https://doi.org/10.31891/2307-5732-2026-363-11
УДК 004.8:340.134

СТИСЛО ТАРАС
ЗВО «Університет Короля Данила»
https://orcid.org/0000-0002-2377-7985
e-mail: taras.styslo@ukd.edu.ua

АЛГОРИТМІЧНА ВІДПОВІДАЛЬНІСТЬ: ТЕХНІЧНІ МОДЕЛІ ПОЯСНЮВАНОГО ШТУЧНОГО ІНТЕЛЕКТУ ТА ПРАВОВІ ВИКЛИКИ ЇХ ВПРОВАДЖЕННЯ

У статті розглянуто проблему алгоритмічної відповідальності в контексті розвитку технологій штучного інтелекту (ШІ) та потреби забезпечення прозорості процесів прийняття рішень. Зростання кількості автоматизованих систем у публічному управлінні, фінансовому секторі, судочинстві та охороні здоров'я актуалізує питання пояснюваності рішень, прийнятих ШІ. Науковий інтерес зосереджено на технічних моделях пояснюваного штучного інтелекту (Explainable Artificial Intelligence, XAI) - таких як SHAP, LIME, Grad-CAM - які дають змогу інтерпретувати роботу складних алгоритмів глибинного навчання. Проаналізовано міжнародні нормативні ініціативи (EU AI Act, OECD AI Principles, ISO/IEC 42001:2023) та українські підходи до регулювання використання ШІ. Показано взаємозв'язок між технічними аспектами прозорості моделей і юридичними вимогами щодо відповідальності за помилки, упередженість та шкоду, спричинену алгоритмічними рішеннями. Зроблено висновок, що ефективне впровадження XAI потребує синергії між технічними стандартами, етичними принципами та правовим регулюванням. Перспективним напрямом подальших досліджень є розробка національної методології оцінювання алгоритмічної відповідальності та її інтеграція в законодавство України у сфері цифрової етики й штучного інтелекту.

Ключові слова: *штучний інтелект; пояснюваний штучний інтелект; алгоритмічна відповідальність; прозорість алгоритмів; XAI; юридична підзвітність; етичні принципи ШІ; EU AI Act; ISO/IEC 42001:2023; аудит моделей; управління ризиками ШІ; правове регулювання ШІ.*

STYSLO TARAS
King Danylo University

ALGORITHMIC ACCOUNTABILITY: TECHNICAL MODELS OF EXPLAINABLE ARTIFICIAL INTELLIGENCE AND LEGAL CHALLENGES OF THEIR IMPLEMENTATION

The article examines the issue of algorithmic accountability in the context of the growing role of Artificial Intelligence (AI) and the urgent need for transparency in decision-making processes. The proliferation of automated systems in public administration, finance, justice, and healthcare highlights the importance of explainability in AI-based decisions, especially in cases where such decisions directly affect human rights, freedoms, and legal responsibilities. The research focuses on technical models of Explainable Artificial Intelligence (XAI), including SHAP, LIME, and Grad-CAM, which enable the interpretation of complex deep learning algorithms and increase trust in AI-driven systems.

International regulatory initiatives such as the EU AI Act, OECD AI Principles, and ISO/IEC 42001:2023 are analyzed alongside Ukrainian approaches to AI governance, with particular attention to issues of compliance, risk management, and accountability mechanisms. The study reveals the interdependence between technical transparency and legal accountability for algorithmic bias, errors, and potential harm caused by automated decision-making systems. Special emphasis is placed on ethical challenges related to fairness, non-discrimination, and human oversight in high-risk AI applications.

It is concluded that the effective implementation of XAI requires synergy between technical standards, ethical principles, and legal regulation. A promising direction for further research involves the development of a national methodology for assessing algorithmic responsibility, the establishment of standardized audit procedures for AI systems, and their integration into Ukrainian legislation on digital ethics and artificial intelligence. The results of the study may be useful for researchers, policymakers, legal experts, and developers involved in the design and regulation of trustworthy AI systems.

Keywords: *artificial intelligence; explainable artificial intelligence; algorithmic accountability; algorithmic transparency; XAI; legal responsibility; AI ethics; EU AI Act; ISO/IEC 42001:2023; model audit; AI risk management; legal regulation of AI.*

Стаття надійшла до редакції / Received 16.01.2026
Прийнята до друку / Accepted 11.02.2026
Опубліковано / Published 26.03.2026

Постановка проблеми

Стрімкий розвиток систем штучного інтелекту (ШІ) упродовж останнього десятиліття кардинально змінив підходи до прийняття рішень у бізнесі, державному управлінні, медицині, транспорті та правовій сфері. Моделі машинного й глибинного навчання здатні здійснювати аналіз великих масивів даних, виявляти закономірності, прогнозувати події й формувати рекомендації з точністю, що перевищує людські можливості. Водночас ці системи приймають рішення у спосіб, який часто неможливо повністю інтерпретувати або пояснити людині - так званий феномен «чорної скриньки».

В умовах цифрової трансформації та автоматизації державних сервісів постає питання алгоритмічної відповідальності - визначення суб'єкта, який несе юридичну відповідальність за наслідки дії автономних систем. Прозорість і пояснюваність моделей штучного інтелекту стають не лише технічним, а й правовим завданням. Без забезпечення зрозумілості алгоритмів неможливо гарантувати ні дотримання прав людини, ні належне судове чи адміністративне оскарження рішень, прийнятих автоматизовано. Отже, проблема має міждисциплінарний характер і потребує поєднання технічних підходів до Explainable AI (XAI) із нормативними принципами відповідальності.

Аналіз останніх досліджень

Питання алгоритмічної прозорості почали активно досліджувати після 2016 року - з появою стандартів етичного використання ШІ та рекомендацій OECD і IEEE [1;6]. Європейська комісія у 2019 році опублікувала «Ethics Guidelines for Trustworthy AI», що визначили принципи пояснюваності, справедливості та підзвітності. На глобальному рівні ISO/IEC у 2023 році ухвалили стандарт ISO/IEC 42001:2023 «AI Management System», який окреслює вимоги до системного управління ШІ у відповідності до принципів ризик-орієнтованого підходу [4;5].

Серед технічних досліджень найбільш поширеними є моделі LIME (Local Interpretable Model-Agnostic Explanations), SHAP (SHapley Additive exPlanations) та Grad-CAM (Gradient-weighted Class Activation Mapping). Вони забезпечують можливість візуалізувати й інтерпретувати рішення нейронних мереж, роблячи їх частково пояснюваними для експертів [7;8;9].

В Україні питання правового регулювання ШІ перебуває на етапі становлення. У 2023-2024 рр. Міністерство цифрової трансформації та Національний центр розвитку ШІ розробили концепцію регулювання ШІ на основі принципів ЄС. Однак відсутні національні стандарти оцінки алгоритмічної відповідальності, що створює ризики невідповідності майбутньому Регламенту ЄС [11] про штучний інтелект (EU AI Act) [3].

Невирішеними залишаються такі питання - чи може алгоритм набувати статусу суб'єкта права, хто саме повинен нести відповідальність у разі завдання збитків - розробник, користувач чи постачальник даних, а також яким чином технічно забезпечити прозорість рішень без втрати точності моделей [12;13].

Метою статті є науковий аналіз взаємозв'язку між технічними моделями пояснюваного штучного інтелекту (XAI) та правовими аспектами алгоритмічної відповідальності в контексті впровадження автоматизованих систем прийняття рішень [2;3;4;5].

Завдання дослідження полягають у комплексному вивченні взаємозв'язку технічних і правових аспектів забезпечення алгоритмічної відповідальності систем штучного інтелекту [1;2;3].

По-перше, передбачається здійснити поглиблений аналіз сучасних технічних підходів до реалізації принципу пояснюваності штучного інтелекту, зокрема моделей XAI, які сприяють підвищенню прозорості процесів прийняття рішень [7;8;9].

По-друге, необхідно з'ясувати правові наслідки відсутності прозорості алгоритмів у контексті європейського та національного законодавства, оцінити ризики для суб'єктів правовідносин і проблеми правозастосування [1;2;3;11].

По-третє, дослідження спрямоване на розроблення концептуальних напрямів інтеграції технічних та юридичних механізмів у єдину систему управління алгоритмічною відповідальністю, що забезпечить баланс між інноваційністю, безпекою та підзвітністю ШІ-рішень [3;4;5;12;13].

Виклад основного матеріалу

Поняття алгоритмічної відповідальності формується на перетині правових, технічних та етичних підходів до управління системами штучного інтелекту (ШІ), які здатні приймати автономні рішення на основі обробки великих масивів даних [1;2]. У науковій літературі воно трактується як комплексна властивість штучних інтелектуальних систем, що забезпечує їх підзвітність, контрольованість і відтворюваність результатів діяльності, незалежно від рівня автоматизації процесів [2;13].

З огляду на міждисциплінарний характер феномену, алгоритмічна відповідальність постає як комплексна категорія, що поєднує правові, етичні та технічні виміри функціонування систем штучного інтелекту [1;2;6;13]. Її зміст не може бути зведений виключно до правового регулювання або технічного контролю, адже ефективність застосування ШІ визначається саме гармонізацією цих трьох компонентів у єдиному концептуальному полі [2;13].

По-перше, юридичний компонент алгоритмічної відповідальності полягає у визначенні суб'єкта, на якого покладається обов'язок за наслідки дії інтелектуальної системи. Йдеться про встановлення чітких меж відповідальності між розробником, оператором, користувачем та власником даних, з урахуванням їхніх ролей у життєвому циклі ШІ. Вирішення цього питання має ключове значення для формування механізмів правозастосування, компенсації збитків та запобігання потенційним технологічним ризикам [3].

По-друге, етичний аспект алгоритмічної відповідальності пов'язаний із забезпеченням дотримання фундаментальних принципів справедливості, недискримінації, прозорості й людської гідності [1;2;6;13]. Саме ці принципи формують довіру суспільства до автоматизованих систем і є необхідною умовою їх прийняття у критично важливих сферах, таких як охорона здоров'я, фінанси, публічне управління та правосуддя. В етичному контексті підзвітність алгоритмів розглядається як складова цифрової етики, що передбачає відповідальне ставлення до проєктування, навчання та експлуатації моделей ШІ.

По-третє, технічний компонент алгоритмічної відповідальності орієнтований на розроблення й впровадження інженерних рішень, які забезпечують пояснюваність, верифікацію та аудит функціонування алгоритмів [4;5]. До таких рішень належать методи Explainable AI (XAI), процедури тестування на надійність, системи журналювання рішень та інструменти контролю якості даних [7;8;9]. Саме вони створюють технічне підґрунтя для здійснення правового контролю, дозволяючи документально підтвердити процеси прийняття рішень і оцінити їх відповідність встановленим нормам.

Таким чином, алгоритмічна відповідальність у сучасному науковому дискурсі розглядається як інтегративна система правових, етичних і технічних механізмів, спрямованих на забезпечення підзвітності та довіри до технологій штучного інтелекту в умовах цифрової трансформації суспільства [1;2;3;4;5;13].

Для реалізації алгоритмічної відповідальності у технічному вимірі ключовим виступає принцип Explainability [7;8;9;13], тобто здатність системи розкривати внутрішню логіку своїх рішень у зрозумілій людині формі. Саме цей принцип є основою для формування довіри до ШІ, можливості проведення експертної оцінки обґрунтованості рішень та юридичної кваліфікації наслідків їх застосування.

На практиці забезпечення алгоритмічної відповідальності вимагає комплексного підходу: від розроблення етичних кодексів і правових стандартів до впровадження технічних процедур документування процесів прийняття рішень [1;2;3;4;5;6]. У провідних країнах світу, зокрема в ЄС, США, Канаді та Японії, розробляються національні стратегії відповідального ШІ [1;2;6], які визначають баланс між інноваційністю технологій і безпекою суспільства.

Пояснюваний штучний інтелект (Explainable Artificial Intelligence, XAI) є базовим технічним інструментом реалізації принципу прозорості. Його мета полягає у тому, щоб результати, отримані за допомогою алгоритмів машинного або глибинного навчання, можна було зрозуміти, інтерпретувати та обґрунтувати [7;8;9].

Серед найпоширеніших методів XAI особливе значення мають LIME (Local Interpretable Model-Agnostic Explanations) та SHAP (SHapley Additive exPlanations). LIME створює локальні апроксимації складної моделі, аналізуючи, як зміна окремих вхідних параметрів впливає на кінцеве рішення. SHAP, у свою чергу, ґрунтується на теорії ігор Шеплі та дозволяє кількісно оцінити внесок кожної ознаки у прийняте рішення [7;8].

Іншим важливим підходом є Grad-CAM (Gradient-weighted Class Activation Mapping), який застосовується для глибинних згорткових нейронних мереж, переважно у системах комп'ютерного зору. Цей метод візуалізує ділянки зображення, що найбільше вплинули на класифікацію, і тим самим дозволяє людині зрозуміти, якими елементами інформації керувалася модель [9].

Попри значний науковий прогрес, жоден із цих методів не забезпечує повної прозорості або формальної доказовості. Вони створюють лише інтерпретаційні гіпотези, тобто ймовірні пояснення поведінки системи. Проте ці інструменти стали основою для розроблення XAI Dashboard-систем - інтерактивних платформ, що надають аудиторам, юристам і регуляторам змогу перевіряти обґрунтованість алгоритмічних висновків [7;8;9;13].

Реалізація XAI на практиці потребує впровадження модулів аудиту моделей (Model Auditing), журналів рішень (Decision Logs) та механізмів відтворюваності результатів (Reproducibility Tools) [4;5]. У поєднанні з методами контролю версій моделей (Model Versioning) вони створюють технічну базу для юридичної відповідальності за наслідки дії ШІ-систем. Таким чином, пояснюваність є не лише питанням інтерпретації, а й інструментом забезпечення правової підзвітності [3;4;5].

Найбільш суттєвим і водночас концептуально складним викликом у контексті формування системи алгоритмічної підзвітності є відсутність правосуб'єктності алгоритмічних систем. Алгоритм, як технічний конструкт, не може розглядатися як самостійний носій прав і обов'язків, а отже - не здатний бути суб'єктом юридичної відповідальності у традиційному розумінні. Це зумовлює необхідність перерозподілу відповідальності між різними учасниками життєвого циклу штучного інтелекту - розробниками, провайдерами, операторами, користувачами та власниками даних. Така багаторівнева модель відповідальності, хоча й відповідає сучасним принципам багатостороннього управління цифровими технологіями, водночас ускладнює встановлення причинно-наслідкових зв'язків між алгоритмічними рішеннями та їх правовими наслідками для фізичних і юридичних осіб [3;12].

У межах Регламенту Європейського Союзу про штучний інтелект (EU AI Act) запропоновано системне врегулювання цього питання шляхом запровадження обов'язку провайдерів ШІ здійснювати комплексну оцінку ризиків та забезпечувати трасованість (traceability) алгоритмічних рішень. Цей підхід передбачає обов'язкове документування усіх етапів життєвого циклу моделі: процесів навчання, використаних наборів даних, архітектури системи, параметрів алгоритму та результатів тестування на упередженість і точність. Таким чином, формується юридично значимий ланцюг відстежуваності, який забезпечує можливість ретроспективного аналізу рішень та покладання відповідальності на конкретного суб'єкта у випадку заподіяння шкоди [3].

В українському правовому полі питання регулювання штучного інтелекту перебуває на етапі становлення. Станом на 2025 рік не існує законодавчо закріплених вимог щодо обов'язкової пояснюваності алгоритмів, процедур сертифікації моделей або визначення юридичних механізмів притягнення до відповідальності у разі шкоди, спричиненої автономною дією системи. Відсутність таких положень створює регуляторну прогалину між фактичним використанням технологій ШІ у державному та приватному секторах і можливостями їх правового контролю. У результаті виникає дисбаланс між інноваційною активністю ІТ-галузі та рівнем правової захищеності користувачів і суспільства загалом [11].

Окремої уваги заслуговує галузь правосуддя, де застосування ШІ для підтримки прийняття судових рішень породжує ризики алгоритмічної упередженості (algorithmic bias). Згідно з дослідженнями Європейської комісії (2023), моделі машинного навчання можуть відтворювати соціальні або гендерні стереотипи, закладені у навчальних даних, що потенційно загрожує принципам справедливого судочинства. У цьому контексті актуальним є запровадження правової оцінки справедливості (fairness assessment) алгоритмічних рішень, а також створення механізмів апеляційного перегляду випадків, коли результати, згенеровані ШІ, впливають на права та обов'язки особи [10].

Загалом, забезпечення алгоритмічної підзвітності неможливе без налагодження сталої інституційної взаємодії між технічними фахівцями, правниками, етичними комітетами, розробниками стандартів і державними регуляторами. Лише комплексна міждисциплінарна модель управління дозволить узгодити технічні стандарти, правові норми та етичні принципи, забезпечивши ефективну реакцію правової системи на технологічні ризики і формування довіри до штучного інтелекту в умовах цифрової держави [1;2;3;4;5;6;13].

Для ефективного поєднання технічних, правових та етичних засад функціонування систем штучного інтелекту доцільним є впровадження комплексної рамкової системи алгоритмічної підзвітності (Algorithmic Accountability Framework). Така система покликана інтегрувати регуляторні вимоги, міжнародні технічні стандарти та організаційні процедури в єдину узгоджену модель управління ризиками, прозорістю та відповідальністю алгоритмічних рішень [12;13].

З науково-практичного погляду, Algorithmic Accountability Framework може розглядатися як багаторівнева архітектура, спрямована на забезпечення технологічної надійності, юридичної визначеності та соціальної довіри до систем штучного інтелекту. Її структурні елементи повинні охоплювати як технічні механізми контролю, так і нормативно-правові гарантії підзвітності суб'єктів, що беруть участь у створенні, експлуатації та моніторингу алгоритмів.

По-перше, важливим компонентом такої системи має стати реєстр алгоритмів високого ризику, який акумулюватиме інформацію про моделі, застосовані у критично важливих галузях - охороні здоров'я, транспортній логістиці, фінансових послугах, судочинстві та публічному управлінні. Наявність такого реєстру забезпечить прозорість і контрольованість впровадження ШІ-рішень, а також дозволить здійснювати державний нагляд за їх впливом на права громадян.

По-друге, доцільно запровадити механізм сертифікації моделей пояснюваного штучного інтелекту (XAI) на відповідність міжнародним стандартам управління ШІ, зокрема ISO/IEC 42001:2023 («Artificial Intelligence Management System») та ISO/IEC 23894:2023 («AI Risk Management»). Сертифікація сприятиме формуванню єдиних технічних критеріїв якості, безпеки та прозорості, а також слугуватиме інструментом довіри між розробниками, державними органами та користувачами технологій [4;5].

По-третє, ключовим елементом правового забезпечення має стати регламентоване ведення журналів рішень (Decision Logs), які фіксуватимуть усі етапи процесу прийняття рішень алгоритмом. Така документація є необхідною для ретроспективного аналізу роботи системи, верифікації її коректності та підтвердження обґрунтованості дій під час розслідування інцидентів чи спорів, пов'язаних із функціонуванням ШІ.

По-четверте, система має передбачати аудит справедливості (Fairness Audit) - регулярну перевірку моделей на предмет наявності дискримінаційних, соціально або гендерно упереджених патернів у навчальних даних і результатах роботи. Проведення таких аудитів відповідно до міжнародних рекомендацій OECD AI Principles дозволить мінімізувати ризики порушення прав людини й забезпечити відповідність діяльності алгоритмічних систем етичним стандартам демократичного суспільства [1].

По-п'яте, важливою складовою є етичний нагляд (Ethical Oversight), що передбачає створення незалежних комітетів із цифрової етики. Їхнім завданням є оцінка суспільних наслідків впровадження нових алгоритмів, розгляд потенційних конфліктів між технічними інноваціями та гуманітарними цінностями, а також вироблення рекомендацій щодо відповідального використання ШІ у публічному секторі.

Впровадження зазначеної системи дозволить сформувати єдину національну екосистему алгоритмічного управління, у межах якої технічна відповідальність розробників буде гармонійно поєднана з юридичною підзвітністю операторів і державних інституцій. Це сприятиме підвищенню рівня довіри до технологій штучного інтелекту, мінімізації ризиків зловживань, запобіганню дискримінації та забезпеченню балансу між інноваційним розвитком і захистом прав людини.

Реалізація подібної моделі в Україні має розглядатися як перспективний напрям державної політики у сфері цифрової трансформації та правового регулювання ШІ. Її впровадження дозволить гармонізувати національну нормативну базу з положеннями Європейського Союзу, стимулювати розвиток інституту цифрової етики, а також закласти основи відповідального та підзвітного використання технологій штучного інтелекту у публічному управлінні, бізнесі та суспільному житті [3;11;13].

Висновки

Проблематика алгоритмічної відповідальності належить до категорії комплексних і міждисциплінарних наукових питань, що охоплюють взаємодію технічних, правових, етичних і соціальних компонентів цифрової екосистеми. Вона постає не лише як питання технологічної довіри до систем штучного інтелекту (ШІ), але й як важливий чинник формування сучасної моделі правової держави у цифрову добу. З огляду на стрімке впровадження інтелектуальних технологій у публічне управління, фінанси, медицину, транспорт і правосуддя, постає потреба у цілісному підході до регулювання відповідальності за прийняті алгоритмами рішення, який поєднує інженерну точність, юридичну визначеність і етичну справедливість [1;2;213].

Проведений у межах дослідження аналіз засвідчує, що пояснюваність (explainability) є ключовою передумовою формування юридичної підзвітності штучного інтелекту. Без здатності системи обґрунтувати логіку своїх рішень неможливо забезпечити ні прозорість процесів прийняття рішень, ні ефективне правозастосування у випадках потенційних порушень прав людини. Тому розроблення методів пояснюваності

є не лише технічним, а й нормативно-правовим завданням, що має бути інтегроване у механізми державного управління та сертифікації ШІ-рішень.

Водночас, наявні технічні підходи - SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations) та Grad-CAM (Gradient-weighted Class Activation Mapping) - хоча й забезпечують часткову прозорість алгоритмів, не створюють умов для формування повної системи алгоритмічної відповідальності. Їхня інтерпретативна природа дозволяє лише наблизитися до розуміння процесів прийняття рішень, проте не гарантує можливості юридичного відтворення або доказовості у разі настання спірних наслідків. Тому ці рішення слід розглядати як базові технологічні інструменти, що мають бути доповнені нормативними вимогами до ведення журналів рішень, аудиту моделей і незалежної верифікації даних [7;8;9;3;4;5].

У контексті національного правового регулювання особливої актуальності набуває необхідність створення державного стандарту або дорожньої карти імплементації Explainable AI (XAI) у правову систему України. Такий документ має бути гармонізований із вимогами Регламенту ЄС про штучний інтелект (EU AI Act), а також із міжнародними стандартами ISO/IEC 42001:2023 (AI Management Systems) та ISO/IEC 23894:2023 (AI Risk Management) [3;4;5;11]. Запровадження єдиного нормативного підходу забезпечить узгодженість між технічними процесами, правовими нормами й етичними принципами, що регулюють застосування ШІ у публічному секторі та приватній сфері.

Подальші дослідження доцільно спрямувати на розроблення моделей оцінювання рівня алгоритмічної відповідальності, що дозволять кількісно вимірювати ступінь прозорості, надійності та підзвітності систем штучного інтелекту. Також перспективним напрямом є створення механізмів державного моніторингу алгоритмів високого ризику, здатних виявляти потенційні порушення прав людини, дискримінаційні патерни або технічні збої, які можуть мати соціально значущі наслідки. Важливою складовою цієї діяльності є сертифікація ШІ-систем, що повинна включати як технічний, так і правовий аудит відповідно до міжнародних стандартів управління ризиками [10;11;12].

Отже, забезпечення пояснюваності штучного інтелекту є не лише технічною проблемою, а фундаментальною умовою розвитку правової держави у цифрову епоху. Вона формує основу для довіри, підзвітності та легітимності алгоритмічних систем, виступаючи водночас запорукою балансу між технологічними інноваціями та захистом прав людини. Розв'язання цієї проблеми вимагає міждисциплінарної координації, поєднання зусиль наукової спільноти, державних інституцій і розробників технологій, що у перспективі дозволить сформувати в Україні сталу екосистему відповідального та етичного використання штучного інтелекту [1;2;3;13].

Література

1. OECD. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. — Paris : OECD Publishing, 2019. — [Електронний ресурс]. — Режим доступу: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (Дата звернення: 23.10.2025).
2. European Commission. Ethics Guidelines for Trustworthy AI / High-Level Expert Group on Artificial Intelligence. — Brussels, 2019. — [Електронний ресурс]. — Режим доступу: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Дата звернення: 23.10.2025).
3. European Commission. EU Artificial Intelligence Act (Regulation (EU) 2024/1689) // Official Journal of the European Union. — 2024. — [Електронний ресурс]. — Режим доступу: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 (Дата звернення: 23.10.2025).
4. International Organization for Standardization (ISO); International Electrotechnical Commission (IEC). ISO/IEC 42001:2023 – Artificial Intelligence Management System. — Geneva : ISO, 2023.
5. International Organization for Standardization (ISO); International Electrotechnical Commission (IEC). ISO/IEC 23894:2023 – Artificial Intelligence — Risk Management. — Geneva : ISO, 2023.
6. IEEE Standards Association. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version II. — 2018. — [Електронний ресурс]. — Режим доступу: https://ethicsinaction.ieee.org (Дата звернення: 23.10.2025).
7. Ribeiro, M. T.; Singh, S.; Guestrin, C. "Why Should I Trust You?" Explaining the Predictions of Any Classifier // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. — New York : ACM, 2016. — DOI: https://doi.org/10.1145/2939672.2939778.
8. Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions (SHAP) // *Advances in Neural Information Processing Systems (NeurIPS)*. — 2017. — [Електронний ресурс]. — Режим доступу: https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (Дата звернення: 23.10.2025).
9. Selvaraju, R. R.; Cogswell, M.; Das, A. та ін. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization // *International Journal of Computer Vision*. — 2020. — Vol. 128. — P. 336–359. — DOI: https://doi.org/10.1007/s11263-019-01228-7.
10. European Commission Joint Research Centre (JRC). Bias in Artificial Intelligence Systems and Big Data Analytics. — Luxembourg : Publications Office of the European Union, 2023. — DOI: https://doi.org/10.2760/503151.

11. Міністерство цифрової трансформації України. Концепція розвитку штучного інтелекту в Україні**.** — Київ, 2023. — [Електронний ресурс]. — Режим доступу: https://thedigital.gov.ua (Дата звернення: 23.10.2025).

12. United States Congress. Algorithmic Accountability Act of 2022 (S.3572)**.** — Washington, D.C. : U.S. Government Publishing Office, 2022. — [Електронний ресурс]. — Режим доступу: https://www.congress.gov/bill/117th-congress/senate-bill/3572 (Дата звернення: 23.10.2025).

13. Floridi, L.; Cowls, J.; Beltrametti, M. та ін. AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations // *Minds and Machines*. — 2018. — Vol. 28(4). — P. 689–707. — DOI: https://doi.org/10.1007/s11023-018-9482-5.

References

1. OECD. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. — Paris : OECD Publishing, 2019. — [Elektronnyi resurs]. — Rezhym dostupu: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (Data zvernennia: 23.10.2025).

2. European Commission. Ethics Guidelines for Trustworthy AI / High-Level Expert Group on Artificial Intelligence. — Brussels, 2019. — [Elektronnyi resurs]. — Rezhym dostupu: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Data zvernennia: 23.10.2025).

3. European Commission. EU Artificial Intelligence Act (Regulation (EU) 2024/1689) // Official Journal of the European Union. — 2024. — [Elektronnyi resurs]. — Rezhym dostupu: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689 (Data zvernennia: 23.10.2025).

4. International Organization for Standardization (ISO); International Electrotechnical Commission (IEC). ISO/IEC 42001:2023 – Artificial Intelligence Management System. — Geneva : ISO, 2023.

5. International Organization for Standardization (ISO); International Electrotechnical Commission (IEC). ISO/IEC 23894:2023 – Artificial Intelligence — Risk Management. — Geneva : ISO, 2023.

6. IEEE Standards Association. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. Version II. — 2018. — [Elektronnyi resurs]. — Rezhym dostupu: https://ethicsinaction.ieee.org (Data zvernennia: 23.10.2025).

7. Ribeiro, M. T.; Singh, S.; Guestrin, C. “Why Should I Trust You?” Explaining the Predictions of Any Classifier // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — New York : ACM, 2016. — DOI: https://doi.org/10.1145/2939672.2939778.

8. Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions (SHAP) // Advances in Neural Information Processing Systems (NeurIPS). — 2017. — [Elektronnyi resurs]. — Rezhym dostupu: https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (Data zvernennia: 23.10.2025).

9. Selvaraju, R. R.; Cogswell, M.; Das, A. ta in. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization // International Journal of Computer Vision. — 2020. — Vol. 128. — P. 336–359. — DOI: https://doi.org/10.1007/s11263-019-01228-7.

10. European Commission Joint Research Centre (JRC). Bias in Artificial Intelligence Systems and Big Data Analytics. — Luxembourg : Publications Office of the European Union, 2023. — DOI: https://doi.org/10.2760/503151.

11. Ministerstvo tsyfrovoi transformatsii Ukrainy. Kontseptsiia rozvytku shtuchnoho intelektu v Ukraini. — Kyiv, 2023. — [Elektronnyi resurs]. — Rezhym dostupu: https://thedigital.gov.ua (Data zvernennia: 23.10.2025).

12. United States Congress. Algorithmic Accountability Act of 2022 (S.3572). — Washington, D.C. : U.S. Government Publishing Office, 2022. — [Elektronnyi resurs]. — Rezhym dostupu: https://www.congress.gov/bill/117th-congress/senate-bill/3572 (Data zvernennia: 23.10.2025).

13. Floridi, L.; Cowls, J.; Beltrametti, M. ta in. AI4People — An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations // Minds and Machines. — 2018. — Vol. 28(4). — P. 689–707. — DOI: https://doi.org/10.1007/s11023-018-9482-5.