УДК 004.8:005.334 DOI: https://doi.org/10.37332/2309-1533.2025.4.32
JEL Classification: D81, M15, O33

Шухманн В.А.,
здобувач* третього (освітньо-наукового) рівня вищої освіти
«доктор філософії» за спеціальністю 051 Економіка,
ORCID: https://orcid.org/0000-0002-1427-3312,
Західноукраїнський національний університет, м. Тернопіль

НЕДОСТОВІРНІСТЬ ІНФОРМАЦІЇ ЦИФРОВОГО ПРОСТОРУ ЯК ВИКЛИК ШТУЧНОМУ ІНТЕЛЕКТУ В ПРОЦЕСІ УПРАВЛІННЯ РИЗИКАМИ

Schuchmann V.A.,
candidate for the third level of higher education
"Doctor of Philosophy" in specialty 051 Economics,
West Ukrainian National University, Ternopil

UNRELIABILITY OF INFORMATION IN THE DIGITAL SPACE AS A CHALLENGE TO ARTIFICIAL INTELLIGENCE IN THE PROCESS OF RISK MANAGEMENT

Постановка проблеми. Сучасний етап розвитку економічних систем характеризується поглибленням процесів цифровізації та активним впровадженням інструментів штучного інтелекту в практику управлінських рішень. Підприємства й органи державної влади дедалі ширше застосовують алгоритми машинного навчання для обробки великих масивів даних, прогнозування ринкової кон'юнктури та виявлення потенційних загроз. Водночас ефективність функціонування таких систем безпосередньо залежить від якості, достовірності та повноти інформаційних ресурсів, що використовуються в процесі аналізу.

В умовах глобального цифрового середовища, для якого характерні інформаційна надмірність, поширення дезінформації та маніпулятивного контенту, формується суперечність, коли технології штучного інтелекту, покликані знижувати рівень невизначеності, за наявності недостовірних даних можуть, навпаки, посилювати її. Нормативно-правове поле України, зокрема Закон України «Про Національну програму інформатизації» [1], визначає інформаційну безпеку одним із пріоритетних напрямів державної політики, проте питання системної верифікації контенту для автоматизованих систем підтримки управлінських рішень залишається недостатньо врегульованим. В результаті недостовірна інформація трансформується з комунікаційної проблеми у вагомий чинник ризику стратегічного управління, що обумовлює необхідність перегляду підходів до ризик-менеджменту в умовах поширення генеративних технологій штучного інтелекту.

Аналіз останніх досліджень і публікацій. У сучасній науковій літературі проблематика управління ризиками в умовах цифровізації розглядається з різних концептуальних позицій. Значна частина досліджень зосереджена на трансформації систем ризик-менеджменту та забезпеченні фінансової безпеки в цифровій економіці. Зокрема, у працях О. Десятнюк і О. Птащенко [2] обґрунтовується необхідність адаптації традиційних підходів до управління ризиками з урахуванням впливу цифрових технологій та зростання фінансових загроз. У цьому контексті О. Скіцько та співавтори [3] здійснили систематизацію основних загроз і ризиків використання штучного інтелекту в економічних системах, що створює теоретичне підґрунтя для подальших досліджень.

Окремий напрям наукових розвідок пов'язаний із процесами прийняття та оцінювання управлінських рішень у складному й динамічному середовищі. Так, А. Мельник і Р. Винокуров [4] аналізують особливості формування управлінських рішень в умовах зростання невизначеності, тоді як С. Онищенко і А. Глушко [5] акцентують увагу на ролі інформаційно-аналітичного забезпечення у підтримці фінансової стабільності підприємств. Важливим доповненням до цих досліджень є праця Н. Лесько [6], в якій розкрито правові механізми забезпечення достовірності інформації як ключової передумови легітимності управлінських рішень, а також робота О. Кравчука [7], присвячена методичним аспектам ідентифікації ризиків, пов'язаних із впровадженням систем штучного інтелекту.

* *Науковий керівник: Буяк Л.М. – д-р екон. наук, професор*

У межах техніко-інформаційного підходу увага науковців зосереджується на специфічних загрозах цифрового середовища. Зокрема, В. Івкова та І. Опірський [8] досліджують ризики, зумовлені використанням OSINT-інструментів.

Важливим об'єктом аналізу сучасних досліджень є когнітивні обмеження та помилки інтелектуальних систем. Феномен «галюцинацій» генеративних моделей та його вплив на достовірність даних розглядається у працях Є. Махна та колективу авторів [9]. Як один із підходів до мінімізації таких ризиків у науковій літературі обґрунтовується концепція «пояснювального штучного інтелекту» (Explainable AI), теоретичні засади якої систематизовано у працях А. Барредо Аррієти та співавторів [10].

Водночас, попри значну кількість наукових напрацювань, у літературі недостатньо дослідженим залишається питання впливу недостовірної зовнішньої інформації на ефективність управлінських рішень, сформованих із використанням систем штучного інтелекту, зокрема в контексті стратегічного управління підприємством.

Постановка завдання. Метою статті є дослідження впливу недостовірної інформації та дезінформаційних потоків цифрового середовища на ефективність використання штучного інтелекту в системі управління ризиками підприємства, а також обґрунтування багаторівневих підходів до перевірки даних задля підвищення якості стратегічних управлінських рішень.

Виклад основного матеріалу дослідження. Коректність управлінських рішень, незалежно від того, чи вони формуються людиною, чи автоматизованою системою, безпосередньо обумовлена відповідністю інформації фактичному стану досліджуваного об'єкта або процесу. У науковій літературі якість інформаційних ресурсів розглядається крізь призму сукупності ключових характеристик, серед яких виділяють рівень повноти та достатності відомостей, їх актуальність, адекватність відображення реальності, точність, стійкість до спотворень, а також своєчасність використання [6, с. 141].

У системі стратегічного управління ризиками підприємства однією з ключових загроз виступає семантична ненадійність інформації, що використовується алгоритмами штучного інтелекту. Механізм викривлення процесу формування управлінських рішень доцільно подати у вигляді узагальненої концептуальної моделі, наведеної на рис. 1.

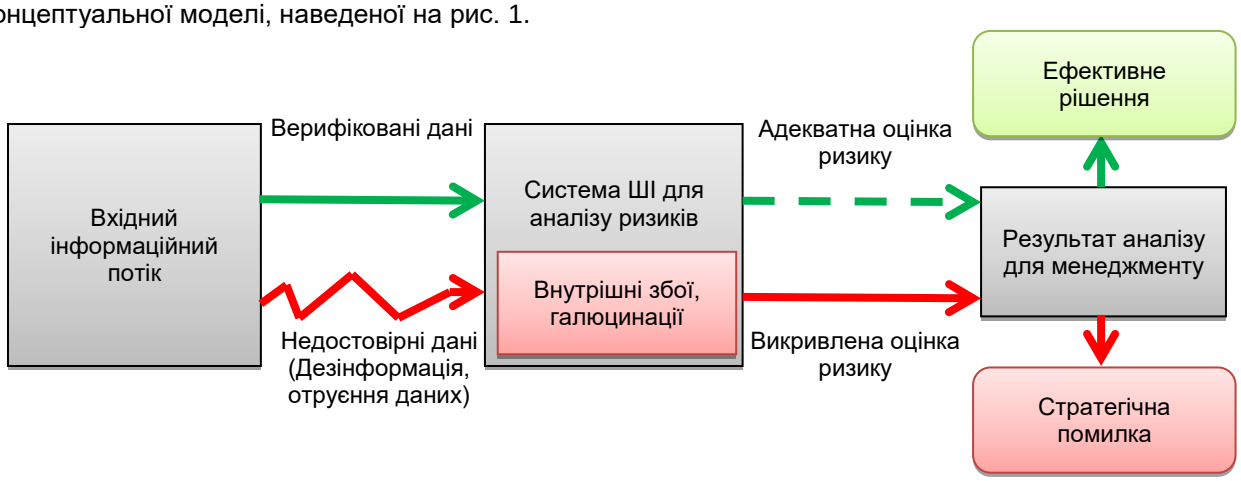

Рис. 1. Концептуальна модель впливу недостовірності інформації на прийняття рішень системою ШІ

Джерело: розроблено автором самостійно

Відповідно до представленої концептуальної моделі, зазначені загрози доцільно групувати залежно від походження, виокремлюючи зовнішні та внутрішні чинники:

1. Зовнішні загрози: викривлення вхідних інформаційних потоків та ризики застосування OSINT. До групи зовнішніх загроз належать дії, що спрямовані на спотворення інформаційного середовища, яке використовується системами штучного інтелекту для аналітичної обробки. Найбільш критичним проявом таких впливів є явище «отруєння даних», що полягає у навмисному включенні до інформаційних джерел спеціально підготовлених некоректних або маніпулятивних відомостей, на основі яких здійснюється навчання.

Особливого значення в цьому контексті набуває використання розвідувальних технологій, заснованих на аналізі відкритих джерел (OSINT). Як зазначають В. Івкова та І. Опірський, попри високу аналітичну цінність таких інструментів, відкриті дані характеризуються підвищеною вразливістю до фальсифікації, що створює передумови для дезінформації автоматизованих систем. У разі використання корпоративними системами ШІ інформації з неперевірених зовнішніх платформ для оцінювання ринкової ситуації цілеспрямований інформаційний вплив з боку конкурентів може призвести до формування помилкових оцінок ризиків [8].

2. Внутрішні загрози: когнітивні викривлення та феномен «галюцинацій» штучного інтелекту. Поряд із зовнішніми чинниками, суттєву небезпеку становлять внутрішні обмеження самих інтелектуальних моделей. Навіть за наявності відносно надійних вхідних даних генеративні системи демонструють схильність до специфічних когнітивних збоїв, відомих як «галюцинації».

У площині управління підприємством такі помилки можуть проявлятися у формуванні неіснуючих ринкових тенденцій або у генерації необґрунтованих прогнозних показників. Основна небезпека полягає у високому ступені правдоподібності подібного контенту, що істотно ускладнює його своєчасне виявлення та перевірку з боку фахівців із ризик-менеджменту.

Узагальнену систематизацію ключових інформаційних загроз для систем штучного інтелекту в процесі управління ризиками подано в табл. 1.

Таблиця 1

Загрози викривлення інформації в алгоритмах штучного інтелекту та їх управлінські наслідки

Категорія загрози	Спосіб реалізації	Наслідки для системи ризик-менеджменту на основі ШІ
Отруєння даних	Навмисне додавання або заміна частини вхідної інформації хибними чи упередженими наборами даних	Деформація логіки роботи моделі, що зумовлює систематичні помилки у виявленні та ранжуванні ризиків
Маніпуляції з OSINT-даними	Масове продукування та поширення неправдивих або маніпулятивних повідомлень у публічному інформаційному просторі, який аналізує ШІ	Викривлене сприйняття зовнішнього середовища (ринкових умов, надійності контрагентів, репутаційних факторів), що веде до неправильних управлінських дій
Галюцинації генеративних моделей	Формування системою переконливих, але фактично недостовірних тверджень унаслідок імовірнісної природи генерації	Використання вигаданих даних або нереалістичних прогнозів як підґрунтя для стратегічних рішень, що створює загрози для довгострокової стійкості підприємства

Джерело: узагальнено на основі джерел [4; 9]

Використання спотворених або ненадійних даних у ШІ-орієнтованій системі ризик-менеджменту запускає ланцюговий ефект негативних наслідків, що проявляється на всіх управлінських рівнях підприємства – від поточної операційної діяльності до формування довгострокової стратегії.

Надходження недостовірної інформації в управлінські процеси підприємства запускає послідовний механізм наростання ризиків, коли початкові інформаційні спотворення під час подальшої алгоритмічної інтерпретації не лише зберігаються, а й посилюються, що в підсумку зумовлює ухвалення управлінських рішень із потенційно негативними наслідками для фінансової стабільності та довгострокових цілей розвитку [3]. Графічне відображення зазначеної логіки ескалації ризиків представлено на рис. 2.

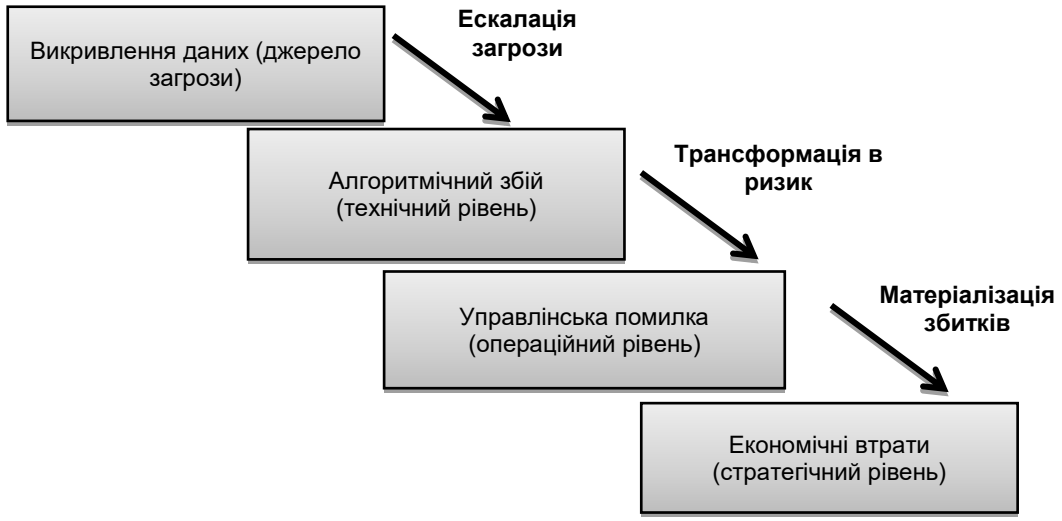

Рис. 2. Модель ескалації інформаційних ризиків в економічні збитки

Джерело: розроблено автором самостійно

За умови використання викривленої інформаційної бази, інтелектуальна система може неправильно класифікувати економічно стабільного партнера як джерело підвищеної загрози внаслідок штучно сформованого негативного інформаційного середовища, що спричиняє

необґрунтоване припинення ділових взаємодій і втрату потенційних доходів. Паралельно з цим зберігається і протилежний ризик – неспроможність своєчасно виявити фінансові загрози, якщо вхідні дані були цілеспрямовано спотворені або якщо модель сформувала надмірно оптимістичні аналітичні висновки, що може мати наслідком прямі економічні збитки.

Ще більш критичними наслідки використання недостовірної інформації стають на стратегічному рівні управління, де ухвалення рішень відбувається в умовах підвищеної складності та невизначеності. У випадку, коли джерелом цієї невизначеності виступає сам аналітичний інструмент, порушується стійкість системи довгострокового планування. З огляду на ключову роль інформаційно-аналітичного забезпечення у підтриманні фінансової рівноваги підприємства, стратегічні рішення, ухвалені на основі некоректних або «галюцинаторних» результатів роботи моделі, можуть спрямувати розвиток компанії у помилковому напрямі та знизити її конкурентні позиції [2].

З метою підтримання надійності системи управління ризиками на підприємстві доцільним є формування інтегрованого механізму нейтралізації інформаційних загроз, що ґрунтується на поєднанні сучасних технологічних рішень із регламентованими процедурами контролю.

У цьому контексті інструменти зниження інформаційних ризиків у системах штучного інтелекту під час роботи з даними доцільно класифікувати відповідно до стадій, на яких вони використовуються:

– інформаційно-верифікаційні, впроваджуються на етапі первинного контролю та очищення інформації;

– алгоритмічно-інтерпретаційні, застосовуються безпосередньо в роботі алгоритмів опрацювання даних;

– організаційно-процедурні, використовуються на стадії підсумкової перевірки результатів.

Узагальнену характеристику зазначених інструментів подано в табл. 2.

Таблиця 2

Інструменти мінімізації ризиків недостовірності інформації в системах управління на основі штучного інтелекту

Група інструментів	Інструменти	Механізм реалізації
1. Інформаційно-верифікаційні	Перехресна верифікація	Зіставлення внутрішньої корпоративної інформації з незалежними зовнішніми реєстрами та базами даних з метою виявлення логічних і фактичних невідповідностей.
	OSINT-валідація	Застосування інструментів аналізу відкритих джерел не для накопичення інформації, а для підтвердження її достовірності та ідентифікації ознак інформаційного впливу або дезінформації.
	Попередня автоматизована фільтрація	Використання допоміжних моделей машинного навчання для виявлення статистичних відхилень і нетипових патернів у масивах вхідних даних до їх обробки основним ШІ-модулем.
2. Алгоритмічно-інтерпретаційні	Explainable AI (XAI)	Залучення моделей, здатних розкривати логіку формування результатів, що дає змогу ідентифікувати помилкові причинно-наслідкові зв’язки та зменшити ризик генерації недостовірних висновків.
	Альтернативне сценарне прогнозування	Формування кількох можливих варіантів розвитку подій замість єдиного прогнозу, що підвищує адаптивність управлінських рішень в умовах невизначеності.
3. Організаційно-процедурні	Валідація експертом	Обов’язкове залучення фахівця з управління ризиками до оцінювання результатів, які мають стратегічні наслідки, з урахуванням пояснень, наданих інтерпретованими моделями.
	Юридична експертиза джерел інформації	Аналіз правомірності використання інформаційних ресурсів відповідно до норм законодавства з метою зниження правових і репутаційних ризиків.

Джерело: розроблено автором самостійно

Основою механізму мінімізації ризиків від недостовірної інформації має стати перехід від використання моделей типу «чорна скринька» до концепції «пояснюваного штучного інтелекту». Технології Explainable AI дозволяють інтерпретувати результати роботи складних алгоритмів, роблячи процес прийняття рішення прозорим для людини-оператора.

Запропонована модель комплексної системи мінімізації ризиків недостовірності інформації, що поєднує верифікацію даних на вході та інтерпретацію рішень на виході, візуалізована на рис. 3.

Запровадження запропонованої моделі безпеки трансформує сам підхід до використання штучного інтелекту в управлінській діяльності. Управлінський персонал переходить від некритичного сприйняття результатів автоматизованого аналізу до формату усвідомленого контролю, за якого рішення, сформовані ШІ, підлягають обов’язковій перевірці та інтерпретації. У такій конфігурації штучний інтелект функціонує не як автономне джерело істини, а як інструмент підтримки прийняття рішень із прозорою логікою роботи.

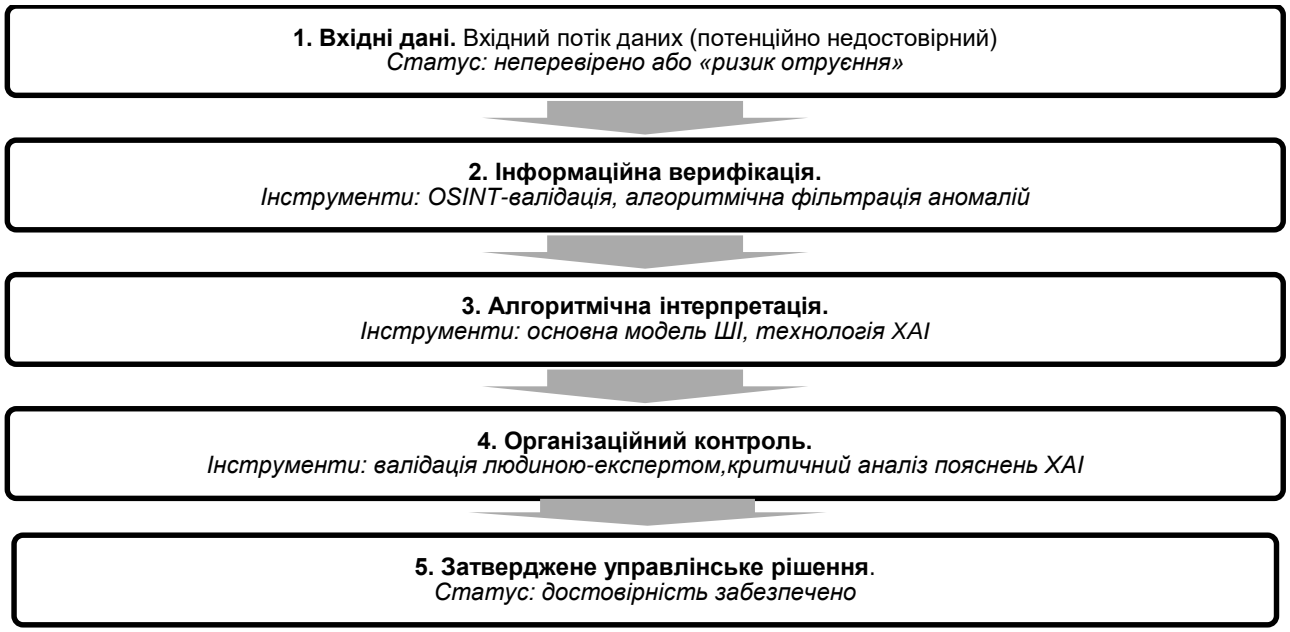

Рис. 3. Концептуальна схема багаторівневої системи мінімізації ризиків недостовірності інформації

Джерело: розроблено автором самостійно

Поєднання високої швидкості обробки великих масивів даних, забезпеченої інтерпретованими моделями, з експертною оцінкою людини дає змогу інтегрувати ризики цифрового інформаційного середовища в систему керованих операційних ризиків підприємства, знижуючи їх критичний вплив.

Висновки з проведеного дослідження. Достовірність інформації в цифровому середовищі виступає ключовою умовою результативного функціонування систем управління ризиками, що використовують інструментарій штучного інтелекту. У ході дослідження встановлено, що за умов надлишку слабоструктурованих та неоднорідних інформаційних потоків формується ефект зворотної дії, оскільки алгоритми, покликані мінімізувати невизначеність у разі опрацювання викривлених даних самі генерують додаткові загрози для стратегічного розвитку підприємства.

Узагальнення деструктивних чинників дало змогу згрупувати їх за походженням на зовнішні впливи, пов'язані з маніпуляціями інформаційним середовищем (зокрема через цілеспрямоване спотворення наборів даних і деформацію OSINT-джерел), та внутрішні обмеження алгоритмічної природи, що виявляються у вигляді помилкових висновків генеративних моделей. Обґрунтовано, що некритичне використання результатів роботи таких систем ініціює поетапне нарощування ризиків – від помилок автоматизованої оцінки на операційному рівні до хибних управлінських рішень у довгостроковому плануванні, що підривають фінансову стійкість підприємства.

У зв'язку з цим, доведено доцільність зміни підходу до використання штучного інтелекту, як і від сприйняття його висновків щодо готових управлінських рішень, так і до концепції постійної перевірки та інтерпретації результатів. Запропонований механізм зниження інформаційних ризиків передбачає інтеграцію багаторівневої фільтрації вхідних даних, застосування інтерпретованих моделей Explainable AI для підвищення прозорості алгоритмічної логіки та залучення експертної оцінки при ухваленні рішень із високим рівнем відповідальності. Реалізація такого підходу забезпечує баланс між аналітичними можливостями автоматизованих систем і вимогами до надійності управлінських рішень.

Подальші дослідження доцільно спрямувати на формування гібридних інструментів перевірки достовірності інформації, що поєднуватимуть автоматизоване виявлення синтетичного контенту з процедурами експертного аудиту, а також на розроблення адаптивних моделей оцінювання надійності інформаційних джерел для їх інтеграції в корпоративні системи управління ризиками.

Література

1. Про Національну програму інформатизації : Закон України від 01.12.2022 № 2807-IX. URL: https://zakon.rada.gov.ua/laws/show/2807-20#Text (дата звернення: 30.08.2025).

2. Десятнюк О. М. Птащенко О. В. Управління ризиками в цифровій економіці: фінансова безпека та трансформаційні зміни. *Європейський науковий журнал Економічних та Фінансових інновацій.* 2024. Т. 2. № 14. С. 238-247. DOI: https://doi.org/10.32750/2024-0223.

3. Загрози та ризики використання штучного інтелекту / Скіцько О., Складанний П., Ширшов Р., Гуменюк М., Ворохоб М. *Кібербезпека: освіта, наука, техніка*. 2023. Т. 2. № 22. С. 6-18. DOI: https://doi.org/10.28925/2663-4023.2023.22.618.

4. Мельник А., Винокуров Р. Особливості прийняття управлінських рішень, їх оцінка та оцінювання в сучасних умовах. *Економіка та суспільство*. 2025. Випуск № 71. DOI: https://doi.org/10.32782/2524-0072/2025-71-166.

5. Онищенко С. В., Глушко А. Д. Інформаційно-аналітичне забезпечення фінансової безпеки підприємств у сучасних умовах. *Науковий Вісник Одеського національного економічного університету.* 2023. № 7–8. С. 145-154. DOI: https://doi.org/10.32680/2409-9260-2023-7-8-308-309-145-154.

6. Лесько Н. В. Особливості реалізації принципу достовірності інформації. *Науково-інформаційний вісник Івано-Франківського університету права імені Короля Данила Галицького: Журнал. Серія Право.* 2023. Вип. 15(27). Т. 1. С. 140-144.

7. Кравчук О. Процедура ідентифікації ризиків впровадження штучного інтелекту в публічне управління. *Науковий вісник Вінницької академії безперервної освіти. Серія «Екологія. Публічне управління та адміністрування»*. 2025. № 1(7). С. 122-126. DOI: https://doi.org/10.32782/2786-5681-2025-1.15.

8. Івкова В., Опірський І. OSINT-технології як загроза кібербезпеці держави. *Кібербезпека: освіта, наука, техніка*. 2025. Т. 3. № 27. С. 165-179. DOI: https://doi.org/10.28925/2663-4023.2025.27.749.

9. Махно Є., Руденко Є., Судніков Є., Тищенко М. Галюцинації штучного інтелекту у сфері освіти та науки: причини, наслідки та методи мінімізації. *Повітряна міць України*. 2025. Т. 1. № 8. С. 111-126. DOI: https://doi.org/10.33099/2786-7714-2025-1-8-111-126.

10. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI / A. Barredo Arrieta et al. *Information fusion*. 2020. Vol. 58. P. 82-115. DOI: https://doi.org/10.1016/j.inffus.2019.12.012.

References

1. The Verkhovna Rada of Ukraine (2022), The Law of Ukraine “On the National Informatization Program” dated 01.12.2022 no. 2807-IX, available at: https://zakon.rada.gov.ua/laws/show/2807-20#Text (access date August 30, 2025).

2. Desiatniuk, O.M. and Ptashchenko, O.V. (2024), “Risk management in the digital economy: financial security and transformational changes”, *Yevropeiskyi naukovyi zhurnal Ekonomichnykh ta Finansovykh innovatsii*, Vol. 2, no. 14, pp. 238-247, DOI: https://doi.org/10.32750/2024-0223.

3. Skitsko, O., Skladannyi, P., Shyrshov, R. et al. (2023), “Threats and risks of using artificial intelligence”, *Kiberbezpeka: osvita, nauka, tekhnika*, Vol. 2, no. 22, pp. 6-18, DOI: https://doi.org/10.28925/2663-4023.2023.22.618.

4. Melnyk, A. and Vynokurov, R. (2025), “Features of management decision-making, their assessment and evaluation in modern conditions”, *Ekonomika ta suspilstvo*, Issue no. 71, DOI: https://doi.org/10.32782/2524-0072/2025-71-166.

5. Onyshchenko, S.V. and Hlushko, A.D. (2023), “Information and analytical support of financial security of enterprises in modern conditions”, *Naukovyi visnyk Odeskoho natsionalnoho ekonomichnoho universytetu*, no. 7–8, pp. 145-154, DOI: https://doi.org/10.32680/2409-9260-2023-7-8-308-309-145-154.

6. Lesko, N.V. (2023), “Features of implementation of the principle of reliability of information”, *Naukovo-informatsiinyi visnyk Ivano-Frankivskoho universytetu prava imeni Korolia Danyla Halytskoho: Zhurnal. Seriia Pravo*, Iss. 15(27), Vol. 1, pp. 140-144.

7. Kravchuk, O. (2025), “Procedure for identifying risks of introducing artificial intelligence into public administration”, *Naukovyi visnyk Vinnytskoi akademii bezperervnoi osvity. Seriia “Ekolohiia. Publichne upravlinnia ta administruvannia”*, no. 1(7), pp. 122-126, DOI: https://doi.org/10.32782/2786-5681-2025-1.15.

8. Ivkova, V. and Opirskyi, I. (2025), “OSINT technologies as a threat to the state’s cybersecurity”, *Kiberbezpeka: osvita, nauka, tekhnika*, Vol. 3, no. 27, pp. 165-179, DOI: https://doi.org/10.28925/2663-4023.2025.27.749.

9. Makhno, Ye., Rudenko, Ye, Sudnikov, Ye. and Tyshchenko, M. (2025), “Artificial intelligence hallucinations in education and science: causes, consequences and minimization methods”, *Povitriana mits Ukrainy*, Vol. 1, no. 8, pp. 111-126, DOI: https://doi.org/10.33099/2786-7714-2025-1-8-111-126.

10. Barredo Arrieta, A. et al. (2020), “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”, *Information Fusion*, Vol. 58, pp. 82-115, DOI: https://doi.org/10.1016/j.inffus.2019.12.012.

Шухманн В.А.

НЕДОСТОВІРНІСТЬ ІНФОРМАЦІЇ ЦИФРОВОГО ПРОСТОРУ ЯК ВИКЛИК ШТУЧНОМУ ІНТЕЛЕКТУ В ПРОЦЕСІ УПРАВЛІННЯ РИЗИКАМИ

Мета. Дослідження впливу недостовірної інформації та дезінформаційних потоків цифрового середовища на ефективність використання штучного інтелекту в системі управління ризиками

підприємства, а також обґрунтування багаторівневих підходів до перевірки даних задля підвищення якості стратегічних управлінських рішень.

Методика дослідження. У процесі дослідження застосовано поєднання загальнонаукових і спеціальних методів. Системний аналіз використано для структурування інформаційних загроз, аналіз і синтез – для встановлення взаємозв'язку між якістю даних та помилками алгоритмів, графічне моделювання – для відображення логіки ескалації ризиків, а методи узагальнення і класифікації – для впорядкування інструментів зниження інформаційних ризиків.

Результати дослідження. Визначено ключові загрози цифрового інформаційного середовища, що впливають на надійність рішень, сформованих із застосуванням штучного інтелекту. Доведено, що ці загрози мають зовнішній характер, пов'язаний з маніпулюванням інформаційними потоками, та внутрішній, обумовлений особливостями роботи генеративних моделей. Обґрунтовано, що спотворення даних запускає ланцюгову реакцію ризиків, яка проявляється у помилках оцінювання на операційному рівні та призводить до стратегічних прорахунків у довгостроковому розвитку підприємства. Запропоновано концептуальну модель мінімізації ризиків, що передбачає поєднання попередньої перевірки даних, інтерпретації результатів роботи алгоритмів та обов'язкової експертної оцінки управлінських рішень.

Наукова новизна результатів дослідження. Подальшого розвитку набуло теоретичне обґрунтування забезпечення інформаційної надійності в системах управління ризиками на основі штучного інтелекту. Запропоновано багаторівневу класифікацію методів верифікації даних відповідно до етапів їх обробки та розроблено модель каскадного перетворення інформаційних спотворень у стратегічні втрати підприємства.

Практична значущість результатів дослідження. Результати дослідження можуть бути використані у діяльності підприємств для вдосконалення систем підтримки прийняття рішень, зниження фінансових втрат від дезінформації та підвищення стійкості стратегічного управління в умовах цифрової невизначеності.

Ключові слова: штучний інтелект, ефективність, цифрова економіка, недостовірність інформації, галюцинації ШІ, Explainable AI, стратегічне управління, верифікація даних.

Schuchmann V.A.

UNRELIABILITY OF INFORMATION IN THE DIGITAL SPACE AS A CHALLENGE TO ARTIFICIAL INTELLIGENCE IN THE PROCESS OF RISK MANAGEMENT

Purpose. The aim of the article is to investigate the impact of unreliable information and disinformation flows in the digital environment on the efficiency of using artificial intelligence in the enterprise risk management system, as well as to substantiate multi-level approaches to data verification in order to improve the quality of strategic management decisions.

Methodology of research. A combination of general scientific and special methods was applied in the research process. System analysis was used to structure information threats; analysis and synthesis – to establish the relationship between data quality and algorithmic errors; graphic modelling – to display the logic of risk escalation; methods of generalization and classification – to organize tools for reducing information risks.

Findings. Key threats of the digital information environment affecting the reliability of decisions formed using artificial intelligence are identified. It is proved that these threats have an external nature related to the manipulation of information flows, and an internal one conditioned by the peculiarities of generative models' operation. It is substantiated that data distortion triggers a chain reaction of risks, which manifests in assessment errors at the operational level and leads to strategic miscalculations in the long-term development of the enterprise. A conceptual model of risk minimization is proposed, which involves a combination of preliminary data verification, interpretation of algorithm results, and mandatory expert assessment of management decisions.

Originality. The theoretical substantiation of ensuring information reliability in AI-based risk management systems has received further development. A multi-level classification of data verification methods according to their processing stages is proposed, and a model of the cascade transformation of information distortions into strategic losses of the enterprise is developed.

Practical value. The research results can be used in the activities of enterprises to improve decision support systems, reduce financial losses from disinformation, and increase the resilience of strategic management in conditions of digital uncertainty.

Key words: artificial intelligence, efficiency, digital economy, information unreliability, AI hallucinations, Explainable AI, strategic management, data verification.

Дата надходження рукопису: 17.10.2025
Дата прийняття рукопису до друку: 24.11.2025
Дата публікації: 26.12.2025