РОЗДІЛ ІІ. ІНФОРМАЦІЙНО-КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

DOI: https://doi.org/10.25140/2411-5363-2026-1(43)-90-104
УДК 004.8:004.056:34:330.46

Юлія Миколаївна Ткач[1], Владислав Олександрович Одноколов[2], Тарас Анатолійович Петренко[3]

[1]доктор педагогічних наук, кандидат технічних наук, професор,
завідувач кафедри кібербезпеки та математичного моделювання
Національний університет «Чернігівська політехніка» (Чернігів, Україна)
E-mail: tkachym79@gmail.com. **ORCID:** https://orcid.org/0000-0002-8565-0525

[2]студент-магістр кафедри кібербезпеки та математичного моделювання
Національний університет «Чернігівська політехніка» (Чернігів, Україна)
E-mail: olexodno@gmail.com. **ORCID:** https://orcid.org/0009-0004-8157-5259

[3]кандидат технічних наук, доцент кафедри кібербезпеки та математичного моделювання
Національний університет «Чернігівська політехніка» (Чернігів, Україна)
E-mail: mail_taras@stu.cn.ua. **ORCID:** https://orcid.org/0000-0001-5571-3815

РИЗИКИ ВПРОВАДЖЕННЯ ШТУЧНОГО ІНТЕЛЕКТУ: БЕЗПЕКОВІ, ПРАВОВІ ТА СОЦІАЛЬНО-ЕКОНОМІЧНІ АСПЕКТИ

У статті здійснено дослідження та систематизацію ризиків інтеграції штучного інтелекту в соціально-економічні та безпекові процеси в умовах цифрової трансформації суспільства. Проаналізовано багатовимірний характер впливу технологій штучного інтелекту на інформаційний простір, ринок праці, захист персональних даних, економічні відносини та національну безпеку. Узагальнено та структуровано основні групи ризиків, пов'язаних із використанням систем штучного інтелекту, зокрема технологічні, правові, етичні, соціальні та безпекові. Розглянуто механізми реалізації зазначених ризиків і потенційні наслідки їх прояву для державних інституцій, бізнес-середовища та суспільства. Окрему увагу приділено міжнародним підходам до регулювання штучного інтелекту, зокрема ризик-орієнтованим моделям, що застосовуються в Європейському Союзі, а також визначено концептуальні напрями мінімізації та регулювання ризиків інтеграції штучного інтелекту.

***Ключові слова**: штучний інтелект; ризики; цифрова трансформація; інформаційна безпека; регулювання штучного інтелекту.*
Рис.: 4. Бібл.: 15

Актуальність теми дослідження. У сучасних умовах штучний інтелект перестав бути лише об'єктом академічних досліджень і перетворився на базову інфраструктурну технологію, від якої залежить повсякденна діяльність мільйонів людей. Це стосується не тільки складних наукових чи військових систем. Це також про звичайні сервіси, якими користуються щодня. Наприклад, стрічка рекомендацій у соцмережах, фільтри спаму, голосові помічники, навігація та інтернет-банкінг. Саме тому будь-які збої, викривлення або зловмисне використання алгоритмів ШІ неминуче набувають суспільного виміру й виходять далеко за рамки суто технічної проблематики.

Як зазначається в класичній монографії С. Рассела та П. Норвіга «Artificial Intelligence: A Modern Approach», а також у Рекомендації Ради Організації економічного співробітництва та розвитку щодо штучного інтелекту (OECD Recommendation of the Council on Artificial Intelligence) і Рекомендації ЮНЕСКО з етики штучного інтелекту (UNESCO Recommendation on the Ethics of Artificial Intelligence) [1; 2; 3], стрімкий розвиток технологій штучного інтелекту (ШІ) є одним із ключових драйверів цифрової трансформації економіки, державного управління та повсякденного життя. Системи машинного навчання вже використовують у банках, медицині, транспорті, освіті, медіа та обороні. Разом із розвитком ШІ з'являються нові ризики. Вони стосуються прав людини, кібербезпеки, інформаційної незалежності та стабільності в суспільстві. Тому потрібно аналізувати ці загрози й мати план дій. Потрібні рішення на кількох рівнях. Це закони, технічний захист, робота організацій і освіта.

Важливо пам'ятати, що ШІ можна використовувати по-різному. З одного боку, він приносить користь. Він допомагає працювати швидше. Він може обробляти великі дані, підказувати рішення і створювати нові сервіси. З іншого боку, ті самі можливості ШІ можна використати зі злим наміром. Їх застосовують для масових інформаційних атак. Вони можуть порушувати приватність. Вони можуть підсилювати дискримінацію. Їх також використовують для кібератак і диверсій. У підсумку це може збільшувати соціальну нерівність.

Аналіз досліджень та публікацій. У науковій та експертній літературі все частіше наголошується, зокрема в «Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)» Європейської комісії [4], що ризики впровадження ШІ не можна звести лише до помилок алгоритмів або помилкових рішень окремих систем. Йдеться про структурні зміни, які відбуваються в економіці, політиці, медіасередовищі та культурі під впливом інтелектуальних технологій. Саме тому аналіз ризиків ШІ має включати не тільки технічні аспекти (точність моделей, надійність інфраструктури, захист даних), а і правові, етичні, соціальні та геополітичні виміри.

У цій логіці доречно говорити про ризик-орієнтований підхід до розвитку та регулювання ШІ. Його сутність полягає в тому, що різні застосування ШІ мають різний потенціал шкоди, а отже вимагають різних за суворістю режимів контролю. Низькоризикові застосування можна регулювати м'якшими інструментами, зокрема професійними стандартами або кодексами етики. Високоризикові системи потребують жорсткіших правил. Це, наприклад, ШІ в медицині, судах, кредитуванні та керуванні критичною інфраструктурою. Тут важливі якісні дані, прозорість, тестування, аудит і контроль з боку людини.

Далі в тексті розглядаються ризики, про які найчастіше пишуть у дослідженнях і міжнародних документах [1–5; 9; 11; 13]. Перший ризик це дезінформація та маніпуляції громадською думкою. Це загрози приватності й захисту персональних даних. Це алгоритмічна упередженість і дискримінація. Це ризики для критичної інфраструктури та сфери безпеки. Це соціально-економічні наслідки змін на ринку праці. Це також правові та інституційні проблеми. Для кожної групи описано, у чому загроза, як вона може проявлятися і що можна зробити, щоб зменшити ризики.

Для опису практичних кроків управління ризиками ШІ доцільно спиратися на рамку NIST AI Risk Management Framework (AI RMF 1.0), яка структурує роботу з ризиками через етапи governance, mapping, measuring та managing [6].

Мета роботи. Метою статті є дослідження та систематизація ризиків розвитку й інтеграції штучного інтелекту в соціально-економічні та безпекові процеси, а також визначення концептуальних підходів до управління зазначеними ризиками.

Основна частина

1. Дезінформація та інформаційні маніпуляції.

Першим з найбільш помітних викликів є ризики дезінформації та маніпуляцій громадською думкою за допомогою генеративних моделей ШІ. Сучасні системи здатні створювати текст, зображення, аудіо- та відеоконтент, практично не відмінний від реального. Через це можна масово поширювати фейкові новини, deepfake-відео та підроблені документи. Також можна запускати скоординовані кампанії впливу в соцмережах.

У результаті люди можуть менше довіряти медіа, державним органам і виборам. Стратегія протидії у цьому вимірі передбачає розвиток технологій автоматичного виявлення підробленого контенту, маркування штучно згенерованих матеріалів, посилення відповідальності за свідоме поширення дезінформації, а також довгострокові програми з медіаграмотності для різних вікових груп населення, на чому, зокрема, наголошується в OECD Recommendation of the Council on Artificial Intelligence [2].

Генеративні моделі можуть за секунди створити текст або зображення. Через це робити переконливі підробки стало набагато дешевше. Раніше масова кампанія дезінформації вимагала багато грошей і людей. Тепер достатньо налаштувати кілька скриптів, щоб штампувати тисячі постів, коментарів і фейкових акаунтів. У результаті в інформаційному просторі йде постійна боротьба між правдивими повідомленнями й фейками.

Ще небезпечніше, коли генерацію поєднують із мікротаргетингом. Тоді повідомлення підбирають дуже точно під окремі групи або навіть під конкретну людину. ШІ аналізує профіль поведінки, історію пошуку, контакти та реакції. Після цього він "вгадує", на що людина найсильніше реагує. У руках політконсультантів або іноземних спецслужб це дає змогу тихо й точково впливати на громадську думку.

Не менш небезпечними є deepfake-відео, які дозволяють створити візуально достовірні ролики з політиками, військовими, лідерами громадської думки. У кризових ситуаціях - наприклад, під час терактів, військових загострень, техногенних катастроф - навіть короткочасна поява такого відео може призвести до паніки, хаотичних дій населення або нелегітимного тиску на органи влади. Чим вищий рівень напруги в суспільстві, тим легше дезінформації «вкорінитися» в масовій свідомості.

Стратегії протидії дезінформації на базі ШІ можна умовно поділити на технологічні, правові та освітньо-культурні. Технічні заходи включають програми, які знаходять підроблений контент. Вони також включають цифрові мітки для оригінальних матеріалів. Ще один напрям це інструменти, які відстежують, як поширюється підозріла інформація. Але треба розуміти, що повного захисту немає. Зловмисники постійно змінюють свої методи і підлаштовуються під новий захист.

Правовий вимір включає посилення відповідальності за свідоме поширення дезінформації, встановлення вимог до платформ щодо прозорості алгоритмів ранжування контенту й модерації, а також визначення мінімальних стандартів маркування політичної реклами та контенту, що створений або модифікований ШІ. Важливий і міжнародний рівень. Інформаційні атаки часто виходять за межі однієї країни. Тому потрібні спільні дії та координація між державами.

Також важливі освіта і культура. Потрібно підвищувати медіаграмотність і критичне мислення. Навіть найкращі системи пошуку фейків мало допоможуть, якщо люди вірять у все, що бачать у стрічці. Тому потрібні навчальні програми в школах і університетах. Корисні також курси поза навчанням і проєкти громадських організацій. Це допомагає суспільству довше триматися проти дезінформації.

У підсумку ШІ сам по собі не створює фейки. Але він робить їх масовішими, швидшими і більш переконливими.

Тому боротьба з фейками в епоху ШІ – це не тимчасова кампанія, а постійний елемент інформаційної безпеки та політики демократичної держави.

Перед тим як перейти до ілюстративного узагальнення, доцільно зафіксувати ключові механізми інформаційного впливу, які посилюються генеративними моделями (масовість, швидкість і переконливість підробок, мікротаргетинг, deepfake-контент). На рис. 1 показано основні канали, через які ШІ впливає на інформаційний простір та довіру до джерел, і як це переходить у суспільні наслідки.

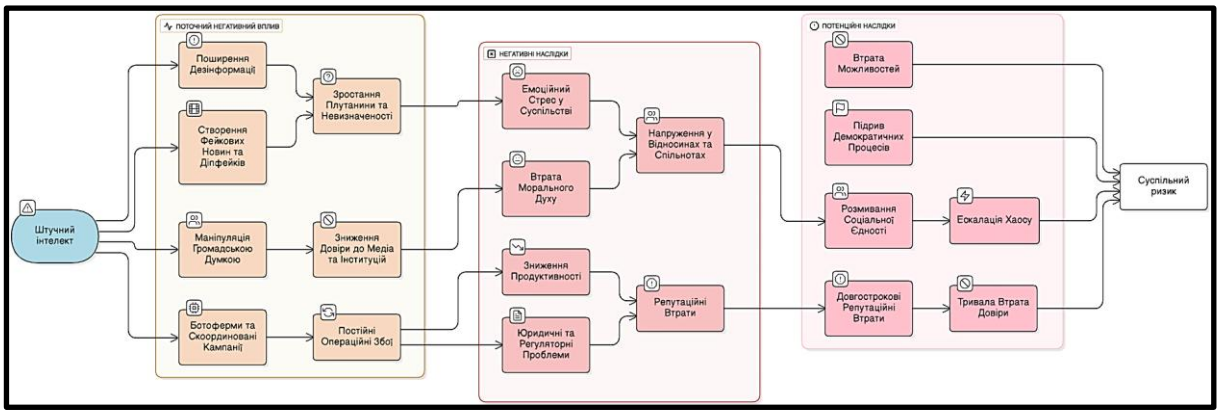

Рис. 1. Схема Впливу Штучного Інтелекту

2. Приватність і персональні дані.

Другий важливий блок ризиків стосується захисту персональних даних і права на приватність. Щоб навчати якісні моделі ШІ, потрібні великі набори даних. Часто там є чутлива інформація про користувачів. Якщо дані погано знеособлені і до них немає контролю доступу, їх можуть вкрасти або використати не так, як треба. Такі набори даних стають ціллю для кібератак і зловживань.

Є ще одна небезпека. Це цифровий нагляд. Алгоритми можуть стежити за людьми в громадських місцях. Вони також можуть стежити за тим, що люди роблять у приватному житті.

Зниження цих ризиків вимагає впровадження принципів “privacy by design”, мінімізації збору даних, проведення оцінки впливу на захист даних для проєктів із використанням ШІ, а також гармонізації національного законодавства з міжнародними стандартами захисту приватності, зокрема з UNESCO Recommendation on the Ethics of Artificial Intelligence та Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [3, 4].

У сучасній цифровій економіці дані стали ключовим ресурсом, на основі якого працює більшість бізнес-моделей. Дані про поведінку людей допомагають ШІ працювати. Це вподобання, переміщення і контакти. На основі цих даних ШІ робить прогнози, підлаштовує сервіси під людину і точніше показує рекламу. Але чим більше даних збирають, тим більший ризик для приватності.

Особливо небезпечно, коли дані з різних місць об’єднують. Окремо вони виглядають дрібницями, як-от історія покупок, геолокація, пошукові запити або лайки. Але разом вони стають одним профілем людини. За таким профілем можна здогадатися про здоров’я, гроші, політичні погляди й особисте життя. ШІ може знаходити в цих даних зв’язки, які людина не помітить. Потім це можуть використати так, що для користувача це буде несподівано або неприємно.

Відсутність прозорості щодо того, хто саме збирає дані, з якою метою, як довго їх зберігає, кому передає, створює підґрунтя для численних зловживань. Дані можуть продаватися третім сторонам, використовуватися для дискримінаційного ціноутворення (наприклад, вищі тарифи для певних груп), агресивного маркетингу, політичного таргетингу, а в авторитарних контекстах - і для переслідування опозиції, придушення інакодумства, масового спостереження.

Особливу увагу привертають системи масового відеоспостереження. Їх доповнюють алгоритми, які розпізнають обличчя, емоції та типову поведінку. Офіційно такі системи ставлять заради безпеки, боротьби зі злочинністю або для кращої роботи міста.

У цьому контексті орієнтиром виступає GDPR, який визначає правові підстави обробки даних, права суб'єктів даних та вимоги до безпеки обробки [11].

Але без чітких законів і правил це може обернутися тотальним наглядом. Якщо немає незалежного контролю та реальної відповідальності, такі дані легко використовувати не за призначенням. Тоді дії людини постійно відстежують, записують і аналізують.

Концепція privacy by design означає, що захист приватності потрібно закладати в систему з самого старту, а не додавати "потім", уже після запуску.Це означає мінімізацію збору даних, відокремлення ідентифікаторів, шифрування, локалізацію обробки (наприклад, на пристрої користувача, а не в хмарі), а також обмеження термінів зберігання. Оцінка впливу на захист даних (Data Protection Impact Assessment) дає змогу ще під час проєктування виявити можливі ризики, оцінити їхню ймовірність і масштаб наслідків та визначити заходи для їх зменшення.

У ЄС вимоги до прозорості алгоритмічного ранжування, модерації та управління системними ризиками онлайн-платформ закріплюються Digital Services Act (DSA), що важливо для протидії масштабній дезінформації [12].

Також важливо, щоб національні закони узгоджувалися з міжнародними стандартами захисту даних і приватності. Про це говорять у Рекомендації ЮНЕСКО з етики штучного інтелекту та в проєкті Регламенту ЄС про ШІ (Artificial Intelligence Act) [3,4]. Орієнтиром тут є європейський підхід до захисту персональних даних. Він визначає права людей на свої дані, обов'язки тих, хто ці дані збирає і обробляє, та контроль з боку наглядових органів. Для України та інших країн, які входять у європейський правовий простір, така гармонізація потрібна для довіри до державних цифрових сервісів і для безпечного обміну даними з іншими країнами.

Загалом, право на приватність у добу ШІ перестає бути лише індивідуальною проблемою. Це стає питанням суспільного блага. Масові порушення приватності можуть змінити баланс сили між людьми, державою і компаніями. Це створює умови для маніпуляцій. Люди можуть почати боятися і більше мовчати. Зростає недовіра.

Тому захист персональних даних у темі ШІ це не лише вимога закону. Це також важлива частина демократичної культури.

Щоб структуровано відобразити, у яких точках життєвого циклу даних виникають загрози приватності під час застосування ШІ, нижче подано схематизацію типових сценаріїв порушення приватності та витоків персональних даних. На рис. 2 узагальнено, як поєднання збору даних, профілювання та цифрового нагляду створює ризики зловживань і потребує принципів privacy by design та оцінки впливу на захист даних.

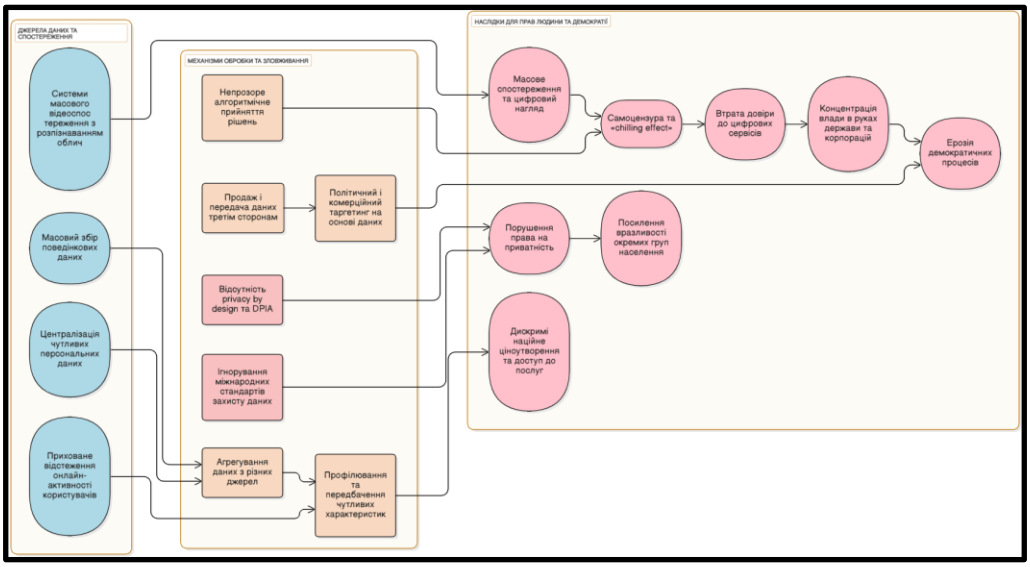

Рис. 2. Схема «Порушення приватності з ШІ»

3. Алгоритмічна упередженість і дискримінація.

Третя група ризиків пов'язана з упередженістю алгоритмів і дискримінацією.

Як зазначається у «Artificial Intelligence: A Modern Approach» та «Ethics Guidelines for Trustworthy AI» [1,5], моделі машинного навчання відображають статистику даних, на яких вони тренуються. Якщо в навчальних даних уже є упередження (за статтю, віком, етнічним походженням, мовою тощо), ШІ може їх повторювати й посилювати. Це особливо небезпечно в таких чутливих сферах, як кредитний скоринг, відбір персоналу, надання соціальної допомоги, правоохоронна діяльність. Щоб запобігти дискримінації, потрібні етичний і технічний аудит моделей, використання метрик справедливості, прозорі критерії ухвалення рішень та збереження принципу, що у високоризикових випадках фінальне рішення має залишатися за людиною.

Алгоритмічна упередженість може виникати на різних етапах життєвого циклу системи ШІ. На етапі збору даних проблема полягає в тому, що історичні дані відображають уже наявні нерівності в суспільстві. Наприклад, якщо в минулому певні групи населення рідше отримували кредити або були недопредставлені в керівних посадах, модель, навчена на цих даних, «сприйматиме» це як норму. У результаті ШІ не лише повторює старі несправедливості, а й може закріплювати їх на майбутнє. Виходить своєрідне "цифрове продовження" дискримінації.

Під час створення моделі упередженість може з'явитися через те, як задають мету системи, як міряють її якість і як побудований сам алгоритм. Наприклад, якщо система налаштована тільки на високу загальну точність, вона може гірше працювати для менших груп. Ці групи рідше є в даних. У підсумку система частіше помиляється саме щодо них.

Така ситуація особливо небезпечна, коли йдеться про захищені категорії, щодо яких діють спеціальні правові гарантії (наприклад, заборона дискримінації за ознакою раси, статі, віку).

Під час впровадження і використання важливо, як люди читають результати і в якому контексті приймають рішення. Навіть "нейтральний" алгоритм може привести до дискримінації. Так буває, коли його висновок сприймають як остаточну правду. Рішення не перевіряють і не враховують обставини. Це особливо небезпечно в поліції, судах і соцслужбах, де рішення сильно впливають на життя людини.

Щоб зменшити упередженість, потрібні кілька кроків. Спочатку треба перевірити дані і зробити їх кращими. Треба прибрати явні перекоси. Треба додати більше даних про менші групи. Також треба стежити, щоб чутливі ознаки не впливали на рішення прямо або через непрямі ознаки. Далі треба міряти справедливість результатів. Це показує, чи модель не працює гірше для одних груп, ніж для інших. І ще потрібні регулярні перевірки моделі. Це має бути технічна перевірка і етична перевірка. Це особливо важливо там, де ризик високий.

У таких випадках фінальне рішення має приймати людина, а не алгоритм. Але це працює тільки тоді, коли людина має час, знання і достатньо інформації. Вона має реально перевіряти поради ШІ. Якщо людина просто натискає "погодитися", це не є контролем і не зменшує ризики.

Алгоритмічна упередженість часто відображає проблеми в суспільстві. Це не завжди проста помилка в програмі. Тому її не виправити лише технічними правками. Потрібні чіткі правила і принципи. Також потрібна розмова про права людини в цифрову епоху.

Алгоритмічна упередженість не є одиничною «помилкою моделі», а часто формується як цикл, у якому історичні перекоси в даних, вибір метрик та практики впровадження взаємно підсилюються; окремим фактором ескалації є отруєння даних і навмисні

маніпуляції. На рисунку 3 показано цей цикл і логіку, чому зниження дискримінаційних ризиків потребує регулярного аудиту даних і моделей та наявності людського контролю у високоризикових застосуваннях.

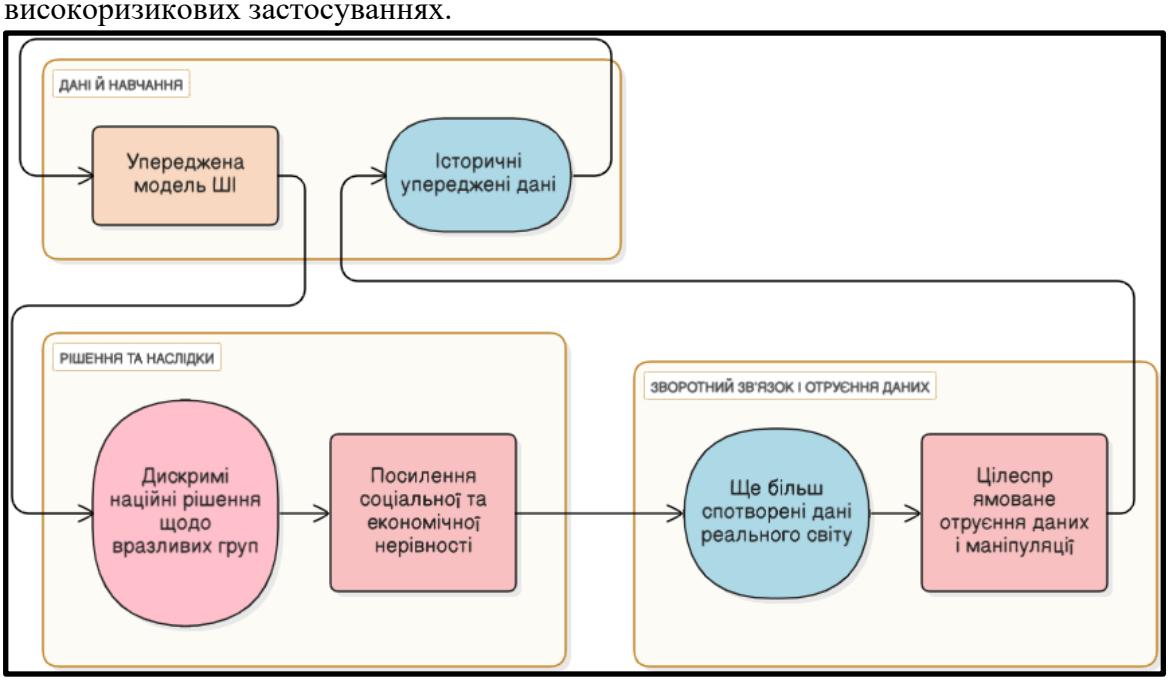

Рис. 3. Схема «Цикл алгоритмічної упередженості та отруєння даних»

4. Критична інфраструктура та безпекова сфера.

Окремий блок ризиків стосується ШІ в критичній інфраструктурі та сфері безпеки. Як зазначено в «OECD Recommendation of the Council on Artificial Intelligence» та «UNESCO Recommendation on the Ethics of Artificial Intelligence» [2,3], такі системи в енергетиці, транспорті, медицині й промисловості можуть підвищувати ефективність. Але водночас вони створюють більше можливостей для кібератак. Злам або підміна роботи таких систем може спричинити масштабні збої й аварії та створити загрози для життя і здоров'я людей. Ще складнішим є питання застосування автономних та напівавтономних систем озброєнь, здатних приймати рішення про застосування сили з мінімальною участю людини. У цьому контексті міжнародні організації наголошують на потребі заборонити повністю автономну летальну зброю, запровадити контроль за експортом таких технологій і виробити міжнародно-правові правила, які визначатимуть відповідальність за наслідки їх застосування.

Критична інфраструктура – це сукупність об'єктів і систем, від функціонування яких залежить життєздатність суспільства й держави. Йдеться про енергетику, транспорт, водопостачання, зв'язок, охорону здоров'я, банківську систему, державне управління. Інтеграція ШІ в такі системи дає низку переваг: можливість прогнозувати аварії, оптимізувати розподіл ресурсів, швидко реагувати на нестандартні ситуації. Водночас будь-яка помилка або злочинне втручання в алгоритм можуть мати катастрофічні наслідки.

Окрему небезпеку становлять кібератаки на ШІ-системи, які керують фізичними процесами. Це може бути "отруєння даних". Зловмисник тихо підміняє навчальні або робочі дані. Через це система поступово починає приймати неправильні рішення. Є й інший тип атак, adversarial-атаки. Тоді роблять спеціальні вхідні сигнали. Для моделі вони виглядають нормальними. Проте модель через них помиляється під час розпізнавання або класифікації.

У транспорті, енергомережах і медичному обладнанні такі атаки можуть викликати серйозні аварії. Для цього не обов'язково мати фізичний доступ до об'єкта.

Описані атаки на ML (зокрема data poisoning та adversarial examples) систематизовані в таксономії NIST, де наведено терміни, типи атак і типові заходи протидії [7].

Окреме питання це ШІ у військовій сфері. Автономні та напівавтономні системи можуть самі знаходити цілі. Вони можуть самі вирішувати, коли застосувати силу. Вони можуть завдавати ударів без прямої команди людини. Прихильники кажуть, що такі системи можуть працювати точніше. Вони також можуть зменшувати ризик для своїх військових. Ще вони можуть діяти швидше в критичні моменти. Але ризик помилок і збоїв дуже високий. Є ризик хакерських атак. Є ризик непередбачуваної поведінки в нових умовах.

Міжнародні організації, правозахисники та експерти часто наголошують на іншому. Вони вважають, що не можна віддавати алгоритмам рішення про життя і смерть. На їхню думку, це суперечить базовим принципам міжнародного гуманітарного права та етики. Тому у світі дедалі частіше говорять про таке. Повністю автономні летальні системи треба заборонити. А якщо ШІ використовують у військовій сфері, має залишатися реальний контроль з боку людини.

Для країн, які стикаються з гібридною агресією, зокрема для України, ШІ в обороні є потрібним, але складним. З одного боку, він допомагає аналізувати розвіддані, виявляти кібератаки і прогнозувати розвиток подій. Це підсилює оборону. З іншого боку, противник може намагатися зламати ці системи або обдурити їх. Для цього він теж може використати ШІ.

ENISA окремо наголошує, що інтеграція ШІ в кіберзахист і критичні системи створює нові поверхні атак (модель, дані, пайплайн, постачальники), що потребує спеціальних вимог до безпеки й тестування [8].

Таким чином, ШІ в критичній інфраструктурі та безпековій сфері – це сфера, де потенціал корисної дії й потенційна шкода є максимальними. Тому саме тут вимоги до надійності, кіберзахисту, прозорості та людського контролю мають бути максимально жорсткими.

5. Соціально-економічні ризики та трансформація ринку праці.

Не менш важливі ризики для економіки та роботи людей. Через автоматизацію рутинних задач і "розумні" системи частина професій може скоротитися або зникнути.

Також може зрости розрив між висококваліфікованими працівниками і тими, хто має низьку зарплату. Якщо держава і бізнес не діятимуть заздалегідь, це може призвести до більшого безробіття, нерівності та невдоволення. Щоб зменшити ці ризики, потрібні програми перекваліфікації та навчання протягом життя. Також важливо розвивати професії, де людина працює разом із ШІ. Ще один крок це стимули для роботодавців, які вкладають гроші в навчання працівників і розвиток їхніх навичок.

У публічних дискусіях щодо ШІ часто звучить питання: "Скільки робочих місць забере ШІ?" Правильніше говорити не про просте "знищення" робочих місць, а про глибоку зміну структури зайнятості. Частина професій справді може зникнути або суттєво скоротитися, але водночас з'являтимуться нові ролі, пов'язані з розробкою, налаштуванням, супроводом, аудитом і етичним контролем ШІ-систем. Проблема полягає в тому, що цей перехід не є автоматичним і безболісним для людей, які сьогодні працюють у професіях, що автоматизуються.

Оцінки впливу генеративного ШІ на кількість і якість робочих місць, а також ризики трансформації зайнятості та умов праці, детально розглянуті в аналітичному звіті Міжнародної організації праці [14].

Найбільш уразливі працівники з низькими цифровими навичками. Також уразливі ті,

чия робота складається з повторюваних і стандартних дій. Це можуть бути оператори контакт-центрів, касири, частина офісних працівників і молодший адміністративний персонал. У багатьох сферах уже зараз з'являються чат-боти, автоматична обробка документів і алгоритми, які попередньо перевіряють заявки. Через це потрібно менше ручної роботи.

Водночас зростає попит на фахівців, які вміють поєднувати технології з конкретною сферою. Це аналітики даних, інженери з машинного навчання, спеціалісти з кібербезпеки, архітектори цифрових платформ, фахівці з етики ШІ та консультанти з цифрової трансформації. Але підготувати таких фахівців складно. На це потрібні час, гроші і якісна освіта. І це доступно не для всіх.

Крім того, ШІ змінює саму організацію праці. Все більше людей працюють через платформи. Це фриланс, віддалена робота і гіг-економіка. Там алгоритми фактично керують роботою. Вони роздають завдання, оцінюють результат і можуть впливати на оплату. Якщо немає нормальних правил і захисту, така робота стає нестабільною. Люди можуть залишатися без соціальних гарантій. Також зростає нерівність між працівником і платформою.

Соціально-економічні ризики через ШІ не означають кінець світу. Але потрібні дії наперед. Потрібні інвестиції в освіту і перекваліфікацію. Потрібні державні програми для розвитку навичок майбутнього. Також важливо зробити справедливий перехід. Люди, які втрачають роботу через автоматизацію, мають отримати реальний шанс перейти в іншу сферу.

Для України є ще один фактор. Економіка багато в чому тримається на ІТ і аутсорсингу. ШІ тут може допомагати ставати сильнішими. Але він також посилює конкуренцію з іншими країнами. Якщо інші країни швидше впровадять ШІ у виробництво і бізнес, частина звичних аутсорсингових послуг може стати менш потрібною. Тоді українським компаніям доведеться переходити на складніші продукти та сервіси.

Це своєю чергою вимагатиме підвищення кваліфікації значної кількості фахівців і створення сприятливих умов для розвитку інноваційних стартапів.

У підсумку ШІ змінює не лише окремі професії. Він також змінює правила між працівником, роботодавцем і державою. Важливо, щоб освіта, соціальний захист і трудові закони оновили вчасно і чесно. Тоді ці зміни можуть дати зростання і нові можливості. Якщо ні, вони можуть привести до нестабільності та конфліктів.

6. Нормативно-правове та інституційне середовище.

Важливою умовою безпечного розвитку ШІ є формування адекватного нормативно-правового та інституційного середовища. На рівні Європейського Союзу вже пропонується регулювання за ризик-орієнтованим підходом, що передбачає жорсткіші вимоги до систем, здатних істотно впливати на права та свободи людини [4]. Для України важливо врахувати ці підходи, коли вона робить свою стратегію розвитку ШІ. Потрібно узгодити стандарти з європейськими правилами. Потрібно створити незалежні органи нагляду. Також варто залучити науковців і громадські організації до обговорення правил використання таких технологій.

Ризик-орієнтований підхід, запропонований у проєкті Регламенту ЄС щодо ШІ (Artificial Intelligence Act) [4], передбачає класифікацію систем ШІ за рівнем ризику. До категорії неприпустимого ризику відносяться практики, які несумісні з правами людини (наприклад, певні форми маніпулятивного впливу на вразливі групи, соціальний скоринг громадян). Такі системи пропонується повністю заборонити.

Категорія високого ризику охоплює системи, які використовують у критично важливих сферах: інфраструктурі, охороні здоров'я, освіті, працевлаштуванні, кредитуванні та правоохоронній діяльності. Для таких систем діють суворі правила. Потрібно правильно керувати даними. Потрібна документація і прозорість. Потрібен контроль з боку людини й сильний кіберзахист. Також потрібна обов'язкова перевірка відповідності. Суть підходу проста. ШІ загалом не треба "душити" правилами. Найбільше уваги треба там, де наслідки можуть бути найважчими.

Системи з обмеженим ризиком здебільшого мають вимоги до прозорості. Наприклад, користувача треба попередити, що він спілкується з ботом. А системи з мінімальним ризиком, як-от ігри чи спам-фільтри, зазвичай не потребують окремого втручання. Це дозволяє одночасно зберігати безпеку і розвивати технології. Також це не створює зайвих вимог для простих продуктів.

Окрім законів є "м'які" правила. Це етичні кодекси, стандарти галузі та поради для розробників і користувачів. Документи ОЕСР, ЮНЕСКО та Європейської комісії [2–5] задають основні принципи. Серед них фокус на людині, повага до прав людини, недискримінація, прозорість, відповідальність і технічна надійність.

Для України узгодження політики щодо ШІ з європейськими підходами є і викликом, і можливістю. Це означає потребу в оновленні законодавства в частині захисту персональних даних, кібербезпеки, електронних комунікацій, стандартизації, а також у створенні спеціалізованих інституцій, відповідальних за нагляд у сфері ШІ. Такими інституціями можуть бути окремі агентства, міжвідомчі ради, етичні комітети, дорадчі органи за участі науковців і представників громадянського суспільства.

Важливо розвивати регуляторні "пісочниці". У них нові ШІ-проєкти тестують в обмежених і контрольованих умовах. За процесом стежить регулятор.

Це дає змогу тестувати нові підходи й не підставляти під ризик багато людей. Також це допомагає державі побачити реальні проблеми та підлаштувати правила.

Але самі закони безпеку не гарантують. Потрібні сильні інституції, які можуть ці закони виконувати. Потрібні фахівці, гроші й ресурси. Потрібні перевірки та покарання за порушення. Також потрібна звичка дотримуватися правил і в державі, і в бізнесі. Тільки за наявності таких умов регулювання ШІ перестає бути декларативним і перетворюється на реальний інструмент управління ризиками.

Результат дослідження. Існуючі підходи до класифікації ризиків використання ШІ часто або (а) описують ризики за доменами застосування (медицина, фінанси тощо), або (б) задають рівні регуляторної суворості залежно від потенційної шкоди. Запропонована в цій статті систематизація доповнює такі підходи тим, що групує ризики за механізмами прояву та окремо виділяє "комбіновані" сценарії, коли ризики різних вимірів зчіплюються в один ланцюжок ескалації. Це дає змогу не тільки перелічити загрози, а й пояснити, чому точкові заходи (наприклад, лише кіберзахист або лише медіаграмотність) можуть бути недостатні.

Оскільки описані групи ризиків у реальних сценаріях рідко проявляються ізольовано, для узагальнення їх взаємного підсилення запропоновано концептуальну схему комбінованих ризиків. На рис. 4 подано діаграму Венна перетину чотирьох вимірів (S1 – S4), яка демонструє типові «ланцюжки ескалації» та пояснює, чому ефективна протидія має бути комплексною, а не зосередженою лише на одному напрямі.

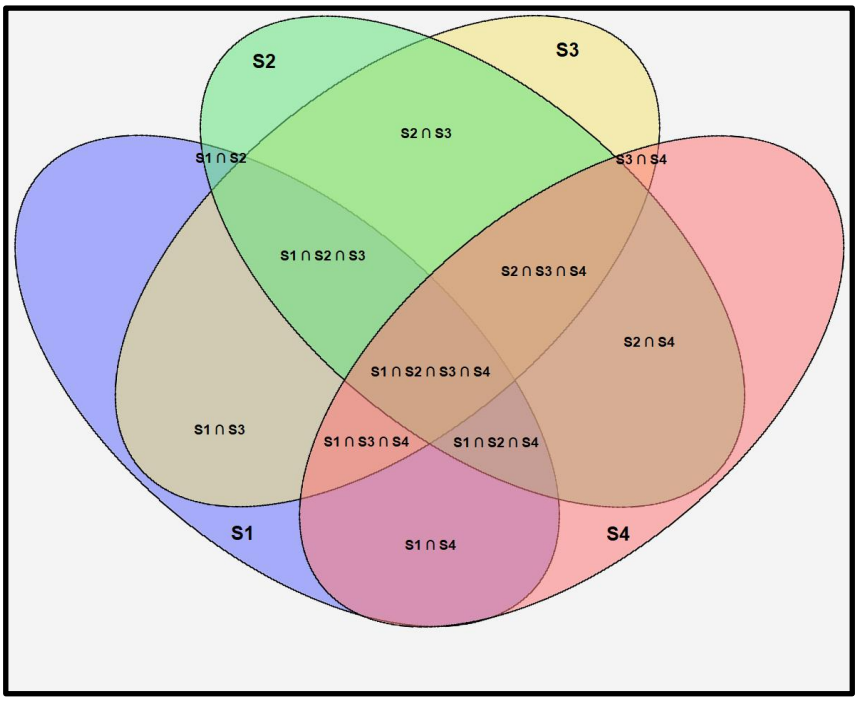

Рис. 4. Групи ризиків впровадження ШІ

(S1) = Інформаційно-комунікативні ризики → «дезінформація, deepfake, підрив довіри»;

(S2) = Приватність, дані, цифровий нагляд → «тотальний збір персональних даних, приховане профілювання, нав'язливий моніторинг і контроль»;

(S3) = Соціально-економічні/дискримінаційні ризики → «алгоритмічна дискримінація, поляризація, нерівність»;

(S4) = Інфраструктурні/безпекові ризики → «критична інфраструктура, аварії/збої, атаки»;

(S1 ∩ S2) → «таргетована дезінформація на основі профілів» / «маніпуляція через персональні дані»;

(S1 ∩ S3) → «радикалізація груп через рекомендаційні системи» / «нормалізація дискримінаційних наративів»;

(S1 ∩ S4) → «інфо-кібератаки + паніка/хаос» / «фейки під час інцидентів інфраструктури»;

(S2 ∩ S3) → «цифрова сегрегація і "соціальні рейтинги"» / «втрата приватності → посилення нерівності»;

(S2 ∩ S4) → «соціальний контроль через дані та критичні системи» / «прицільні атаки й шантаж операторів»;

(S3 ∩ S4) → «нерівний доступ до безпечних технологій і сервісів» / «вразливість груп і регіонів»;

(S1 ∩ S2 ∩ S3) → «маніпуляції + профілювання + дискримінація» → керований вплив, що закріплює нерівність

(S1 ∩ S2 ∩ S4) → «профілі + вплив + важелі інфраструктури» → тиск/шантаж + інформаційне прикриття;

(S1 ∩ S3 ∩ S4) → «фейки + соціальний конфлікт + збої» → поляризація на тлі аварій/атак;

(S2 ∩ S3 ∩ S4) → «дані + сегрегація + інфраструктурний примус» → керована дестабілізація й цифровий контроль;

(S1 ∩ S2 ∩ S3 ∩ S4) → «цифровий авторитаризм: тотальний контроль, руйнація довіри, згортання демократії».

Запропонована діаграма Венна демонструє, що негативні наслідки розвитку генеративного ШІ та зловживання самонавчальними алгоритмами не існують окремо в «ізольованих» сферах, а взаємно підсилюються у чотирьох вимірах: (S1) інформаційно-комунікативному (дезінформація, deepfake, підрив довіри), (S2) приватності й цифрового нагляду (масовий збір даних, профілювання), (S3) соціально-економічному/дискримінаційному (нерівність, поляризація, алгоритмічна дискримінація) та (S4) інфраструктурно-безпековому (атаки, збої критичних систем). Перетини кіл підкреслюють «комбіновані» ризики: наприклад, профілювання робить дезінформацію таргетованою й ефективнішою; рекомендаційні механізми можуть підживлювати радикалізацію та конфлікт; дані й контроль над критичними системами створюють інструменти тиску, шантажу й керованої дестабілізації.

Центральний перетин (S1 ∩ S2 ∩ S3 ∩ S4) узагальнює системний сценарій: «цифровий авторитаризм» - коли одночасно руйнується суспільна довіра, посилюється нерівність, зростає тотальний нагляд і з'являються важелі примусу через інфраструктуру, що веде до згортання демократичних процесів. Практичний сенс схеми: протидія має бути комплексною (медіастійкість + приватність/захист даних + недискримінаційні механізми + кіберстійкість критичної інфраструктури), інакше «латання» лише однієї сфери залишає інші канали для ескалації ризиків.

Висновок. Отже, ризики використання ШІ різні і не зводяться лише до помилок у алгоритмах. Це цілий набір проблем. Сюди входять фейки, втрата приватності, дискримінація, кібератаки і соціальна напруга. Щоб цьому протидіяти, потрібні різні заходи разом. Потрібен технічний захист. Потрібні етичні правила. Потрібні чіткі закони. Потрібні сильні державні органи. Потрібна цифрова грамотність у суспільстві. Лише тоді ШІ може допомагати розвитку, а не створювати нові великі кризи.

Проведена систематизація груп ризиків у вигляді діаграми Венна показує, що ключові негативні наслідки виникають саме у "перетинах": профілювання робить дезінформацію таргетованою; рекомендаційні механізми можуть підсилювати поляризацію й дискримінаційні наративи; контроль над даними та критичною інфраструктурою створює інструменти тиску, шантажу й керованої дестабілізації. Центральний перетин узагальнює системний сценарій "цифрового авторитаризму", коли одночасно зростає нагляд, руйнується довіра, посилюється нерівність і з'являються важелі примусу через інфраструктуру.

Розглянуті в статті групи ризиків - інформаційні, приватнісні, дискримінаційні, інфраструктурні, соціально-економічні та нормативно-правові - тісно пов'язані між собою. Посилення контролю над даними без достатніх гарантій прав людини може породити загрози приватності та зловживання владою. Якщо не звертати уваги на упередженість алгоритмів, люди можуть перестати довіряти цифровим сервісам. Це може статися навіть тоді, коли з кіберзахистом усе добре. Якщо автоматизувати роботу занадто швидко і не робити програм перекваліфікації, зросте соціальна напруга. Це також може підштовхнути частину людей до радикальних настроїв.

Тому стратегія безпечного розвитку й використання ШІ повинна бути міждисциплінарною та міжсекторальною. Вона має об'єднувати інженерів, правників, філософів, економістів, соціологів, освітян, представників громадянського суспільства та бізнесу. Тільки так можна нормально розібратися в проблемі з усіх боків і зробити рішення, які не будуть просто "для вигляду".

Для України це особливо важливо. Країна розвиває ІТ, входить у європейський правовий простір і водночас протистоїть гібридній агресії. Тому управління ризиками ШІ має велике значення. ШІ може стати потужним інструментом відновлення, модернізації та інтеграції в глобальну економіку, але за умови, що його розвиток відбуватиметься в руслі прав людини, демократії та верховенства права.

У цьому контексті міжнародні документи [1–15] можна сприймати як стратегічний орієнтир, однак конкретний шлях кожна країна має визначати самостійно, з урахуванням власних історичних, культурних, економічних і безпекових умов. Якщо цей шлях буде побудований на прозорості, підзвітності, участі громадян і повазі до гідності кожної людини, штучний інтелект стане не загрозою, а партнером у досягненні цілей сталого розвитку.

Нові покоління програмно-апаратного забезпечення повинні бути оснащені сильнішими та зручнішими вбудованими засобами захисту. Слід підвищувати рівень безпеки комп'ютерних мереж, які використовуються для роботи з секретними даними.

Приділяти увагу треба розвитку безпечної і повсюдної електронної ідентифікації (eID), що полегшить транскордонне використання онлайн-послуг та створить умови для інтеграції України у світовий електронний інформаційний простір. Оскільки найгострішим питанням надання електронних послуг провайдерами різних сфер на сьогодні є питання захисту персональних даних споживачів таких послуг, слід посилити контроль за дотриманням вимог законодавства щодо унеможливлення доступу зловмисників до конфіденційних даних споживачів та забезпечення анонімності при eID за рахунок впровадження новітніх технічно-програмних рішень реалізації електронних транзакцій.

З масовим розповсюдженням технології інтернету речей, переходом у хмарні сховища даних, формуванням обліку FinTech, зокрема цифрових та криптовалют, криптобірж, електронних виборів та «розумних контрактів», для зниження небезпечних вразливостей треба ретельно захищати метадані від можливого викрадення унаслідок зловмисних атак.

Найближчими роками важливо не лише "наздогнати" технології, а й навчитися їх контролювати. Для цього треба вкладати гроші в освіту і науку. Треба визначити національні правила, за якими ШІ можна вважати надійним. Треба підтримувати відкриті дослідження ризиків. Також потрібен зв'язок між користувачами, розробниками й державними регуляторами.

Список використаних джерел

1. Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.) Prentice Hall. https://api.pageplace.de/preview/DT0400.9781292153971_A27091185/preview-9781292153971_A27091185.pdf.

2. OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449) Organisation for Economic Co-operation and Development. https://oecd.ai/en/assets/files/OECD-LEGAL-0449-en.pdf.

3. UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

4. European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* (COM/2021/206 final).EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

5. High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf.

6. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

7. National Institute of Standards and Technology. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (NIST AI 100-2e2025). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf.

8. ENISA. (2020). *Artificial Intelligence Cybersecurity Challenges*. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges.

9. European Union. (2024). *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. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

10. Council of Europe. (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). Council of Europe. https://rm.coe.int/1680afae3c.

11. European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

12. European Union. (2022). *Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

13. Council of Europe. (2020). *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*. Council of Europe. https://rm.coe.int/09000016809e1154.

14. Gmyrek, P., Berg, J., & Bescond, D. (2023). *Generative AI and Jobs: A global analysis of potential effects on job quantity and quality* (ILO Working Paper No. 96). International Labour Organization. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40inst/documents/publication/wcms_890761.pdf.

15. ISO/IEC. (2023). *ISO/IEC 23894:2023 – Artificial intelligence: Guidance on risk management*. International Organization for Standardization & International Electrotechnical Commission. https://www.iso.org/standard/77304.html.

References

1. Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.) Prentice Hall. https://api.pageplace.de/preview/DT0400.9781292153971_A27091185/preview-9781292153971_A27091185.pdf.

2. OECD. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449) Organisation for Economic Co-operation and Development. https://oecd.ai/en/assets/files/OECD-LEGAL-0449-en.pdf.

3. UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

4. European Commission. (2021). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* (COM/2021/206 final).EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

5. High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf.

6. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

7. National Institute of Standards and Technology. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations* (NIST AI 100-2e2025). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf.

8. ENISA. (2020). *Artificial Intelligence Cybersecurity Challenges*. European Union Agency for Cybersecurity. https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges.

9. European Union. (2024). *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. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

10.Council of Europe. (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). Council of Europe. https://rm.coe.int/1680afae3c.

11.European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

12.European Union. (2022). *Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng.

13. Council of Europe. (2020). *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*. Council of Europe. https://rm.coe.int/09000016809e1154.

14. Gmyrek, P., Berg, J., & Bescond, D. (2023). *Generative AI and Jobs: A global analysis of potential effects on job quantity and quality* (ILO Working Paper No. 96). International Labour Organization. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40inst/documents/publication/wcms_890761.pdf.

15. ISO/IEC. (2023). *ISO/IEC 23894:2023 – Artificial intelligence: Guidance on risk management*. International Organization for Standardization & International Electrotechnical Commission. https://www.iso.org/standard/77304.html.

Дата першого надходження статті до видання: 23.12.2025
Дата прийняття статті до друку після рецензування: 15.01.2026

UDC 004.8:004.056:34:330.46

Yuliia Tkach[1], Vladyslav Odnokolov[2], Taras Petrenko[3]

[1]Doctor of Pedagogical Sciences, PhD in Technical Sciences, Professor,
Head of the Department of Cybersecurity and Mathematical Modeling
Chernihiv Polytechnic National University (Chernihiv, Ukraine)
E-mail: tkachym79@gmail.com. **ORCID:** https://orcid.org/0000-0002-8565-0525

[2]Master's Student, Department of Cybersecurity and Mathematical Modeling
Chernihiv Polytechnic National University (Chernihiv, Ukraine)
E-mail: olexodno@gmail.com. **ORCID:** https://orcid.org/0009-0004-8157-5259

[3]PhD, associate professor of the Department of Cybersecurity and Mathematical Modeling
Chernihiv Polytechnic National University (Chernihiv, Ukraine)
E-mail: mail_taras@stu.cn.ua. **ORCID:** https://orcid.org/0000-0001-5571-3815

RISKS OF IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE: SECURITY, LEGAL, AND SOCIO-ECONOMIC ASPECTS

The article examines and systematizes the risks associated with the development and integration of artificial intelligence into socio-economic and security processes in the context of digital transformation. The multidimensional impact of artificial intelligence technologies on the information environment, labor market, personal data protection, economic relations, and national security is analyzed. The main categories of risks related to the use of artificial intelligence systems are generalized and structured, including technological, legal, ethical, social, and security risks. The mechanisms through which these risks materialize and their potential consequences for public institutions, the business sector, and society as a whole are considered. Particular attention is paid to international approaches to the regulation of artificial intelligence, especially risk-based regulatory models implemented within the European Union. Conceptual directions for minimizing and regulating the risks associated with the integration of artificial intelligence are identified.

Keywords: *artificial intelligence; risks; digital transformation; information security; AI regulation.*
Fig.: 4. References: 15.