нових інформаційних ризиків. На сьогодні описано понад 700 ризиків, які розділені на 7 типів: дискримінація і токсичність; конфіденційність і безпека; дезінфор-мація; зловмисники та зловживання; взаємодія людини та комп'ютера; соціально-економічний і екологічній збиток; безпека, недоліки та обмеження систем ШІ [1].

На думку вітчизняних науковців основні загрози використання ШІ наці-ональній безпеці держави в інформаційно-комунікаційній сфері включають:

- підвищення ефективності виявлення вразливостей в системах захисту мереж і здійснення кібератак;
- утворення штучних каналів витоку конфіденційної інформації з мереж і маскування їхнього впровадження;
- покращення ефективності цільових атак отруєння даних (Data Poi-soning) для модифікації і спотворення даних;
- містифікація даних – створення недостовірних (галюцінованих) фак-тів [2].

Серед основних загроз використання ШІ інформаційній безпеці організацій фахівці називають: підпільну розробку великих мовних моделей (LLM); оновлення «Script Kiddies»; голосове шахрайство для соціальної інженерії [3].

Управління інформаційними ризиками використання ШІ в інформаційно-комунікаційних мережах слід здійснювати по трьом напрямкам: нормативно-правовим; адміністративно-організаційним і програмно-технічним [2].

Нормативно-правові заходи передбачають, насамперед, прийняття державної політики застосування ШІ та створення законодавчих норм розроблення і використання ШІ в Україні, зокрема у за-

кладах вищої освіти. Міністерство освіти України і кожний університет мають визначити сферу використання ШІ в навчальному процесі.

Видобування ознак та попередня обробка є невід'ємною частиною [2], але вибір методів та алгоритмів машинного навчання значущо впливає на продуктивність системи розпізнавання голосу. Глибоке розуміння цих технік є важливим для ефективної розробки систем розпізнавання голосу.

Для цифрової обробки сигналів застосовується дискретне перетворення Фур'є (Discrete Fourier Transform, DFT) до періодично дискретизованої часової області [3]. Тут будь-який цифровий звук f (сигнал, який може бути розкладений на різні компоненти частот) є сумою чистих звуків на різних частотах. Застосування DFT до неперервних сигналів включає кілька кроків, як показано на рисунку

Адміністративно-організаційні заходи полягають, по-перше, у створенні ефективної моделі управління ризиками використання ШІ, до розроблення якої мають бути залучені не тільки представники СБУ, ДСТЗІ і Кіберполіції, але й представники бізнесу та Міністерства цифрової трансформації України. Модель має постійно оновлюватися та удосконалюватися. По-друге, керівництво установ має проводити постійне навчання персоналу навичкам роботи з ШІ і протидії інформаційним ризикам. По-третє, має бути запроваджене обґрунтоване обмеження доступу ШІ до критичної інформації.

Програмно-технічні заходи мають: допомагати здійснювати розроб-лення та впровадження безпечних систем штучного інтелекту з вбудованими засобами захисту; встановлювати програмно-технічні запобігання використання ШІ для доступу до конфіденційної інформації, особливо до персональних даних; створення систем ШІ для аудиту знань прикладних систем ШІ; розроблення механізмів оперативного виявлення та активної протидії потенційно небезпечним системам штучного інтелекту.

Слід пам’ятати, що ШІ має стати дієвим помічником людського розуму, а не його замінником і сліпо довіряти йому не можна. Ключовим етапом у системі управління ризиками при використанні є оцінювання потенційного ризикового середовища та їх ідентифікація. Цей процес є ітеративним і передбачає виявлення нових видів ризиків та аналіз їхніх основних характеристик для подальшої інтерпретації, аналізу та обробки.

Література

1. What are the risks from Artificial Intelligence? [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://airisk.mit.edu/

2. Скіцько О., Складанний П., Ширшов Р., Гуменюк М., Ворохоб М. Загрози та ризики використання штучного інтелекту / Кібербезпека: освіта, наука, техніка. 2023, № 2 (22), с. 6-18

3. Легомінова С.В., Гайдур Г.І. Аналіз сучасних загроз інформаційній безпеці організацій та формування інформаційної платформи протидії їм / Кібербезпека: осві-та, наука, техніка. 2023, № 2 (22), с. 54-67