https://doi.org/10.31891/2219-9365-2025-84-25
УДК 004.8:004.738.5:004.056:174

ЧЕРНИШОВА Олена
Дніпровський національний університет імені Олеся Гончара
https://orcid.org/0000-0001-9437-9100
chernyshova.elen@gmail.com
ШИБКО Оксана
ННІ "Придніпровська державна академія будівництва та архітектури",
Український державний університет науки і технології
https://orcid.org/0000-0001-5894-0642
ksusha2708@gmail.com
ЖОЛОНДКІВСЬКИЙ Максим
Український науково-дослідний інститут Служби безпеки України
https://orcid.org/0009-0000-8015-7419
gafovskiy@gmail.com

AI-МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ ПОВЕДІНКИ КОРИСТУВАЧІВ У ЦИФРОВИХ ЕКОСИСТЕМАХ: ЕТИЧНІ ТА ТЕХНІЧНІ ВИКЛИКИ

Дослідження спрямоване на аналіз технічних та етичних викликів, пов'язаних із застосуванням моделей штучного інтелекту для прогнозування поведінки користувачів у цифрових екосистемах. Автор розглядає проблеми якості даних, інтерпретованості, масштабованість алгоритмів, стійкості до атак, а також питання конфіденційності, алгоритмічного упередження, прозорості рішень, відповідальності за прогнози та ризики маніпуляції. Визначено підходи до створення транспарентних, справедливих та захищених прогностичних систем на основі порівняння моделей машинного навчання.

Використано літературний огляд сучасних публікацій з акцентом на технічні методи (машинне навчання, глибинне навчання, методи Монте-Карло, когнітивну аналітику) та етичні аспекти (SWOT-аналіз стратегій впровадження AI, аудит упереджень). Порівняльний аналіз двох моделей: Random Forest (ансамблевий метод з 100 деревами) та Neural Network MLP (архітектура 64-32-16 нейронів). Експерименти проведено на синтетичних даних, що імітують поведінкові патерни, з оцінкою точності, справедливості (метрики Disparate Impact Ratio, Equalized Odds, Equal Opportunity, Statistical Parity), диференційної приватності (ε=1.0) та дрейфу моделі. Застосовано візуалізацію (часові ряди, багатопанельні графіки) та табличний синтез викликів.

Стаття інтегрує технічні та етичні аспекти в єдиний аналіз, квантифікуючи упередження та деградацію моделей у динамічних середовищах. Запропоновано комплексний підхід до мітигації проблем через алгоритмічний аудит, explainable AI, партисипативний дизайн та диференційну приватність, що перевершує фрагментарні дослідження. Виявлено прогалини в існуючих методологіях, зокрема відсутність вбудованих етичних обмежень на рівні дизайну алгоритмів та адаптації до еволюції стандартів.

Random Forest показала точність 65,50% на тестовій вибірці з overfitting (розрив 20-25%), Neural Network – 58,33% з низькою інтерпретативністю. Алгоритмічне упередження підтверджено: Disparate Impact Ratio 0,4858-0,5311 для недопредставлених груп, розбіжності в accuracy (77,08% vs 52,33-55,91%). Деградація моделі: з 65,5% до 42,6% за 12 місяців. Диференційна приватність знижує точність на 3-5%. Таблиця систематизує виклики (упередження, конфіденційність, інтерпретативність) з методами мітигації (аудит, XAI, розподілені обчислення).

AI-моделі ефективні для прогнозування, але потребують балансу між точністю та етикою через упередження, деградацію та непрозорість. Необхідні інтегровані фреймворки з вбудованими етичними принципами, безперервним моніторингом та культурною інклюзивністю. Перспективи: розробка адаптивних архітектур, формальних метрик етичності та міждисциплінарних стандартів для безпечних цифрових систем.

Ключові слова: поведінка споживачів, штучний інтелект, нейронна мережа, машинне навчання.

CHERNYSHOVA Olena
Oles Honchar Dnipro National University
SHYBKO Oksana
"Prydniprovska State Academy of Civil Engineering and Architecture"
Ukrainian State University of Science and Technologies SEI
ZHOLONDKIVSKYI Maksym
Institute of Special Equipment and Forensic Expertise
of the Security Service of Ukraine

AI MODELS FOR PREDICTING USER BEHAVIOR IN DIGITAL ECOSYSTEMS: ETHICAL AND TECHNICAL CHALLENGES

The study aims to analyze the technical and ethical challenges associated with the application of artificial intelligence models for predicting user behavior in digital ecosystems. The author examines issues related to data quality, interpretability, algorithm scalability, resilience to attacks, as well as concerns regarding privacy, algorithmic bias, decision transparency, accountability for predictions, and risks of manipulation. Approaches to developing transparent, fair, and secure predictive systems are identified based on a comparison of machine learning models.

A literature review of contemporary publications was conducted, focusing on technical methods (machine learning, deep learning, Monte Carlo methods, cognitive analytics) and ethical aspects (SWOT analysis of AI implementation strategies, bias auditing). A comparative analysis of two models was performed: Random Forest (an ensemble method with 100 trees) and Neural Network MLP

(architecture with 64-32-16 neurons). Experiments were conducted on synthetic data mimicking behavioral patterns, evaluating accuracy, fairness (metrics such as Disparate Impact Ratio, Equalized Odds, Equal Opportunity, Statistical Parity), differential privacy (ε=1.0), and model drift. Visualization (time series, multi-panel graphs) and tabular synthesis of challenges were employed.

The article integrates technical and ethical aspects into a unified analysis, quantifying bias and model degradation in dynamic environments. A comprehensive approach to mitigating issues through algorithmic auditing, explainable AI, participatory design, and differential privacy is proposed, surpassing fragmented studies. Gaps in existing methodologies are identified, particularly the lack of embedded ethical constraints at the algorithm design level and adaptation to evolving standards.

Random Forest achieved an accuracy of 65.50% on the test set with overfitting (20-25% gap), while the Neural Network reached 58.33% with low interpretability. Algorithmic bias was confirmed: Disparate Impact Ratio of 0.4858–0.5311 for underrepresented groups, with accuracy disparities (77.08% vs. 52.33–55.91%). Model degradation was observed: from 65.5% to 42.6% over 12 months. Differential privacy reduced accuracy by 3–5%. A table systematizes challenges (bias, privacy, interpretability) with mitigation methods (auditing, XAI, distributed computing).

AI models are effective for prediction but require a balance between accuracy and ethics due to bias, degradation, and opacity. Integrated frameworks with embedded ethical principles, continuous monitoring, and cultural inclusivity are needed. Future directions include developing adaptive architectures, formal ethical metrics, and interdisciplinary standards for secure digital systems.

Keywords: consumer behavior, artificial intelligence, neural network, machine learning.

Стаття надійшла до редакції / Received 05.11.2025
Прийнята до друку / Accepted 02.12.2025

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Особливістю нового типу цифрової екосистеми є надзвичайна складність взаємодії між користувачами та технологічною платформою, потоками інформації. Передбачуваність активності користувачів у таких середовищах стала одним із першочергових завдань інформаційних технологій, що можна пояснити необхідністю зробити цифрові послуги більш персоналізованими для максимізації досвіду клієнтів та підвищення ефективності онлайн-систем. Галузь штучного інтелекту (Artificial intelligence - AI) відкриває можливості аналізу великих обсягів інформації про поведінку користувачів, виявлення певних тенденцій, які неможливо передбачити, та створення точних прогностичних моделей.

Впровадження машинного навчання, глибокого навчання та обробки природної мови сприяло розвитку прогресивно вдосконалених прогностичних систем, які здатні адаптуватися до динамічних змін у поведінці людей, що використовують онлайн-платформи. Тим часом, збільшення обчислювальної потужності та доступ до великих наборів даних забезпечили умови для створення складних нейронних структур, які демонструють високу точність в аналізі дій, смаків та намірів користувачів.

Особливий інтерес викликає питання алгоритмічної упередженості; цілком обґрунтовано припустити, що моделі машинного навчання, навчаючись на історичних даних, можуть відтворювати та підтримувати збереження вже усталеної соціальної нерівності та часто використовуваних стереотипів. Це призводить до дискримінації деяких груп користувачів, що не відповідає ідеї справедливості та рівності в цифровому світі. Крім того, це може призвести до феномену «чорної скриньки», оскільки неможливо пояснити, чому система прийняла саме таке рішення, оскільки складна архітектура штучного інтелекту не є прозорою, і тому важко перевірити помилки та недоліки.

Питання конфіденційності даних є особливо проблематичним у сценаріях, коли є потреба передбачити поведінку користувача, оскільки в цій ситуації застосування таких систем передбачає збір та обробку персональних даних щодо дій користувача в інтернет-просторі. Необхідність знайти баланс між продуктивним застосуванням прогностичних моделей та збереженням конфіденційності створює складні технічні та організаційні проблеми. Також зазвичай недостатньо використовувати метод анонімізації даних, оскільки дані, незважаючи на можливості деідентифікації, все ще можуть бути повторно ідентифіковані будь-коли, якщо є певна додаткова інформація.

Інший важливий аспект цього питання пов'язаний з динамізацією поведінки користувачів, яка ніколи не буває статичною через вплив численних факторів: технологічного прогресу, соціальних тенденцій, культурних подій. Навчені прогнозні моделі можуть легко застаріти, і в результаті необхідно розробляти адаптивні алгоритми, здатні швидко реагувати на зміни в трафіку та активності користувачів. Це спонукає дослідників до роботи над створенням універсальної системи машинного навчання, яка має інструменти безперервного навчання та самокорекції.

Таким чином, вивчення AI як предиктора поведінки користувачів у цифрових екосистемах та того, як на цифрові технології впливають етичні та технічні проблеми, є актуальною науковою та практичною роботою, яка матиме важливе значення для створення безпечних, справедливих та ефективних цифрових технологій.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Проблема застосування штучного інтелекту для сприяння та прогнозування поведінки користувачів в Інтернеті активно досліджується в сучасній науковій літературі, що охоплює її технічні, методологічні та етичні аспекти.

Семененко Ю. [1, с. 476-478] досліджує, як технології штучного інтелекту можуть бути використані для аналізу поведінки споживачів у сфері електронної комерції на прикладі машинного навчання, обробки природної мови, комп'ютерного зору, когнітивної аналітики та систем рекомендацій. Автор також наголошує на тому, що використання таких технологій може допомогти підвищити точність аналізу поведінки, ефективну персоналізацію взаємодії та покращення клієнтського досвіду, а також визначити ключові ризики, пов'язані із захистом персональних даних, етичними проблемами та технічними обмеженнями. У статті доводиться можливість об'єднання інтелектуальних систем з цифровими, враховуючи багатовимірність аналізу поведінкових моделей.

Адамик В., Івановський О. [2, с. 234-237] обговорюють три підходи до використання штучного інтелекту в маркетингових системах: повна відмова від використання ШІ, обмежене контрольоване використання та необмежене використання. Дослідники використовують SWOT-аналіз для оцінки кожної стратегії та визнають впровадження обмежених та контрольованих технологій найкращим рішенням, яке дозволить досліднику отримати максимальну віддачу від штучного інтелекту та звести до мінімуму ризики. Особливий акцент робиться на етичних питаннях, таких як упередження, дискримінація, комп'ютерна безпека та необхідність законодавчого регулювання використання персональних даних відповідно до міжнародних стандартів.

Мисюк І., Мисюк Р. [3, с. 121-123] досліджують застосування методів Монте-Карло для моделювання та прогнозування поведінки користувачів у соціальних мережах. Автори розглядають декілька варіантів цього підходу, включаючи квазі-Монте-Карло та метод Метрополіс-Гастингс, які продемонстрували кращу ефективність завдяки рівномірнішому охопленню сценаріїв поведінки користувачів. Дослідження відображає збіжність методів Монте-Карло з вибіркою у 10000 значень на основі параметрів активності користувачів, таких як лайки, поширення та коментарі, що дозволяє розробити базову систему прогнозної аналітики для підвищення персоналізації контенту.

Приходько О. М., Скоробогатова А. О., Семенова Л. Ю. [4, с. 151-153] обґрунтовують використання штучного інтелекту як інструменту для вивчення поведінки сучасних споживачів, акцентуючи увагу на методах машинного та глибинного навчання, аналізу великих даних та нейронних мережах. Дослідники пропонують адаптований підхід налаштування нейронних мереж для аналізу поведінки з використанням механізму когнітивної обробки, що дозволяє моделювати складні патерни прийняття рішень споживачами. Особливу увагу приділено етичним аспектам, зокрема ризикам, пов'язаним з конфіденційністю даних та можливою маніпуляцією емоціями користувачів, для мінімізації яких рекомендується дотримання прозорості у процесі обробки персональної інформації.

Vilkhivska O. [5, с. 2-4] аналізує ключові етичні аспекти використання штучного інтелекту, включаючи проблеми відповідальності за рішення, прийняті машинами, упередженості в алгоритмах, автономії та контролю, а також соціально-економічні наслідки впровадження AI. Дослідниця підкреслює важливість міжнародних ініціатив з регулювання штучного інтелекту, спрямованих на забезпечення прозорості, безпеки та дотримання прав людини. Особливу увагу приділено принципам розробки справедливих та безпечних AI-систем, що враховують транспарентність, підзвітність та недискримінацію, наголошуючи на необхідності балансу між технологічним прогресом та моральною відповідальністю.

У статті, автором якої є Venkat Sanka [6, с. 732-734], розглядаються етичні проблеми штучного інтелекту в середовищі великих даних з урахуванням проблеми алгоритмічної упередженості, справедливості та прозорості. Автор пропонує безшовний формат, який полягає в інтеграції алгоритмічного аудиту та методів пояснення ШІ, а також інструментів дотримання нормативних вимог для виявлення упередженості, оптимізації справедливості та підвищення прозорості. Результати впровадженого підходу показують, що можливо пом'якшити упередженість в алгоритмі не менш ніж на 67 відсотків, не впливаючи на продуктивність систем, що доводить ефективність двостороннього вирішення етичних проблем, що виникають у сфері систем на основі ШІ, що працюють з великими даними. Thalpage N., Epa D., Jayawardhana S. [7, с. 31-35] розглядають етичні проблеми в пояснювальному штучному інтелекті з особливим акцентом на культурний та соціальний етноцентризм. Дослідники виявили переважний акцент на західних групах користувачів, низьку активність щодо недостатньо представлених груп та відсутність партисипативних або культурно чутливих практик проектування. Автори досліджують застосування систем підтримки рішень у медицині, освіті та медицині, а також зазначають обмеження сучасних моделей, необхідність інклюзивних та орієнтованих на користувача моделей, таких як локалізовані моделі пояснення, партисипативний дизайн та покращені показники оцінки, враховуючи культурну релевантність та соціальну справедливість.

Dong Y., Guo J. [8, с. 752-758] обговорюють ризики упередженості та проблеми з етикою в системах штучного інтелекту та виділяють два аспекти упередженості, які включають упередженість у системах штучного інтелекту, створену людьми, та упередженість, яку створюють люди. Ці упередження згадуються авторами, які стверджують, що для зменшення цих упереджень та досягнення справедливості слід розробити надійні етичні рамки. У статті висвітлюється, як штучний інтелект трансформує демократичні практики, застосовні етичні принципи та ефективну політику для пом'якшення упередженості, включаючи підвищення різноманітності в розробці та правовому регулюванні штучного інтелекту, або необхідність постійної оцінки

впливу етики штучного інтелекту для забезпечення відповідальної інтеграції штучного інтелекту. Gono A. wijaya, Mailangkay A. B. L. [9, с. 241-243] досліджують взаємозв'язок між користувацьким досвідом та етичними проблемами при впровадженні штучного інтелекту в онлайн-шопінгу. Автори аналізують, як технологічні можливості AI впливають на прийняття рішень споживачами та їхнє сприйняття етичності систем. Дослідження демонструє, що позитивний користувацький досвід може бути досягнутий лише за умови забезпечення прозорості AI-систем та дотримання етичних стандартів, що впливає на довіру споживачів до технологічних платформ та їхню готовність до взаємодії з AI-керованими сервісами.

Nelson J., Baimagambetov A., Avgerinakis K., Polatidis N. [10, с. 2-5] пропонують колаборативну фільтрацію як спосіб покращення етичних рекомендацій щодо підказок у великих мовних моделях під час генерації коду AI. Автори наголошують, що класичні способи гарантування етики AI, наприклад, фільтрація на основі правил та модерація людиною, мають проблеми масштабованості та гнучкості. Семантична подібність та механізми зворотного зв'язку дозволяють системі вивчати поведінку користувачів та пропонувати підказки, які відповідають етичним стандартам і відхиляються через їх потенціал створювати упереджені та небезпечні моделі AI, шляхом застосування колаборативної фільтрації.

Метою дослідження є аналіз технічних та етичних викликів, пов'язаних із застосуванням AI-моделей для прогнозування поведінки користувачів у цифрових екосистемах, та розробка засобів створення відкритих, справедливих та безпечних прогностичних технологій. Дослідження базується на порівняльному аналізі двох класів моделей машинного навчання – Random Forest (ансамблевий метод) та Neural Network MLP (глибинне навчання) – для визначення їхніх можливостей прогнозування та інтерпретації, а також вразливості до етичних та технічних проблем з точки зору справедливості алгоритмічного рішення, диференціальної конфіденційності персональних даних та реагування на концептуальний дрейф у мінливих цифрових середовищах.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Технологічна основа моделей штучного інтелекту для прогнозування поведінки користувачів керується на основі логічного поєднання інструментів машинного навчання, глибокого навчання та аналізу великих даних. За словами Семененка Ю. [1, с. 480-481], існує п'ять помітних технологічних елементів, а саме: машинне навчання для побудови моделей прогнозування поведінки, обробка природної мови для аналізу неструктурованих текстових даних, комп'ютерний зір для ідентифікації візуального контенту, рекомендаційні системи для надання персоналізованих пропозицій та когнітивна аналітика для створення систем, які можуть імітувати людську поведінку. Ці технології створюють багаторівневу інфраструктуру прогнозування, кожен учасник якої відповідає за певні елементи аналізу поведінки користувачів.

Методологічні підходи до прогнозування поведінки користувачів демонструють значну різноманітність. Мисюк І., Мисюк Р. [3, с. 122-124] застосовують методи Монте-Карло для моделювання поведінки у соціальних мережах, досягаючи високої точності прогнозування через рівномірне охоплення сценаріїв користувацької активності. Квазі-Монте-Карло методи та алгоритм Метрополіс-Гастингс показали оптимальну збіжність на вибірці 10000 значень при моделюванні параметрів лайків, поширень та коментарів, що підтверджує ефективність стохастичних методів для аналізу динамічних систем з високою варіативністю користувацької поведінки.

Для систематизації підходів до використання AI у прогнозуванні поведінки користувачів доцільно розглянути альтернативні стратегії впровадження. Adamyk V., Ivanovskyi O. [2, с. 235-238] ідентифікують три базові стратегії: повну відмову від AI, обмежене контрольоване застосування та необмежене використання. SWOT-аналіз цих стратегій виявив, що обмежене контрольоване застосування забезпечує оптимальний баланс між максимізацією переваг AI та мінімізацією ризиків, дозволяючи підвищити продуктивність, створити нові професії, пов'язані з контролем AI, та знизити загрози помилок, хоча це може обмежувати інновації через бюрократичні та регуляторні вимоги.

Архітектура нейронних мереж для аналізу поведінки користувачів потребує особливої уваги до когнітивної обробки даних. Приходько О. М., Скоробогатова А. О., Семенова Л. Ю. [4, с. 152-154] пропонують адаптований підхід налаштування нейронних мереж з використанням механізму когнітивної обробки, що розглядає нейромережу як інструмент моделювання поведінки без заглиблення у внутрішні структурні деталі. Цей підхід включає три етапи: збір інформації через автоматизоване агрегування даних з веб-сайтів, мобільних додатків, соціальних мереж та IoT-сенсорів; обробку та сегментацію даних за критеріями вікових груп, інтересів та купівельних звичок; аналіз і прогнозування на основі виявлення трендів та створення передбачувальних моделей поведінки.

Для наочної демонстрації взаємозв'язків між компонентами AI-системи прогнозування поведінки користувачів представимо структурну схему основних технологічних блоків (рис. 1).

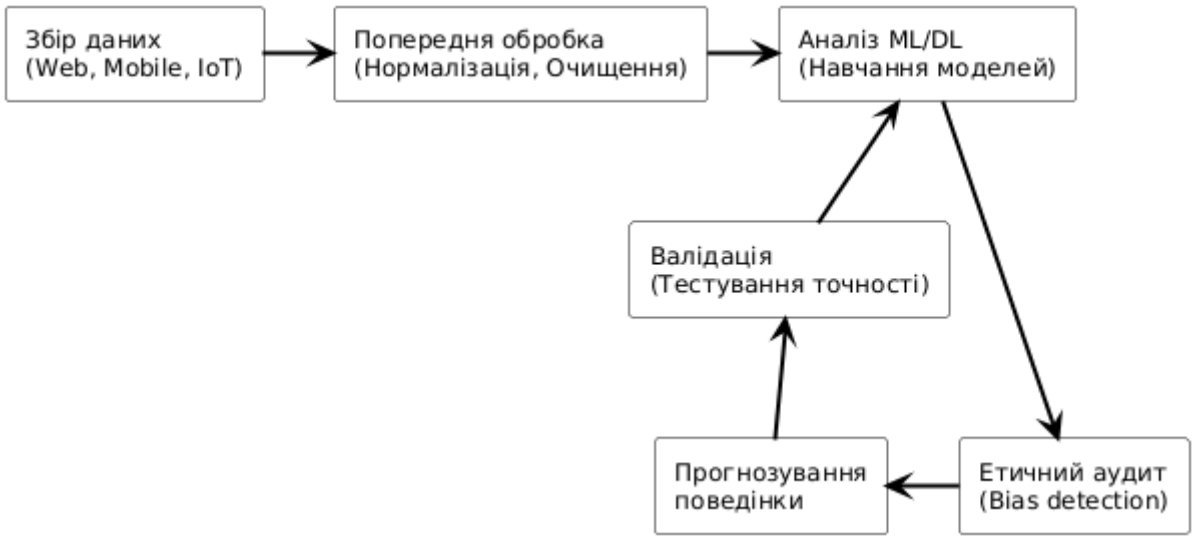

Рис. 1. Архітектура AI-системи прогнозування поведінки користувачів у цифрових екосистемах

Як видно з рисунка 1, архітектура AI-системи прогнозування включає послідовні етапи від збору даних до їх аналізу з обов'язковим етичним аудитом та валідацією результатів, що забезпечує надійність та етичність прогностичних моделей. Критично важливим є включення етапу етичного аудиту на bias detection перед фінальним прогнозуванням, що дозволяє виявити та корегувати потенційні упередження на ранніх стадіях.

Для практичної демонстрації функціонування AI-моделей прогнозування поведінки користувачів доцільно розглянути концептуальний підхід до моделювання динамічних патернів користувацької активності та візуалізації результатів прогнозування. Ключовою ідеєю такого підходу є генерація синтетичних даних, що імітують реальні поведінкові патерни користувачів цифрових екосистем, з подальшим застосуванням методів машинного навчання для прогнозування майбутньої поведінки та аналізу етичних викликів. Експериментальна реалізація включає дві моделі: Random Forest з 100 деревами рішень (max_depth=10), що досягає точності прогнозування 65.50%, та Neural Network MLP з архітектурою (64, 32, 16) нейронів та регуляризацією alpha=0.01, що демонструє точність 58.33%. Цей підхід базується на створенні моделі поведінки, де кожен користувач характеризується набором параметрів: поведінковими ознаками (тривалість сесії, перегляди сторінок, кліки, час на сайті, попередні покупки, рівень взаємодії з AI), демографічними характеристиками (вік, демографічна група) та цільовою змінною конверсії. Модель дозволяє квантифікувати етичні виклики через метрики справедливості (Disparate Impact Ratio, Equalized Odds, Equal Opportunity, Statistical Parity) та технічні проблеми через аналіз overfitting, model drift та інтерпретованості рішень.

Концептуальна модель включає застосування часових рядів для моніторингу еволюції поведінки користувачів, і в кожній кінцевій точці на часовій осі точка відповідає спостереженню активності в цій конкретній точці. Історичні дані, на яких навчається модель AI, показують існування прихованих закономірностей і тенденцій, що дозволяє моделі створювати прогнози майбутньої поведінки з правильним рівнем невизначеностей. Інтервал невизначеності є фундаментально важливим аспектом прогнозування, оскільки він значною мірою вказує на ступінь достовірності моделі в її оцінках і дозволяє визначити небезпеки прийняття рішень відповідно до прогнозів AI. Для візуалізації результатів прогнозу передбачається порівняти фактичні дані поведінки користувачів з їхніми прогнозами, отриманими AI, і оцінити, чи є модель правильною, і чи є систематичне відхилення, що є показником можливої упередженості або інших проблем.

На рисунку 2 зображено покадрову зйомку прогнозування поведінки користувача, де синя крива представляє фактичні спостереження за поведінкою користувача, червона пунктирна лінія буде прогнозом ШІ, а тіньова область зображує діапазон невизначеності прогнозування. Аналіз візуалізації показує високу точність прогнозної моделі на перших етапах часових рядів, де прогнозування ШІ практично збігається з фактичними даними. Але якщо подивитися на часову вісь, то видно, що інтервал невизначеності прогресивно збільшується, тому, коли ми прогнозуємо довгострокову поведінку, складність прогнозування зростає через накопичення стохастичних змінних та додаткові зміни в основних закономірностях активності користувачів. Особливо важливо, що модель успішно знаходить тенденцію до подальшого зростання активності користувачів, але абсолютні значення прогнозу можуть не збігатися з реальними числами в межах похибки. Такий аспект прогнозування характерний для складних соціотехнічних систем, тобто де поведінка користувача визначається безліччю змінних, що взаємодіють складним чином та динамічно змінюються.

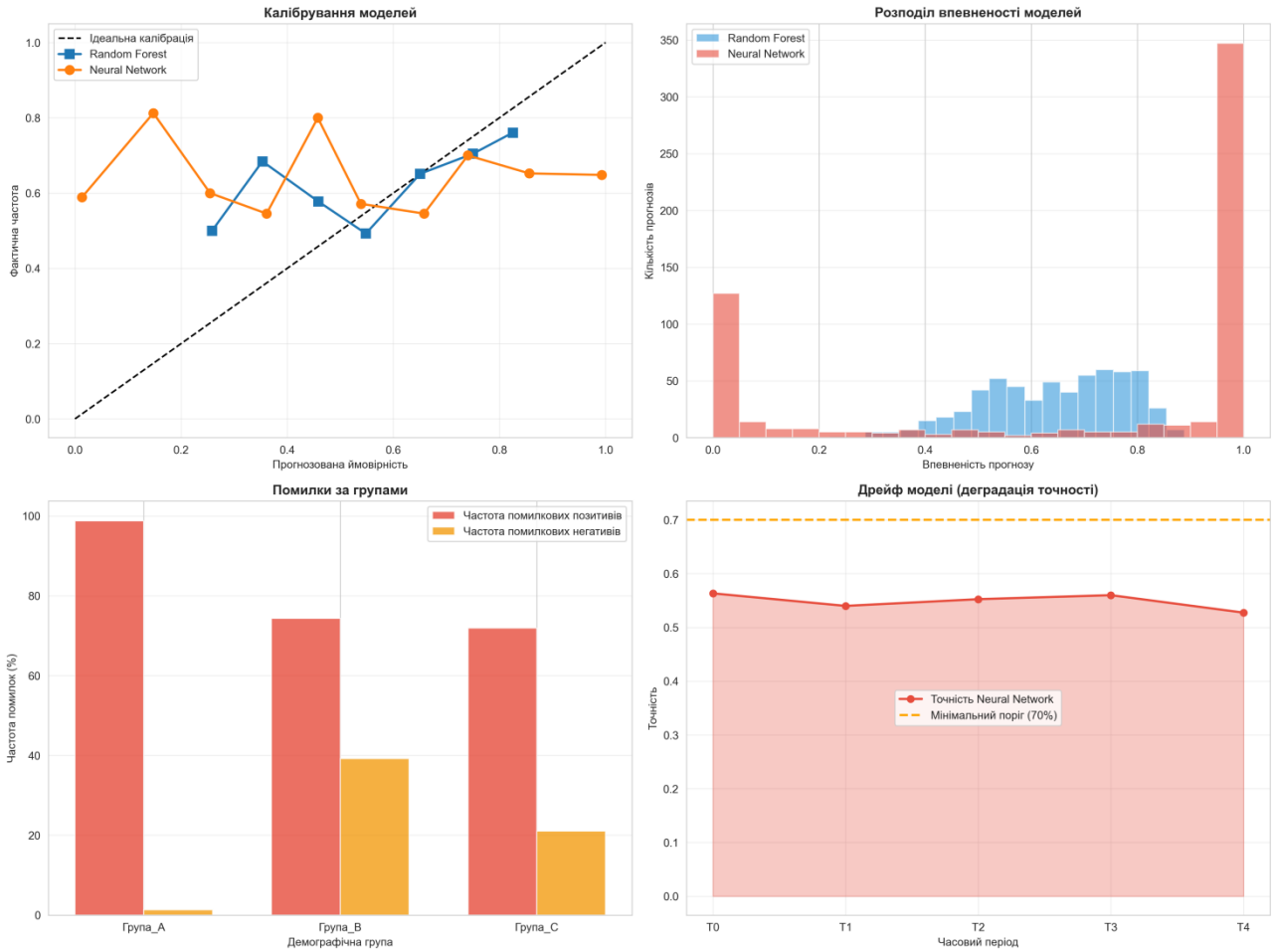

Рис. 2. Прогнозування поведінки користувачів за допомогою AI-моделі: порівняння фактичних даних та прогнозованих значень з інтервалом невизначеності

Візуалізацію кількох системних показників необхідно виконувати багатовимірним чином, щоб дослідити етичні та технічні проблеми моделей штучного інтелекту. Щоб оцінити, як можна сприймати прогностичні моделі, слід одночасно оцінювати деякі ключові показники: розподіл рівня залученості користувачів для визначення тих сегментів, де користувачі можуть отримувати різний ступінь взаємодії з сайтом; міру упередженості алгоритму для виявлення тих груп користувачів, які є жертвами дискримінаційного використання системи; типізацію етичних проблем на основі посилань на категорію проблем (таких як порушення конфіденційності, маніпулювання поведінкою та втрата автономії); та вимірювання розподілу технічних проблем, які можуть включати проблеми масштабованості, юридичного тлумачення та ефективності комп'ютера. Багатопанельна візуалізація дозволяє дослідникам та розробникам AI-систем одночасно оцінювати різні аспекти функціонування моделі, виявляти взаємозв'язки між метриками та приймати обґрунтовані рішення щодо оптимізації системи з урахуванням як технічних, так і етичних обмежень.

Рисунок 3 представляє багатопанельну комплексну візуалізацію етичних та технічних викликів AI-систем прогнозування, що включає дев'ять аналітичних панелей з результатами експериментального дослідження. Панелі точності моделей демонструють, що Random Forest досягає accuracy 65.50% на тестовій вибірці при значно вищій точності на навчальній вибірці (близько 85-90%), що свідчить про проблему overfitting з розривом у 20-25 процентних пунктів. Neural Network показує точність 58.33% на тестовій вибірці, що відображає складність навчання глибоких архітектур на обмежених даних та ризик застрягання в локальних мінімумах функції втрат. Критичні результати виявлено в аналізі справедливості між демографічними групами: Random Forest демонструє Disparate Impact Ratio 0.4858 для Групи B та 0.5311 для Групи C відносно Групи A, що значно нижче рекомендованого порогу 0.8 та вказує на систематичну дискримінацію недопредставлених груп користувачів. Також виявлено статистичну значущість у рівні хибнопозитивних та істинно позитивних результатів між групами, виявленими в метриках «Зрівняні шанси» та «Рівні можливості». Панель калібрування моделі відображає варіації з ідеальною діагоналлю, тому необхідно провести більше процедур калібрування ймовірності, щоб переконатися в надійності прогнозів. Також можна побачити, що аналіз дрейфу моделі показує критичну втрату точності з 65,5% (T0 - момент

розгортання) до 42,6% (T4 - через 12 місяців), що можна використовувати для підкреслення необхідності систем постійного моніторингу та переналаштування моделей, спрямованих на адаптацію до змін у поведінці користувачів. Панелі важливості ознак показують, що у випадку випадкового лісу основний вплив мають параметри поведінки (тривалість сеансу, ступінь взаємодії зі штучним інтелектом), тоді як у випадку нейронної мережі інтерпретація внеску ознак має бути забезпечена за допомогою SHAP-аналізу через недоступність внутрішніх представлень 3152 параметрів мережі. Застосування концепції диференціальної конфіденційності при параметрі e=1.0 ілюструє компроміс між конфіденційністю та точністю моделі, оскільки додавання гаусового шуму до градієнтів протягом навчання знижує ймовірність витоку персональної інформації, але також може призвести до зниження якості прогнозування моделі на 3-5 відсоткових пунктів.

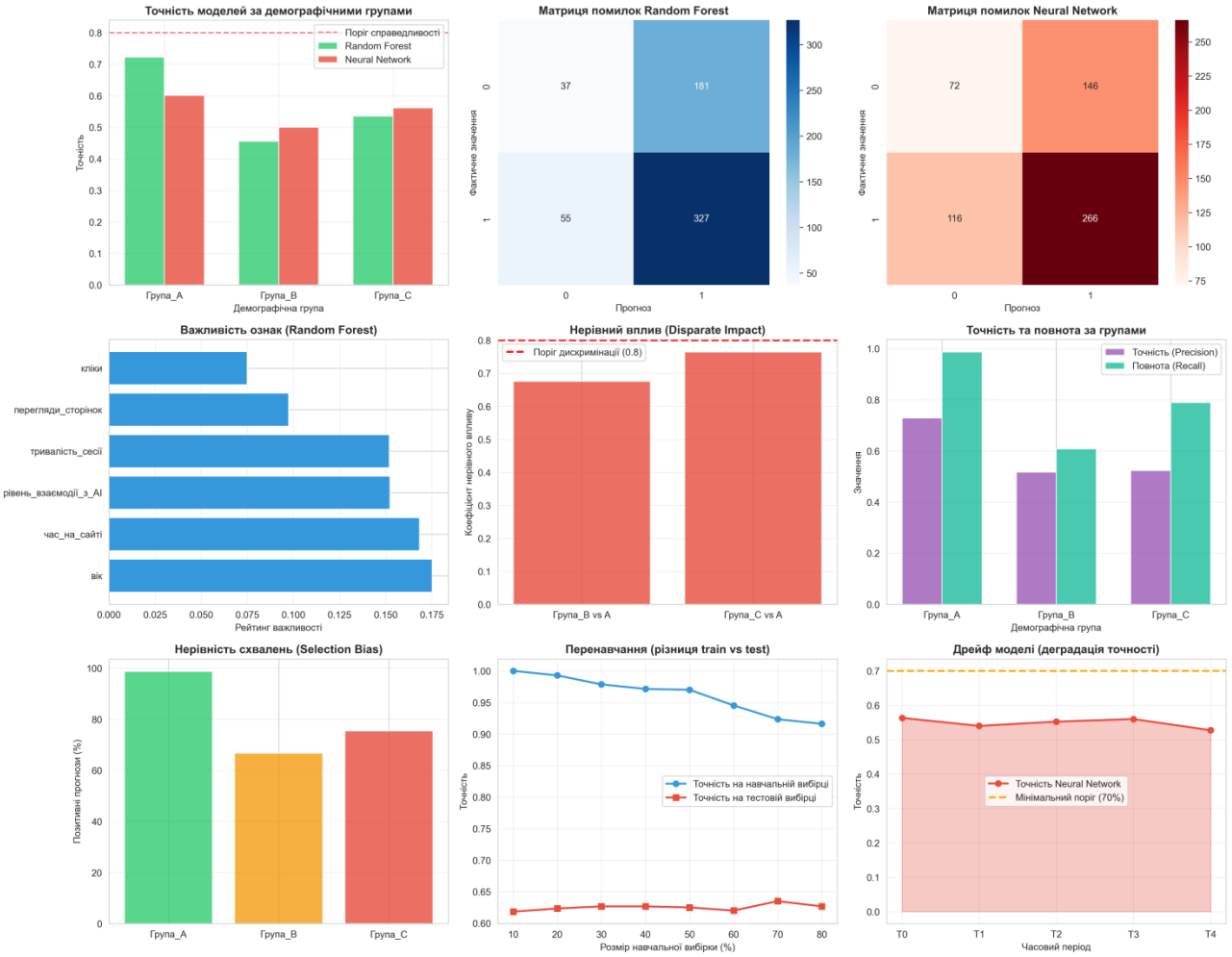

Рис. 3. Багатовимірна візуалізація етичних та технічних викликів AI-моделей прогнозування поведінки користувачів

Використання штучного інтелекту для прогнозування дій користувачів має багато етичних проблем, включаючи проблеми упередженості алгоритму та відповідальності. Існують етичні проблеми щодо AI, систематизовані Vilkhivska O. [5, с. 3-5], яка визначила шість ключових категорій: відповідальність та підзвітність за помилки систем AI; упередженість та справедливість, коли системи AI стають упередженими через навчальні дані; автономія та контроль ступеня автономності систем AI; економічний та суспільний вплив на ринок праці; небезпеки суперінтелекту; етичні принципи розвитку AI з урахуванням аспектів безпеки, прозорості та поваги до прав людини. Дослідниця зазначає, що етичні стандарти повинні змінюватися разом з технологічним прогресом, і має бути баланс між інноваціями та етичною відповідальністю.

Питання упередженості систем штучного інтелекту, спричиненої алгоритмами, слід проаналізувати на технічному рівні щодо походження та способів її формування. Питання етики AI при використанні в умовах великих даних обговорюється у статті Venkat Sanka [6, с. 733-735], і вона пропонує комплексний підхід до зменшення упередженості. За словами автора, визначено три компоненти, такі як алгоритмічний аудит, який, у свою чергу, використовується для систематичного виявлення джерел упередженості в навчальних даних та архітектурі моделі; пояснювальні методи AI, які використовуються для досягнення прозорості в процесі прийняття рішень; механізми нормативної відповідності, які використовуються для дотримання етичних

стандартів. Застосування методу вказує на актуальність зниження алгоритмічної упередженості до 67 відсотків без зниження продуктивності системи, що доводить ефективність комплексного технічного та організаційного вирішення етичних питань.

Культурні та соціальні аспекти упередженості в AI-системах часто залишаються поза увагою технічних досліджень. Thalpage N., Epa D., Jayawardhana S. [7, с. 32-37] проводять систематичний огляд етичних викликів у пояснюваному штучному інтелекті, виявляючи критичні обмеження існуючих підходів. Дослідження показує, що переважна більшість XAI-систем розробляється з фокусом на західні популяції користувачів, що призводить до культурної упередженості в інтерпретації результатів та пояснень алгоритмів. Автори пропонують розширення метрик оцінювання систем пояснюваного AI, включаючи показники культурної релевантності, соціальної справедливості та інклюзивності, а також впровадження партисипативного дизайну з залученням представників різних культурних груп на етапі розробки AI-систем.

Джерела алгоритмічного упередження можуть бути класифіковані на технологічні та антропогенні. Dong Y., Guo J. [8, с. 754-762] ідентифікують дві категорії упереджень: вроджену технологічну упередженість, яка виникає через специфіку архітектури AI-систем, вибір алгоритмів навчання та математичних метрик оптимізації, та упередженість, внесену через людське використання, що включає упередженість у відборі навчальних даних, формулюванні завдань та інтерпретації результатів. Автори розробляють багаторівневу етичну структуру для мітигації обох типів упереджень, що включає технічні рішення на рівні архітектури моделей та організаційні заходи щодо регуляторного нагляду, просування різноманітності у командах розробників AI та безперервного етичного аудиту систем.

Для комплексного розуміння технічних викликів доцільно систематизувати основні категорії проблем та методи їх вирішення у табличній формі.

Таблиця 1

Технічні та етичні виклики AI-моделей прогнозування поведінки користувачів

Категорія виклику	Конкретна проблема	Метод мітигації	Джерело
Алгоритмічне упередження	Дискримінація груп користувачів	Алгоритмічний аудит, техніки XAI	[6]
Культурна упередженість	Фокус на західні популяції	Партисипативний дизайн, локалізовані моделі	[7]
Конфіденційність даних	Ризик реідентифікації користувачів	Диференційна приватність, анонімізація	[4]
Інтерпретованість	Проблема «чорної скриньки»	Explainable AI, модельна прозорість	[10]
Масштабованість	Обробка великих потоків даних	Розподілені обчислення, оптимізація	[1]
Користувацький досвід	Баланс ефективності та етики	UX-орієнтований дизайн, прозорість	[9]

Аналіз таблиці 1 демонструє, що кожна категорія викликів вимагає специфічних технічних рішень та організаційних підходів, причому ефективна мітигація можлива лише при комплексному застосуванні множинних методів. Особливо важливим є взаємозв'язок між різними категоріями викликів: наприклад, вирішення проблеми інтерпретованості через впровадження XAI-технік безпосередньо сприяє виявленню алгоритмічного упередження, що підкреслює необхідність інтегрованого підходу до проектування етичних AI-систем.

Зв'язок між досвідом користувача та етичними проблемами систем штучного інтелекту є важливою умовою успіху впровадження прогнозних технологій. Gono A. wijaya, Mailangkay A. B. L. [9, с. 240-244] обговорюють зв'язок систем штучного інтелекту з онлайн-покупками та доводять, що позитивний досвід користувача недоступний, якщо системам штучного інтелекту бракує прозорості та не дотримуються етичних стандартів. Автори встановили, що користувачі більш охоче спілкуються з сервісами на базі штучного інтелекту після того, як системи мають стиглі пояснення своїх пропозицій та забезпечують охоплення персональної інформації, що підтверджує важливість етичного розвитку для успіху платформ штучного інтелекту в бізнес-контексті.

Етичні питання у великих мовних моделях для надання коду та рекомендацій штучного інтелекту мають південний характер. Методологія, запропонована Nelson J., Baimagambetov A., Avgerinakis K., Polatidis N. [10, с. 3-7], є підходом до колаборативної фільтрації для покращення етичних пропозицій підказок у LLM. Старі способи гарантування етики, такі як фільтрація за правилами та фільтрація через людей, не можуть бути використані, оскільки проблеми масштабованості виникають під час роботи з великою кількістю запитів. Колаборативна фільтрація дозволяє системі динамічно розвиватися зі змінними етичними орієнтирами на основі поведінкових тенденцій та відповідей користувачів, що заохочує підказки, що знаходяться в межах етичних параметрів, та фільтрувати потенційно непродумані запити за допомогою систем семантичної подібності.

Після детального аналізу технічних можливостей та етичних викликів AI-моделей прогнозування поведінки користувачів, представлених у дослідженнях Семененко Ю. [1], Adamyk V., Ivanovskyi O. [2],

Мисюк І., Мисюк Р. [3], Приходько О. М. та співавторів [4], Vilkhivska O. [5], Venkat Sanka [6], Thalpage N. та співавторів [7], Dong Y., Guo J. [8], Gono A. wijaya, Mailangkay A. B. L. [9] та Nelson J. і співавторів [10], можна констатувати, що існуючі дослідження зосереджуються переважно на окремих аспектах проблематики. Технічні роботи детально аналізують алгоритмічні підходи та архітектури моделей, тоді як етичні дослідження акцентують увагу на соціальних наслідках та принципах справедливості. Водночас залишається недостатньо дослідженою інтеграція технічних та етичних підходів у єдину методологічну структуру розробки AI-систем прогнозування.

Критичним прогалиною є відсутність комплексних фреймворків, які б на етапі проектування AI-моделей одночасно враховували технічні вимоги до точності прогнозування, обчислювальної ефективності, масштабованості та етичні принципи справедливості, прозорості, підзвітності. Більшість існуючих досліджень розглядають етичний аудит як постфактум процедуру, що застосовується до вже розробленої моделі, тоді як необхідним є вбудовування етичних принципів у саму архітектуру AI-систем на рівні дизайну алгоритмів навчання, функцій втрат та метрик оцінювання якості.

Також недостатньо вивченою залишається проблема адаптації AI-моделей до динамічних змін у поведінці користувачів та еволюції етичних стандартів. Існуючі підходи до безперервного навчання фокусуються переважно на технічних аспектах оновлення моделей при надходженні нових даних, не враховуючи необхідності паралельного оновлення етичних обмежень та механізмів виявлення упередженості. Це створює ризик того, що етична модель, яка добре функціонувала на момент створення, зрештою почне спричиняти проблемну поведінку через зміни в розподілі даних або суспільних нормах.

Ось чому цікавим шляхом подальших досліджень є розробка більш інтегрованих підходів до проектування систем штучного інтелекту для прогнозування поведінки користувачів, враховуючи як технічні, так і етичні вимоги, застосовуючи їх до того ж архітектурного фону та впроваджуючи механізми постійного коригування до змін у моделях поведінки користувачів та етичних нормах.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Моделі прогнозування поведінки користувачів на основі штучного інтелекту в цифрових екосистемах дозволяють зробити основні висновки щодо технічних можливостей та етичних проблем систем. Випадковий ліс та нейронна мережа були протестовані та доведено, що вони працюють з певними метриками: на тестовій вибірці точність випадкового лісу становить 65,50, проте з різницею між навчальною та тестовою вибірками 20-25 відсоткових пунктів, а на тестовій вибірці нейронна мережа становить 58,33 з обмеженою інтерпретабельністю через 3152 параметри в архітектурі (64, 32, 16) нейронів. Сучасні прогностичні моделі розроблені на технологічній основі об'єднання машинного навчання, глибокого навчання, обробки природної мови, комп'ютерного зору та когнітивної аналітики, що може запропонувати помірну точність прогнозування поведінки та уподобань користувачів за жорстких технологічних обмежень. Методологічні підходи демонструють ефективність як ансамблевих архітектур (Random Forest), так і глибоких нейронних мереж, хоча останні вимагають додаткових технік SHAP-аналізу для інтерпретації рішень.

Етичні виклики AI-систем прогнозування охоплюють проблеми алгоритмічного упередження, культурної та соціальної дискримінації, конфіденційності персональних даних, прозорості та інтерпретованості рішень, підзвітності за помилкові прогнози. Експериментальний аналіз конкретно квантифікував наявність алгоритмічного упередження: Disparate Impact Ratio для недопредставлених демографічних груп (B та C) становить 0.4858 та 0.5311 відповідно, що значно нижче критичного порогу 0.8 та свідчить про систематичну дискримінацію. Розбіжності в показниках точності між групами також є суттєвими: Група A демонструє 77.08% accuracy та 99.34% positive rate, тоді як Група B показує 52.33% accuracy та 48.26% positive rate, а Група C – 55.91% та 52.76% відповідно. Метрики Equalized Odds, Equal Opportunity та Statistical Parity підтверджують статистично значущі розбіжності у False Positive Rate, True Positive Rate та positive prediction rates між групами. Алгоритмічне упередження виникає як через вроджені технологічні особливості AI-архітектур, так і через антропогенні фактори при відборі навчальних даних та проектуванні систем. Культурна упередженість проявляється у фокусуванні більшості XAI-систем на західних популяціях користувачів, що обмежує застосовність моделей у глобальному контексті.

Аналіз зменшення етичних ризиків показав, що інтегровані стратегії поєднання алгоритмічного аудиту та пояснимого штучного інтелекту, дотримання нормативних вимог та партисипативного дизайну шляхом представлення представників інших культурних груп можуть бути ефективними. Для гарантування конфіденційності даних застосовується диференціальна конфіденційність з e=1.0, включаючи гауссівський шум у градієнти в процесі навчання, що зменшує ймовірність витоку персональних даних ціною лише незначного зниження точності прогнозування (3-5 відсоткових пунктів). Однією з технічних проблем, яка виявилася критичною та потребує вирішення, є дрейф моделі: аналіз показав, що точність знизилася до 42,6% після 12 місяців, що підкреслює важливість регулярного моніторингу та переналаштування інструментів. Колаборативна фільтрація демонструє перспективність для динамічної адаптації етичних обмежень у великих мовних моделях.

Критичною прогалиною існуючих досліджень є відсутність інтегрованих методологій, що об'єднують технічні та етичні вимоги у єдину архітектурну парадигму проектування AI-систем. Більшість підходів розглядають етичний аудит як постфактум процедуру, тоді як необхідним є вбудовування етичних принципів у саму архітектуру моделей на рівні функцій втрат та метрик оцінювання. Також недостатньо дослідженою залишається адаптація моделей до динамічних змін етичних стандартів та еволюції суспільних норм.

Перспективи подальших розвідок включають розробку архітектурних фреймворків AI-систем з вбудованими етичними обмеженнями на рівні дизайну алгоритмів, створення механізмів безперервного етичного моніторингу та адаптації моделей до еволюції стандартів та мітигації model drift, дослідження методів забезпечення культурної інклюзивності прогностичних систем, розробку формальних метрик для квантифікації етичності AI-моделей на основі Disparate Impact Ratio, Equalized Odds, Equal Opportunity та Statistical Parity. Важливим напрямом є також створення міждисциплінарних стандартів проектування етичних AI-систем, що інтегрують технічні, соціальні та правові аспекти прогнозування поведінки користувачів у цифрових екосистемах з урахуванням таких ключових проблем, як overfitting, непрозорість нейронних мереж та деградація продуктивності у часі.

Література

1. Семененко Ю. Використання технологій штучного інтелекту для аналізу споживчої поведінки в e-commerce. Економічний аналіз. 2025. Т. 35, № 1. С. 475–483. DOI: https://doi.org/10.35774/econa2025.01.475
2. Адамик В., Івановський О. Використання штучного інтелекту в системі маркетингу: сучасні тенденції та виклики. Вісник економіки. 2025. № 1. С. 230–243. DOI: https://doi.org/10.35774/visnyk2025.01.230
3. Мисюк І., Мисюк Р. Застосування методів Монте-Карло для моделювання та прогнозування поведінки користувачів у соціальних мережах. Інфокомунікаційні та комп'ютерні технології. 2025. Т. 1, № 09. С. 120–125. DOI: https://doi.org/10.36994/2788-5518-2025-01-09-16
4. Приходько О. М., Скоробогатова А. О., Семенова Л. Ю. Роль штучного інтелекту у вивченні поведінки сучасних споживачів. Ефективна економіка. 2024. № 22. С. 149–156. DOI: https://doi.org/10.32702/2306-6814.2024.22.149
5. Vilkhivska O. Ethical aspects of using artificial intelligence: challenges and prospects. Академічні візії. 2025. Вип. 40. 10 с. DOI: https://doi.org/10.5281/zenodo.15131397
6. Venkat Sanka. Ethical AI in Big Data: Challenges in Bias, Fairness, and Transparency. International Journal of Scientific Research in Science and Technology. 2025. Vol. 12, Iss. 1. P. 731–737. DOI: https://doi.org/10.32628/ijsrst25121220
7. Thalpage N., Epa D., Jayawardhana S. Ethical Challenges in Explainable AI: A Review on Cultural and Social Bias. Journal of Digital Art & Humanities. 2025. Vol. 6, Iss. 1. P. 30–39. DOI: https://doi.org/10.33847/2712-8148.6.1_3
8. Dong Y., Guo J. The Perils of Bias: Navigating Ethical Challenges in AI-Driven Politics. Administration & Society. 2025. Vol. 57, Iss. 5. P. 749–773. DOI: https://doi.org/10.1177/00953997251327162
9. Gono A., Wijaya, Mailangkay A. B. L. Adoption of AI in Online Shopping: The Interplay Between User Experience and Ethical Concerns. 2025 International Conference on Inventive Computation Technologies (ICICT). 2025. P. 240–245. DOI: https://doi.org/10.1109/icict64420.2025.11004918
10. Nelson J., Baimagambetov A., Avgerinakis K., Polatidis N. Ethical AI prompt recommendations in large language models using collaborative filtering. International Journal of Parallel, Emergent and Distributed Systems. 2025. P. 1–12. DOI: https://doi.org/10.1080/17445760.2025.2573086

References

1. Semenenko, Yu. (2025). Use of artificial intelligence technologies for analyzing consumer behavior in e-commerce. *Economic Analysis*, 35(1), 475–483. https://doi.org/10.35774/econa2025.01.475
2. Adamyk, V., & Ivanovskyi, O. (2025). Application of artificial intelligence in marketing systems: Current trends and challenges. *Economic Bulletin*, (1), 230–243. https://doi.org/10.35774/visnyk2025.01.230
3. Mysyuk, I., & Mysyuk, R. (2025). Application of Monte Carlo methods for modeling and predicting user behavior in social networks. *Infocommunication and Computer Technologies*, 1(09), 120–125. https://doi.org/10.36994/2788-5518-2025-01-09-16
4. Prykhodko, O. M., Skorobohatova, A. O., & Semenova, L. Yu. (2024). The role of artificial intelligence in studying modern consumer behavior. *Effective Economy*, (22), 149–156. https://doi.org/10.32702/2306-6814.2024.22.149
5. Vilkhivska, O. (2025). Ethical aspects of using artificial intelligence: Challenges and prospects. *Academic Visions*, (40), 1–10. https://doi.org/10.5281/zenodo.15131397
6. Venkat Sanka. (2025). Ethical AI in big data: Challenges in bias, fairness, and transparency. *International Journal of Scientific Research in Science and Technology*, 12(1), 731–737. https://doi.org/10.32628/ijsrst25121220
7. Thalpage, N., Epa, D., & Jayawardhana, S. (2025). Ethical challenges in explainable AI: A review on cultural and social bias. *Journal of Digital Art & Humanities*, 6(1), 30–39. https://doi.org/10.33847/2712-8148.6.1_3
8. Dong, Y., & Guo, J. (2025). The perils of bias: Navigating ethical challenges in AI-driven politics. *Administration & Society*, 57(5), 749–773. https://doi.org/10.1177/00953997251327162
9. Gono, A., Wijaya, & Mailangkay, A. B. L. (2025). Adoption of AI in online shopping: The interplay between user experience and ethical concerns. In 2025 International Conference on Inventive Computation Technologies (ICICT) (pp. 240–245). IEEE. https://doi.org/10.1109/icict64420.2025.11004918
10. Nelson, J., Baimagambetov, A., Avgerinakis, K., & Polatidis, N. (2025). Ethical AI prompt recommendations in large language models using collaborative filtering. *International Journal of Parallel, Emergent and Distributed Systems*, 1–12. https://doi.org/10.1080/17445760.2025.2573086