8. Volatility Foundation. Volatility 3 Windows Plugins Documentation. URL: https://volatility3.readthedocs.io/en/stable/volatility3.plugins.windows.html (дата звернення: 15.08.2026).

9. Harvey P. ExifTool; ReFirm Labs. Binwalk; zed-0xff. zsteg. URL: https://exiftool.org/ExifTool.html ; https://github.com/ReFirmLabs/binwalk ; https://github.com/zed-0xff/zsteg (дата звернення: 15.08.2026).

10. Aleph One. Smashing The Stack For Fun And Profit // Phrack Magazine. 1996. Vol. 7, Issue 49. URL: https://phrack.org/issues/49/14_md

11. Shacham H. The Geometry of Innocent Flesh on the Bone: Return-into-libc without Function Calls (on the x86) // Proceedings of the 14th ACM Conference on Computer and Communications Security. 2007. P. 552-561. DOI: 10.1145/1315245.1315313.

УДК 004.8:316.77

Бубенцова Л.В., Царьов Р.Ю.

ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ З ВИСОКИМ РІВНЕМ РИЗИКУ

Державний університет інтелектуальних технологій і зв'язку
ceis@suitt.edu.ua

Анотація. *У роботі досліджено особливості систем штучного інтелекту з високим рівнем ризику. Проаналізовано основні сфери застосування таких систем, потенційні ризики їх використання та ключові механізми їх мінімізації. Особливу увагу приділено ризик-орієнтованому підходу до регулювання штучного інтелекту, закріпленому в Регламенті Європейського Союзу 2024/1689 (Artificial Intelligence Act). Визначено основні проблеми, пов'язані з алгоритмічною дискримінацією, якістю даних, непрозорістю алгоритмів, захистом персональних даних, кібербезпекою та недостатнім людським контролем. Обґрунтовано необхідність комплексного управління ризиками протягом усього життєвого циклу високоризикової системи штучного інтелекту.*

Ключові слова: *штучний інтелект, високоризикові системи, оцінювання ризиків, AI Act, алгоритмічна дискримінація, людський нагляд, кібербезпека, захист даних.*

Стрімкий розвиток технологій штучного інтелекту (ШІ) є однією з визначальних тенденцій сучасної цифрової трансформації. Системи ШІ активно впроваджуються в економіці, промисловості, освіті, медицині, державному управлінні, транспорті, фінансовій сфері та інших галузях. Використання алгоритмів машинного навчання та інших технологій ШІ дає змогу автоматизувати значну кількість процесів, здійснювати аналіз великих обсягів інформації та підтримувати прийняття управлінських рішень.

Водночас розширення сфери застосування ШІ зумовлює появу нових технологічних, правових та соціальних ризиків. Особливої уваги потребують системи, результати роботи яких можуть істотно впливати на життя, здоров'я, безпеку або фундаментальні права людини. У таких випадках технічна помилка алгоритму, некоректні дані або неправильно сформована модель можуть мати значно серйозніші наслідки, ніж при використанні ШІ для виконання звичайних інформаційних завдань.

У зв'язку з цим у сучасному регулюванні ШІ формується ризик-орієнтований підхід, відповідно до якого вимоги до систем визначаються залежно від потенційного рівня шкоди. Одним із найбільш системних прикладів такого підходу є Регламент Європейського Союзу 2024/1689 — Artificial Intelligence Act (AI Act) [1].

Метою роботи є аналіз особливостей систем штучного інтелекту з високим рівнем ризику; визначення основних сфер їх застосування, потенційних загроз та механізмів їх мінімізації, вимог до забезпечення безпечного та відповідального використання таких систем та напрямів їх подальшого розвитку.

Класифікація високоризикових систем штучного інтелекту

Відповідно до статті 6 Регламенту (ЄС) 2024/1689 (Artificial Intelligence Act, AI Act), високоризикові системи штучного інтелекту поділяються на дві основні категорії [1,2].

До першої належать системи, які є компонентами безпеки продукції або самі є продукцією, що підпадає під визначене законодавство Європейського Союзу щодо гармонізації, за умови обов'язкового проведення оцінювання відповідності із залученням третьої сторони. Наприклад, системи, пов'язані з обладнанням для потенційно вибухонебезпечних середовищ, радіообладнанням, медичними виробами, автомобільним та авіаційним транспортом. Високий рівень ризику таких систем обумовлений тим, що помилка або неналежне функціонування системи може безпосередньо вплинути на безпеку людини.

До другої категорії належать системи, які використовуються у визначених суспільно значущих сферах. До них належать (табл.1):

Таблиця 1 - Системи, які використовуються у визначених суспільно значущих сферах

Сфера	Приклади застосування ШІ
Біометрія	дистанційна біометрична ідентифікація, біометрична категоризація
Критична інфраструктура	управління та експлуатація критичної інфраструктури
Освіта	оцінювання, вступ, визначення рівня знань
Зайнятість	підбір працівників, прийняття кадрових рішень, оцінювання працівників
Основні приватні та державні послуги	оцінювання права на отримання послуг, кредитна оцінка
Правоохоронна діяльність	оцінювання ризиків, аналіз доказів
Міграція та контроль кордонів	аналіз заяв, перевірка документів, біометрична ідентифікація
Правосуддя та демократичні процеси	допомога судовим органам, окремі системи, пов'язані з виборчими процесами

Основні ризики використання високоризикових систем та механізми їх мінімізації

Одним із найбільш суттєвих ризиків є **ризик дискримінації та упередженості**. Системи ШІ значною мірою залежать від якості даних, на яких вони навчаються та функціонують. Якщо вихідні дані містять історичні або соціальні упередження, алгоритм може не тільки відтворювати їх, а й масштабувати на велику кількість користувачів. Особливо небезпечною така проблема стає у сфері працевлаштування, освіти, кредитування або надання соціальних послуг. Наприклад, автоматизоване ранжування кандидатів чи оцінювання їхніх характеристик під час відбору на роботу може призвести до необґрунтованого обмеження можливостей певних груп осіб через статистичні особливості історичних даних, навіть якщо така дискримінація безпосередньо не закладена розробником.

Другим важливим ризиком є непрозорість алгоритмічних рішень**.** Складні моделі машинного навчання можуть формувати результати, причини яких важко пояснити навіть розробникам системи. Проблема «чорної скриньки» набуває особливого значення тоді, коли рішення ШІ безпосередньо впливає на права та інтереси людини. Якщо користувач не має можливості зрозуміти логіку використання системи та оцінити отриманий результат, виникає проблема відповідальності за прийняте рішення.

Важливим фактором ризику є якість даних**.** Дані повинні бути релевантними, достатньо репрезентативними та належним чином організованими.

Окрему групу становлять ризики, пов'язані із захистом даних і кібербезпекою**.** Високоризикові системи можуть бути об'єктами цілеспрямованих атак, спрямованих на зміну їхньої поведінки або результатів роботи. Вони можуть обробляти значні обсяги персональної інформації. Несанкціонований доступ, підміна даних, маніпулювання вхідною інформацією або втручання в алгоритм можуть призвести до неправомірних чи небезпечних результатів.

Слід виділити проблему надмірної довіри до автоматизованих рішень**.** Користувач може сприймати результат роботи системи як об'єктивний лише тому, що його сформував алгоритм. Проте алгоритм не гарантує абсолютної правильності рішення. Тому людський контроль є необхідною складовою безпечного використання високоризикових систем.

Ще одним важливим аспектом є технічна надійність системи**.** Високоризикова система повинна забезпечувати належний рівень точності, стійкості та кібербезпеки протягом усього життєвого циклу. Система має бути здатною коректно функціонувати в умовах помилок, несправностей або непередбачених ситуацій. /

Результати аналізу основних потенційних ризиків високоризикових систем [1] подані у табл.2.

Таблиця 2 - Системи, які використовуються у визначених суспільно значущих сферах та основні потенційні ризики

Сфера	Приклади застосування ШІ	Основні потенційні ризики
Біометрія	дистанційна біометрична ідентифікація, біометрична категоризація	помилкова ідентифікація, дискримінація, порушення приватності
Критична інфраструктура	управління та експлуатація критичної інфраструктури	загроза безпеці людей та безперервності життєво важливих послуг
Освіта	оцінювання, вступ, визначення рівня знань	необ'єктивність оцінювання, дискримінація, обмеження доступу до освіти
Зайнятість	підбір працівників, прийняття кадрових рішень, оцінювання працівників	дискримінація, непрозорість кадрових рішень
Основні приватні та державні послуги	оцінювання права на отримання послуг, кредитна оцінка	необгрунтована відмова у послугах, фінансові та соціальні ризики
Правоохоронна діяльність	оцінювання ризиків, аналіз доказів	помилкові висновки, порушення прав людини
Міграція та контроль кордонів	аналіз заяв, перевірка документів, біометрична ідентифікація	дискримінація, помилки у прийнятті рішень
Правосуддя та демократичні процеси	допомога судовим органам, окремі системи, пов'язані з виборчими процесами	вплив на фундаментальні права та демократичні процеси

Згідно з ризик-орієнтованою класифікацією, запропонованою AI Act [1], високоризиковою вважається не технологія ШІ сама по собі, а система, яка використовується у визначеному контексті та потенційно може завдати істотної шкоди здоров'ю, безпеці або фундаментальним правам людини. Ризик-орієнтоване регулювання дозволяє встановлювати різні вимоги залежно від потенційної небезпеки системи і високоризикові системи можуть застосовуватись за умови виконання визначених вимог безпеки, прозорості та відповідальності. Важливим інструментом мінімізації ризиків є система управління ризиками. Відповідно до статті 9 AI Act, вона повинна бути встановлена, впроваджена, документована та підтримуватися протягом усього життєвого циклу високоризикової системи. Управління ризиками має мати безперервний ітеративний характер і включати ідентифікацію відомих та передбачуваних ризиків, їх оцінювання, визначення відповідних заходів зменшення ризику та регулярний перегляд ефективності таких заходів.

Для зменшення ризиків дискримінації та упередженості в такій системі навчальні, валідаційні та тестові набори даних повинні відповідати призначенню, бути належно сформованими та перевірятися на наявність упередженості. Процеси збирання, підготовки, маркування та оброблення даних мають бути належним чином організовані та задокументовані. Для мінімізації ризику неправильного **або** непрозорого прийняття рішень AI Act встановлює вимоги щодо прозорості, технічної документації та ведення записів [3]**.** Документація повинна містити інформацію, необхідну для оцінювання відповідності системи встановленим вимогам, а експлуатаційні журнали мають забезпечувати можливість відстеження функціонування системи. Це створює основу для аудиту та контролю її роботи. Людський контроль є одним із ключових механізмів зменшення ризику надмірної довіри до автоматизованих рішень [4]**.** Для забезпечення технічної надійності системи можуть застосовуватися резервні механізми, дублювання компонентів та процедури безпечного припинення роботи. Забезпечення вимог кібербезпеки повинно враховувати специфічні властивості технологій штучного інтелекту.

Особливістю регулювання високоризикових систем ШІ відповідно до AI Act є перехід від реактивного підходу до превентивного управління ризиками [5]**.** Безпека, неупередженість, прозорість, людський контроль, якість даних, стійкість і кібербезпека повинні враховуватися не як окремі елементи після створення системи, а як складові єдиного процесу її життєвого циклу. Такий підхід дозволяє підвищити надійність систем ШІ та одночасно забезпечити належний рівень захисту фундаментальних прав людини.

Ризик-орієнтований підхід як основа безпечного використання штучного інтелекту

Ризик-орієнтоване регулювання дозволяє встановлювати різні вимоги залежно від потенційної небезпеки технології. Такий підхід має принципову відмінність від загальної заборони використання ШІ: високоризикові системи можуть застосовуватися за умови виконання визначених вимог безпеки, прозорості та контролю. При цьому необхідно відмежовувати високоризикові системи від практик використання ШІ, які вважаються неприйнятними та підлягають забороні. AI Act встановлює окремий перелік заборонених практик, пов'язаних, зокрема, з маніпулятивними технологіями, експлуатацією вразливостей людини та іншими видами використання ШІ, що створюють неприйнятний рівень ризику [1,6].

Для України досвід Європейського Союзу має особливе значення. Розвиток національної системи регулювання ШІ потребує врахування європейських принципів безпеки, прозорості, відповідальності та захисту фундаментальних прав. Це важливо не лише з погляду правового наближення до ЄС, а й для створення умов довіри до цифрових технологій. Провідні українські IT-компанії (зокрема Grammarly, MacPaw, LetsData та інші) вже підписали добровільну Декларацію про саморегулювання та етичне використання ШІ [7].

У березні 2026 року Україна отримала від Європейської Комісії умови вступу за всіма переговорними кластерами. Однією із таких умов є гармонізація законодавства у сфері штучного інтелекту. Крім цього, у квітні цього року було затверджено Національну програму адаптації законодавства України до права Європейського Союзу, якою визначено завдання з розробки Закону України про штучний інтелект [8]. Саме тому в межах подальшої євроінтеграції Україна має формувати власну нормативно-правову базу у цій сфері.

Водночас завдання нашої держави полягає у створенні умов для розвитку інновацій в Україні, залученні інвестицій, підтримці українських розробників та впровадженні сучасних технологій в публічному управлінні та бізнесі.

Одним із перспективних напрямів є запровадження процедур оцінювання впливу ШІ на права людини**,** алгоритмічного аудиту та постійного моніторингу систем після їх упровадження. Особливу увагу необхідно приділяти системам, що використовуються у державному секторі та можуть впливати на доступ громадян до суспільно важливих послуг.

Висновки. Проведений аналіз свідчить, що високоризикові системи штучного інтелекту мають специфічні особливості, пов'язані насамперед із потенційно значним впливом результатів їх роботи на людину та суспільство.

Основними ризиками є алгоритмічна дискримінація**,** недостатня якість даних, непрозорість прийняття рішень, порушення конфіденційності, кіберзагрози, технічні помилки та надмірна залежність користувача від автоматизованих рекомендацій.

Забезпечення безпечного використання таких систем потребує комплексного підходу, який включає управління ризиками протягом усього життєвого циклу системи, належне управління даними, технічне документування, тестування, моніторинг, забезпечення точності та кібербезпеки, а також ефективний людський нагляд.

Досвід Європейського Союзу, закріплений у AI Act, демонструє перехід від загальних етичних рекомендацій до конкретних нормативних вимог щодо розроблення та використання високоризикового ШІ. Для України врахування цього досвіду є важливим напрямом формування сучасної системи регулювання штучного інтелекту.

Отже, подальший розвиток високоризикових систем ШІ має здійснюватися за принципом «інновації за умови безпеки, прозорості та відповідальності». Ефективність технології не може бути єдиним критерієм її впровадження. Не менш важливими є захист прав людини, можливість реального людського контролю та чітке визначення відповідальності за наслідки використання автоматизованих систем.

Література

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. 2024. URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj (дата звернення: 05.07.2026).
2. European Commission. AI Act Service Desk. Article 6: Classification rules for high-risk AI systems. URL: https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-6 (дата звернення: 05.07.2026).
3. European Commission. AI Act Service Desk. Article 10: Data and data governance. URL: https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-10 (дата звернення: 11.07.2026).
4. European Commission. AI Act Service Desk. Article 14: Human oversight. URL: https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-14 (дата звернення: 08.08.2026).
5. Sousa e Silva N. The Artificial Intelligence Act: critical overview. 2024. URL: https://arxiv.org/abs/2409.00264 (дата звернення: 08.08.2026).
6. Kelly J., Zafar S., Heidemann L. et al. Navigating the EU AI Act: A Methodological Approach for Safety-critical Products. 2024. URL: https://arxiv.org/abs/2403.16808 (дата звернення: 08.08.2026).
7. Українські IT-компанії розроблять правила використання ШІ в технологічних продуктах URL: https://thedigital.gov.ua/news/business/ukrainski-it-kompanii-rozroblyat-pravila-vikoristannya-shi-v-tekhnologichnikh-produktakh (дата звернення: 09.08.2026).
8. Чіткі правила регулювання штучного інтелекту: у Комітеті з питань цифрової трансформації відбулося засідання робочої групи. URL: https://www.rada.gov.ua/news/news_kom/273621.html (дата звернення: 10.08.2026)

УДК 004.056:004.8:62-5

Хлапонін Ю.І.

ГУМАНОЇДНІ РОБОТИ В ЦИФРОВІЙ ЕКОСИСТЕМІ SMART CITY: МОЖЛИВОСТІ, ТЕХНОЛОГІЇ ТА КІБЕРБЕЗПЕКОВІ РИЗИКИ

Державний торговельно-економічний університет
y.khlaponin@knute.edu.ua

Анотація. *Розглянуто перспективи інтеграції гуманоїдних роботів у цифрову екосистему Smart City як нового класу інтелектуальних кіберфізичних систем. Проаналізовано основні технології їх функціонування, зокрема штучний інтелект, сенсорне сприйняття, автономну навігацію та мережеву взаємодію. Визначено ключові проблемні питання безпеки: фізичну безпеку людино-роботичної взаємодії, кібербезпеку програмно-апаратної інфраструктури, захист даних і приватність, а також безпеку систем штучного інтелекту. Обґрунтовано необхідність комплексного підходу до забезпечення кіберстійкості гуманоїдних роботів у міському цифровому середовищі.*

Ключові слова*: гуманоїдний робот, Smart City, кіберфізична система, штучний інтелект, кібербезпека, кіберстійкість, робототехніка, цифрова міська інфраструктура.*

Стрімкий розвиток робототехніки та штучного інтелекту зумовлює перехід гуманоїдних роботів від переважно експериментальних систем до технологій, орієнтованих на практичне застосування. Сучасні підходи до розвитку інтелектуальних роботів ґрунтуються на поєднанні фізичного втілення, сенсорного сприйняття, штучного інтелекту та здатності автономно взаємодіяти з навколишнім середовищем. Концепція embodied intelligence розглядає інтелект робота як результат взаємодії алгоритмів, фізичної конструкції та середовища функціонування [1].

Одним із перспективних напрямів застосування гуманоїдних роботів є цифрове міське середовище. Концепція Smart City передбачає інтеграцію інформаційно-комунікаційних технологій, IoT, цифрових