https://doi.org/10.31891/2219-9365-2026-86-43
УДК 004.8

ЛІП'ЯНІНА-ГОНЧАРЕНКО Христина
Західноукраїнський національний університет
https://orcid.org/0000-0002-2441-6292
e-mail:kh.lipianina@wunu.edu.ua
КОМАР Мирослав
Західноукраїнський національний університет
https://orcid.org/0000-0001-6541-0359
e-mail: mko@wunu.edu.ua
БИКОВИЙ Павло
Західноукраїнський національний університет
https://orcid.org/0000-0002-5705-5702
e-mail: pb@wunu.edu.ua
ЮРКІВ Христина
Західноукраїнський національний університет
https://orcid.org/0009-0007-4917-3251
e-mail: kh.yurkiv@wunu.edu.ua

КОРПОРАТИВНЕ УПРАВЛІННЯ ВІДПОВІДАЛЬНИМ ШТУЧНИМ ІНТЕЛЕКТОМ У ПРОВІДНИХ ІТ-КОРПОРАЦІЯХ: ПОРІВНЯЛЬНИЙ АНАЛІЗ ПРИНЦИПІВ, ВНУТРІШНІХ СТАНДАРТІВ І ПРАКТИК

У статті здійснено порівняльний аналіз корпоративного управління відповідальним штучним інтелектом у провідних ІТ-корпораціях (Google, Microsoft, IBM, Meta, Amazon, Apple, OpenAI) шляхом зіставлення задекларованих принципів, внутрішніх стандартів і практик їхнього практичного втілення. Дослідження ґрунтується на якісному контент-аналізі публічно доступних корпоративних матеріалів та рамковій логіці «принципи → корпоративне управління → процеси → контрольні механізми → моніторинг → зворотний зв'язок». Запропоновано порівняльну матрицю оцінювання зрілості практик за дев'ятьма критеріями (справедливість, прозорість/пояснюваність, безпека/стійкість, приватність, підзвітність/людський нагляд, суспільна користь, корпоративне управління, моніторинг/звітування, продуктові обмеження) із трирівневою шкалою 0–2. Показано, що попри наявність спільного «ціннісного ядра» відповідального ШІ, рівень інституціоналізації та операціоналізації принципів істотно різниться між компаніями й залежить від продуктового профілю та ризиків застосування. Виявлено стійкі патерни: перехід до ризик-орієнтованого управління як операційної норми, стандартизація артефактів прозорості (паспортизація моделей/даних) та добровільні самообмеження у високоризикових сферах як індикатор підзвітності. Узагальнений контур корпоративного впровадження відповідального ШІ та результати порівняння можуть бути використані для інтерпретації зрілості корпоративних моделей і адаптації кращих практик в організаціях в умовах посилення регуляторних і суспільних вимог.

Ключові слова: відповідальний штучний інтелект; корпоративне управління; управління ризиками; прозорість і підзвітність; внутрішні стандарти; саморегуляція.

LIPIANINA-HONCHARENKO Khrystyna, KOMAR Myroslav,
BYKOVYY Pavlo , YURKIV Khrystyna
West Ukrainian National University

CORPORATE GOVERNANCE OF RESPONSIBLE ARTIFICIAL INTELLIGENCE IN LEADING IT CORPORATIONS: A COMPARATIVE ANALYSIS OF PRINCIPLES, INTERNAL STANDARDS, AND PRACTICES

This article presents a comparative analysis of corporate governance of responsible artificial intelligence in leading IT corporations (Google, Microsoft, IBM, Meta, Amazon, Apple, OpenAI) by examining declared principles, internal standards, and practices of their practical implementation. The study is based on a qualitative content analysis of publicly available corporate materials and follows a framework logic of "principles → corporate governance → processes → control mechanisms → monitoring → feedback." A comparative matrix for assessing the maturity of practices is proposed, based on nine criteria (fairness, transparency/explainability, safety/robustness, privacy, accountability/human oversight, social benefit, corporate governance, monitoring/reporting, product restrictions) using a three-level scale (0–2). The findings show that despite a shared "value core" of responsible AI, the level of institutionalization and operationalization of principles varies significantly across companies and depends on product profiles and risk exposure. Several устойчиві patterns are identified: the transition to risk-based governance as an operational norm, the standardization of transparency artifacts (model/data documentation), and voluntary self-restrictions in high-risk areas as an indicator of accountability. The generalized framework for corporate implementation of responsible AI and the comparative results can be used to interpret the maturity of corporate models and to adapt best practices in organizations operating under increasing regulatory and societal demands.

Keywords: responsible artificial intelligence; corporate governance; risk management; transparency and accountability; internal standards; self-regulation.

Стаття надійшла до редакції / Received 18.03.2026
Прийнята до друку / Accepted 04.05.2026
Опубліковано / Published 31.05.2026

© ЛІП'ЯНІНА-ГОНЧАРЕНКО Христина, КОМАР Мирослав, БИКОВИЙ Павло, ЮРКІВ Христина

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

Штучний інтелект (ШІ) став базовою технологічною основою цифрової трансформації та масштабно інтегрується в продукти, послуги й управлінські процеси організацій. Паралельно розширюється спектр ризиків: дискримінаційна упередженість моделей, непрозорість прийняття рішень, уразливості інформаційної безпеки, витоки даних, порушення прав людини, а також репутаційні й регуляторні наслідки для компаній, що впроваджують системи ШІ й машинного навчання (МН) у критичних або соціально чутливих сферах. У відповідь на ці виклики формується концепція відповідального ШІ, яка передбачає не лише етичні декларації, а й практичні механізми забезпечення безпечності, законності, підзвітності та керованості ШІ-систем на всіх етапах їхнього життєвого циклу.

У провідних ІТ-корпораціях відповідальний ШІ дедалі частіше розглядають як складову корпоративного управління: від формулювання принципів і «червоних ліній» – до запровадження внутрішніх стандартів, створення органів нагляду, визначення процедур контролю ризиків, застосування інженерних засобів перевірки моделей і організації моніторингу після впровадження. Водночас реальна зрілість таких підходів визначається не стільки кількістю задекларованих принципів, скільки ступенем їх практичного втілення у процесах розроблення та прийняття управлінських рішень.

У цьому контексті актуальним є порівняльне дослідження корпоративних моделей відповідального ШІ у провідних технологічних корпораціях, оскільки саме вони фактично задають галузеві орієнтири управління ризиками, впливають на очікування регуляторів і ринку та формують прикладні практики, придатні до адаптації іншими організаціями.

Попри активне поширення практик відповідального ШІ в корпоративному секторі, зберігається методологічна й прикладна проблема: між проголошеними принципами та їх реальним виконанням часто існує розрив. Він зумовлений неоднаковим рівнем формалізації внутрішніх стандартів, різною організаційною архітектурою системи управління, відмінностями у продуктових ризиках і ступені інтеграції контрольних механізмів в інженерні процеси. Унаслідок цього одна й та сама ціннісна рамка (справедливість, прозорість, безпека, приватність, підзвітність) може мати істотно різний процедурний і технічний зміст у різних компаніях: від розгорнутих механізмів експертизи та аудиту – до переважно декларативного рівня.

Проблема ускладнюється тим, що сучасні системи ШІ та МН, зокрема генеративні моделі, впроваджуються швидко та в широкому спектрі застосувань, де наслідки помилок або зловживань здатні спричиняти масштабний соціальний ефект. Це підсилює потребу в порівнюваних підходах до управління ризиками, стандартизованих артефактах прозорості (документуванні моделей і даних), технічних запобіжниках безпеки та процедурах післявпроваджувального моніторингу.

Актуальність дослідження визначається необхідністю: (1) системно описати, як провідні ІТ-корпорації переводять відповідальний ШІ із рівня принципів на рівень стандартів і процедур; (2) виявити спільні закономірності корпоративного управління та ключові відмінності, пов'язані з продуктовою специфікою і профілем ризиків; (3) сформувати узагальнену модель (контур) упровадження відповідального ШІ, придатну для інтерпретації зрілості корпоративних практик і подальшої адаптації в організаціях, що впроваджують системи ШІ/МН в умовах посилення регуляторних і суспільних вимог.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У наукових публікаціях останніх років простежується перехід від нормативно-декларативного трактування відповідального ШІ до дослідження його інституційного та операційного забезпечення на рівні організацій. Якщо перші праці зосереджувалися переважно на етичних принципах (справедливість, прозорість, підзвітність тощо), то сучасні дослідження акцентують увагу на питаннях корпоративного управління, механізмах управління ризиками та формалізації процедур контролю впродовж життєвого циклу систем ШІ.

Системний огляд і рамкова модель управління відповідальним ШІ, запропоновані в [1], підкреслюють, що зрілість корпоративного підходу визначається не кількістю проголошених принципів, а наявністю формалізованих управлінських структур, процедур оцінювання ризиків, механізмів внутрішнього аудиту та чітко визначених сфер відповідальності. Автори наголошують на потребі інтегрування відповідального ШІ на стратегічному рівні організації через створення спеціалізованих управлінських структур, запровадження стандартизованих процедур експертизи високоризикових застосувань і регулярного звітування. Такий підхід узгоджується з практиками провідних ІТ-компаній, які розробляють внутрішні стандарти та створюють профільні комітети як інституціоналізовані механізми контролю.

Важливим напрямом є практичне втілення принципів відповідального ШІ в прикладних

управлінських моделях. У роботі [2] запропоновано поетапний підхід до впровадження відповідального використання ШІ, що охоплює: (i) формування нормативно-ціннісної рамки, (ii) розподіл ролей і відповідальностей, (iii) інтеграцію процедур оцінювання ризиків у продуктовий цикл, (iv) створення механізмів комунікації та підзвітності. Дослідники зазначають, що відповідальний ШІ має бути «вбудованим» у процеси розроблення, а не функціонувати як зовнішній контрольний контур. Це підтверджує доцільність узагальненої логіки, за якою принципи послідовно трансформуються в корпоративне управління, процеси, контрольні механізми, моніторинг і зворотний зв'язок, що в сукупності відображає контур корпоративного впровадження відповідального ШІ.

Окрему групу становлять дослідження, присвячені стандартизації прозорості та документування систем ШІ. Класичною працею в цьому напрямі є концепція паспортів моделей (Model Cards) [3], яка передбачає структуроване описання характеристик моделі, її призначення, обмежень, результатів перевірок на упередженість та потенційних ризиків застосування. Паспорти моделей розглядають як інструмент підвищення можливості аудиту та підзвітності, що зменшує інформаційну асиметрію між розробниками й користувачами. Аналогічно, концепція паспортів наборів даних (Datasheets for Datasets) [4] спрямована на систематичне документування походження, складу, методів збирання та можливих ризиків використання наборів даних. Обидва підходи заклали підґрунтя для формування корпоративних практик паспортизації моделей і наборів даних (картки систем/моделей, інформаційні паспорти, примітки щодо прозорості), які нині активно впроваджують провідні технологічні корпорації.

Подальший розвиток тематики пов'язаний з емпіричним оцінюванням дієвості механізмів корпоративного управління у сфері відповідального ШІ. У роботі [5] на основі великомасштабного міжгалузевого аналізу показано, що наявність формалізованих механізмів управління інформацією в контексті відповідального ШІ позитивно пов'язана з показниками корпоративної результативності. Автори доводять, що впровадження процедур контролю ризиків, внутрішніх стандартів і прозорого звітування не лише знижує регуляторні та репутаційні ризики, а й формує довгострокові конкурентні переваги. Отже, відповідальний ШІ постає не лише як етична вимога, а як стратегічний чинник сталого розвитку компанії.

Узагальнення сучасних досліджень дає змогу виокремити кілька ключових тенденцій. По-перше, відбувається інституціоналізація відповідального ШІ через формування управлінських структур і процедур ризик-орієнтованого контролю [1, 2]. По-друге, стандартизація документування моделей і даних (паспорти моделей і паспорти наборів даних) створює основу для прозорості й аудиту [3, 4]. По-третє, емпіричні результати демонструють економічну доцільність інтеграції відповідального ШІ в корпоративні стратегії [5].

Зазначені наукові напрацювання формують теоретичне підґрунтя для аналізу внутрішніх стандартів і практик провідних ІТ-компаній та дають підстави розглядати їхні підходи як частину ширшого процесу еволюції корпоративного управління ШІ – від декларативних принципів до системно організованих механізмів відповідальності.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Мета дослідження полягає у здійсненні порівняльного аналізу корпоративного управління у сфері відповідального ШІ в провідних ІТ-корпораціях шляхом зіставлення: (1) задекларованих принципів, (2) внутрішніх стандартів і практик їх практичного втілення, (3) участі у зовнішніх ініціативах саморегуляції, а також у формуванні узагальненого контуру впровадження відповідального ШІ впродовж життєвого циклу систем ШІ/МН.

Для досягнення мети визначено такі завдання:

систематизувати корпоративні принципи відповідального ШІ та визначити їхнє спільне «ціннісне ядро» і специфічні акценти в різних компаніях;

проаналізувати механізми внутрішнього корпоративного управління (органи та функції у сфері відповідального ШІ, політики, стандарти, ролі й відповідальність) і процесні процедури контролю ризиків (експертиза чутливих застосувань, аудит, тестування на зловживання, валідація);

оцінити інженерні засоби практичного втілення відповідального ШІ (інструменти забезпечення справедливості, пояснюваності, стійкості та приватності; технічні запобіжники; артефакти прозорості), а також підходи до післявпроваджувального моніторингу й звітування;

визначити роль зовнішніх ініціатив саморегуляції у зближенні практик безпеки, прозорості та управління ризиками;

сформувати порівняльну матрицю підходів і узагальнений контур корпоративного впровадження відповідального ШІ та інтерпретувати спільні закономірності й ключові відмінності.

У роботі застосовано порівняльний якісний аналіз змісту публічно доступних корпоративних матеріалів, присвячених відповідальному ШІ, що охоплюють принципи, політики, внутрішні стандарти та методичні настанови, звітні матеріали, інструментальні ініціативи, а також декларативні документи провідних ІТ-компаній: Google, Microsoft, IBM, Meta, Amazon, Apple, OpenAI. Методологічна логіка дослідження ґрунтується на рамковому підході «принципи → корпоративне управління → процеси → контрольні механізми → моніторинг → зворотний зв'язок», який дає змогу зіставляти задекларовані цінності з їх практичним втіленням у процесах і підтримкою інженерними засобами контролю впродовж життєвого циклу

систем ШІ та МН.

Одиницею аналізу є компанія (рядок матриці) та критерій зрілості практик відповідального ШІ (стовпчик матриці). Для систематизації результатів сформовано порівняльну матрицю (табл. 1), у межах якої кодування здійснено за дев'ятьма критеріями: недискримінація (справедливість), прозорість і пояснюваність, безпечність і стійкість, приватність і захист даних, підзвітність і людський нагляд, інклюзивність і суспільна користь, корпоративне управління (стандарти, комітети, процедури експертизи), моніторинг і звітування, продуктові обмеження. Базу джерел для кодування становлять як нормативно-ціннісні декларації (принципи, хартії), так і операційні артефакти впровадження (процедури експертизи, тестування на зловживання, обмеження застосувань, інструменти оцінювання справедливості, пояснюваності та стійкості, практики звітування).

Для кожного критерію застосовано трирівневу шкалу 0–2 (табл. 1): 0 – ознака не задекларована або не підтверджується виявленими корпоративними артефактами; 1 – ознака задекларована або частково відображена, проте без достатньої формалізації (наприклад, на рівні загальних принципів без опису процедур, ролей, інструментів чи режимів звітування); 2 – ознака формалізована та підтримана управлінськими й інженерними механізмами (наявні стандарти, комітети або відповідальні ролі; визначені процедури експертизи, аудиту та встановлення обмежень; застосовуються інструментальні засоби контролю та/або практики моніторингу після впровадження і звітування). Оцінювання здійснювали на підставі сукупності свідчень у документах: для рівня 2 вимагалася наявність не лише декларативної заяви, а й ознак інституціоналізації (корпоративного управління) та практичної реалізації (процесів, контрольних механізмів і моніторингу), що узгоджується з підходом, за яким зрілість практик відповідального ШІ визначається не кількістю проголошених принципів, а наявністю процедур оцінювання ризиків, аудиту, розподілу відповідальностей і регулярного звітування.

Для зниження суб'єктивності кодування застосовано процедурну валідацію: (і) первинне кодування за критеріями матриці; (іі) повторну перевірку відповідності присвоєних значень 0–2 на основі повторного перегляду джерел і узгодження оцінок між критеріями (зокрема між корпоративним управлінням, моніторингом і звітуванням та продуктовими обмеженнями), оскільки саме ці виміри є найбільш чутливими до декларативності й маркетингових формулювань. Додатково виконано інтерпретаційне узагальнення результатів у формі якісного порівняння закономірностей практичного втілення принципів між компаніями, зокрема переходу від «м'яких» декларацій до процесно-інженерних механізмів контролю ризиків.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Провідні технологічні корпорації сформували власні публічні принципи та внутрішні стандарти етичного й відповідального використання ШІ, які виконують роль корпоративних механізмів саморегулювання та зниження ризиків під час розроблення й експлуатації систем ШІ та МН. У більшості випадків ці підходи концентруються навколо спільного «ядра» вимог: етичність, прозорість, безпечність, недопущення дискримінаційної упередженості, підзвітність і людський контроль, а також захист приватності та даних. У цьому контексті показовими є практики Google, Microsoft, IBM, Meta, Amazon, Apple та OpenAI, які поєднують декларативні принципи з конкретними інструментами впровадження та участю у міжнародних ініціативах.

Google однією з перших кодифікувала корпоративні принципи ШІ: ще у 2018 році було оприлюднено сім принципів [6], а в оновленому вигляді 2025 року вони охоплюють соціальну корисність, запобігання несправедливій упередженості, безпеку, підзвітність людині, приватність, наукову досконалість і відповідність допустимим цілям. Важливою ознакою підходу Google є формалізація «червоних ліній»: компанія декларує відмову від розроблення зброї на основі ШІ, систем незаконного стеження та застосувань, що суперечать міжнародному праву й правам людини [6]. Для практичного втілення принципів створено внутрішню команду відповідальних інновацій, яка забезпечує експертизу проєктів на відповідність принципам [7], а також запроваджено інструменти для підвищення прозорості й можливості аудиту (Explainable AI, Model Cards, засоби TensorFlow для аудиту). Додатково Google підтримує регулярну звітність (щорічні доповіді з відповідального ШІ; станом на 2024 рік – шостий звіт) [8] і поширює практики через освітні матеріали та курси [9]. На міжнародному рівні компанія є співзасновником Partnership on AI та ініціатором Frontier Model Forum (2023) разом з OpenAI, Microsoft та Anthropic [10].

Microsoft вибудовує підхід навколо шести принципів відповідального ШІ – справедливість, надійність і безпека, приватність і захист, інклюзивність, прозорість, підзвітність [11] – і підкріплює їх детальним внутрішнім стандартом: Responsible AI Standard (публічно відома версія v2, 2022). Корпоративне управління реалізується через спеціалізовані структури (Office of Responsible AI; раніше – рада Aether), які здійснюють формування політик, навчання команд і процедури експертизи чутливих застосувань. На практичному рівні Microsoft розвиває інженерні засоби контролю (контрольні переліки, інструменти забезпечення відповідності, засоби тестування упередженості), зокрема відкриті бібліотеки Fairlearn та InterpretML. Показовим прикладом ризик-орієнтованого саморегулювання є обмеження функціоналу розпізнавання облич (вилучення визначення емоцій тощо) та позиція щодо продажу таких технологій поліції

до появи відповідного правового регулювання [12]. Компанія також публікує щорічні звіти з прозорості ШІ та Transparency Notes для сервісів Azure AI. У зовнішньому контурі Microsoft є засновником Partnership on AI, співпідписантом «Rome Call for AI Ethics» (2020) [13] та учасником Frontier Model Forum (2023) [10].

IBM акцентує корпоративну етику ШІ через «Принципи довіри та прозорості» (2018) [14], у яких ключовими є:

- ШІ має доповнювати, а не замінювати людину;
- дані та інсайти належать їхнім творцям (клієнтам);
- технології повинні бути прозорими й пояснюваними [14].

Реалізацію забезпечує AI Ethics Board [15], а також технологічні інструменти для забезпечення справедливості, пояснюваності та стійкості: AI Fairness 360, AI Explainability 360, Adversarial Robustness Toolbox і концепція FactSheets як «паспортів» моделей для прозорості та аудиту [15]. Публічно значущим рішенням стала відмова IBM від продажу та досліджень універсального програмного забезпечення для розпізнавання облич у 2020 році з чітким запереченням масового стеження, расового профайлінгу та порушень прав і свобод [16]. IBM бере участь у міжнародних ініціативах (зокрема Partnership on AI) і підкреслює роль альянсів, таких як Data & Trust Alliance у виробленні запобіжників алгоритмічної упередженості [15], а також підтримує «Rome Call» [13].

Meta структурує відповідальний ШІ через п'ять складових: приватність і безпека [17], справедливість та інклюзія, стійкість і безпечність, прозорість [18], підзвітність і корпоративне управління [19]. У 2018 році компанія створила централізовану команду з відповідального ШІ, однак у 2022 році провела реорганізацію та розподілила відповідні функції між продуктовими командами, що викликало публічні дискусії [20]. На рівні інструментів Meta розвиває внутрішні методичні настанови (Responsible Use Guide) [18], практики аудитів впливу алгоритмів, механізми пояснюваності для користувачів та зовнішні перевірки (зокрема у сфері соціальних платформ і модерації контенту). У міжнародному вимірі Meta приєдналася до Frontier Model Forum у 2023 році [10], оголосила AI Alliance [21] та відображає компоненти прав людини й управління у Human Rights Report [22].

Amazon формалізує відповідальність через набір із восьми пріоритетів (2023): справедливість, пояснюваність, приватність і безпека, безпечність, керованість, достовірність і стійкість, корпоративне управління, прозорість [10]. Впровадження підтримується міжфункціональною експертизою та екосистемою AWS, зокрема через Responsible AI Center і публічні настанови для розробників [23]. Компанія підкреслює масштаб внутрішніх інвестицій (інструменти та навчання), а також розвиває технічні запобіжники для генеративних моделей. Як приклад реакції на суспільні ризики, Amazon запровадила мораторій на використання Rekognition поліцією (2020) [24, 25], а згодом підтвердила відмову надавати технологію правоохоронним органам до появи федерального регулювання [12]. На зовнішньому рівні компанія також долучається до міжкорпоративних ініціатив, зокрема Frontier Model Forum [10].

Apple традиційно вибудовує «етичну вісь» довкола приватності та безпеки, а з 2023–2024 років чіткіше артикулює підхід до ШІ. Зокрема, у зв'язку з Apple Intelligence компанія заявила чотири принципи відповідальної розробки:

наділяти користувачів інтелектуальними інструментами;

представляти інтереси користувачів з урахуванням різноманіття та недопущення стереотипів;

здійснювати проєктування з обережністю (оцінювання зловживань і ризиків на кожному етапі);

забезпечувати захист приватності завдяки локальній обробці на пристрої та інфраструктурі Private Cloud Compute [26].

Окремо підкреслюється трактування приватності як фундаментального права [27] та використання підходів, що зберігають приватність, зокрема диференційної приватності (з 2016 року) [28]. У контексті великих мовних моделей Apple декларує невикористання приватних даних користувачів чи їхнього контенту для навчання моделей [29, 30], а також застосування технік підвищення безпеки (фільтрація токсичного вмісту й персональних даних, посттренувальні процедури, наприклад RLHF) [26].

OpenAI вплинула на корпоративний дискурс відповідального ШІ через OpenAI Charter (2018) [31], який визначає місію – забезпечити, щоб загальний ШІ приніс користь усьому людству, із пріоритетом довгострокової безпеки над короткостроковою вигодою та уникненням «гонки озброєнь» у цій сфері. Хартія також фіксує баланс відкритості досліджень і безпеки, зокрема готовність обмежувати повну публікацію за наявності ризиків зловживань [31], і містить принцип кооперації: у разі, якщо інша організація першою наближається до створення безпечного загального ШІ, OpenAI декларує готовність допомогти та сповільнити конкурентну динаміку [32]. Практики операціоналізації включають тестування на зловживання перед релізами, публікацію System Cards (наприклад, для GPT-4), використання Moderation API і політик використання (Usage Policies) [31], а також ініціативи залучення суспільства до нормування поведінки моделей (грантова програма 2023 року) [33]. Підхід поступового розгортання і «відповідальної публікації» (зокрема історично для GPT-2) описується як зразок обережності та підтримується галузевими нормами [34]; на рівні міжкорпоративної координації OpenAI є серед засновників Frontier Model Forum [10].

Узагальнюючи, спільний знаменник корпоративних принципів відповідального ШІ включає: (і)

справедливість і недискримінацію, (ii) прозорість і пояснюваність, (iii) безпеку та надійність, (iv) приватність і захист даних, (v) підзвітність і людський контроль, (vi) інклюзивність і соціальну корисність, а в окремих випадках – акцент на науковій етиці та співпраці [6–11, 14, 17–19, 31]. Практична імплементація цих принципів зазвичай реалізується через поєднання: етичних комітетів і відповідальних осіб; процедур експертизи чутливих застосувань перед релізом; інженерних інструментів забезпечення справедливості, пояснюваності та стійкості; моніторингу після розгортання; політик використання та механізмів блокування порушень; корпоративної освіти й формування культури повідомлення про ризики; а також добровільного обмеження або відмови від небезпечних продуктів (зокрема у сфері розпізнавання облич) [7, 12, 15-16, 20, 24–25].

Водночас корпоративні моделі управління не розвиваються ізольовано: компанії активно долучаються до міжнародних альянсів і багатосторонніх ініціатив (Partnership on AI, Global Partnership on AI, Frontier Model Forum), підтримують глобальні рамки та беруть участь у діалозі з регуляторами щодо правил безпеки, маркування вмісту та стандартизації практик управління ризиками [10, 13].

Перехід від декларативних принципів до інституціоналізованих механізмів відповідального ШІ у провідних корпораціях відбувається через формування внутрішніх стандартів, процедур контролю ризиків і інженерних інструментів, що охоплюють увесь життєвий цикл систем ШІ – від постановки завдання та проєктування до розгортання й післявпроваджувального моніторингу. У більшості розглянутих компаній простежується поєднання трьох рівнів управління: (i) управлінський контур (комітети, офіси, ради), (ii) процесний контур (експертиза чутливих застосувань, аудит, тестування на зловживання), (iii) технічний контур (інструментальні засоби забезпечення справедливості, пояснюваності, стійкості та приватності). Зв'язок між принципами, рівнем корпоративного управління, процесними процедурами, технічними контрольними механізмами та моніторингом і звітуванням, а також роль зворотного зв'язку для оновлення політик і стандартів узагальнено на рис. 1.

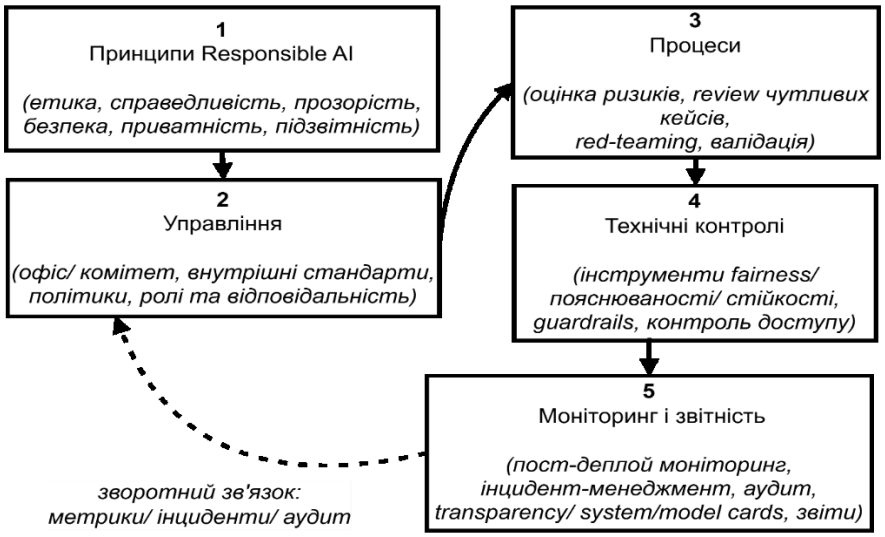

Рис. 1. Узагальнений контур корпоративного впровадження Responsible AI: від принципів до контролів, моніторингу та підзвітності

На рівні корпоративного управління суттєвим є створення спеціалізованих органів або функцій у сфері відповідального ШІ. У Google це реалізовано через команду Responsible Innovation як центр компетенцій, що забезпечує експертизу проєктів на відповідність принципам [7]. У Microsoft функціонує Office of Responsible AI, а історично – Aether Committee, які формують політики, здійснюють навчання команд та проводять процедури експертизи чутливих застосувань для високоризикових сценаріїв [11]. IBM інтегрувала етичний контроль через AI Ethics Board, який забезпечує застосування принципів довіри й прозорості у продуктових рішеннях [15]. Meta розвивала централізовану Responsible AI team (2018–2022), однак згодом перейшла до моделі розподіленої відповідальності, інтегруючи відповідні функції у продуктові команди [20]. Amazon підтримує міжфункціональні механізми оцінювання продуктів ШІ у зв'язці напрямів AWS із безпекою, приватністю, політиками та правовим супроводом [23].

На рівні процесів ключовими є експертиза «чутливих застосувань» і ризик-орієнтовані рішення щодо продуктового портфеля. Microsoft демонструє підхід «від політик до продукту», поєднуючи внутрішній стандарт Responsible AI Standard з інструментами забезпечення відповідності вимогам та експертизою ризикових випадків, що відображається, зокрема, в обмеженні окремих функцій розпізнавання облич (зокрема визначення емоцій) та у задекларованій позиції щодо продажу таких технологій правоохоронним органам до появи регуляторних норм [12]. IBM у 2020 році публічно припинила продаж і дослідження універсальних систем розпізнавання облич, мотивуючи це неприйнятністю масового стеження, расового профайлінгу та

порушень прав і свобод [16]. Amazon запровадила мораторій на використання Rekognition поліцією (2020) з подальшим продовженням обмежень до появи федерального регулювання [12, 24–25]. Такі приклади свідчать, що внутрішні стандарти відповідального ШІ в окремих випадках призводять до добровільної відмови від комерційно привабливих, але високоризикових застосувань.

На рівні інженерної реалізації компанії роблять ставку на стандартизовані артефакти прозорості та інструменти з відкритим програмним кодом. Google однією з перших масштабувала підхід Model Cards та інструменти Explainable AI як основу документування моделей і підвищення прозорості [8]. Microsoft розвиває інструменти Fairlearn та InterpretML для оцінювання й пом'якшення упередженості та для підвищення пояснюваності моделей [11]. IBM запропонувала комплексні набори інструментів AI Fairness 360, AI Explainability 360 та Adversarial Robustness Toolbox, а також концепцію FactSheets як паспортизацію моделей із фіксацією даних, тестів, ризиків та результатів аудиту [15]. Amazon у хмарній екосистемі AWS просуває інструменти для виявлення упередженості й забезпечення керованості (зокрема через практики відповідального ШІ та настанови для клієнтів) [23]. У контексті генеративних моделей окремої ваги набувають технічні запобіжники та політики фільтрації небажаного вмісту, що прямо позиціонуються як частина стандартів безпеки й керованості [23].

Нарешті, внутрішні стандарти закріплюються через регулярну звітність, навчання та формування культури відповідального проєктування. Google публікує щорічні звіти з відповідального ШІ як інструмент підзвітності та фіксації прогресу [8], а IBM декларує масштабні програми навчання партнерів етиці ШІ [15]. У сукупності це формує організаційний механізм, у межах якого відповідальний ШІ стає не декларацією, а операційною нормою, інтегрованою в інженерні процеси та управління продуктами.

Корпоративні практики відповідального ШІ суттєво посилюються завдяки участі компаній у міжнародних ініціативах, що виконують функції координації саморегуляції, вироблення добровільних норм та узгодження підходів до безпеки передових моделей. Одним із найстійкіших міжкорпоративних інструментів є Partnership on AI (PAI), до якого входять Amazon, Apple, Meta (Facebook), Google/DeepMind, IBM, Microsoft, OpenAI та інші, а також робочі групи за напрямами справедливості, цілісності медіа й відповідальної публікації результатів [34].

Другою ключовою платформою саморегуляції для генеративних і передових моделей став Frontier Model Forum, започаткований у 2023 році (учасники: OpenAI, Microsoft, Google, Anthropic; надалі – Meta і Amazon), який зосереджується на координації досліджень безпеки, обміні результатами стрес-тестів і взаємодії з урядами щодо правил для найбільш потужних моделей [10]. У цьому ж контексті важливою формою саморегуляції є добровільні зобов'язання компаній перед державними інституціями (зокрема ініціативи США щодо безпеки ШІ та маркування вмісту, створеного ШІ), які стимулюють розвиток незалежного тестування, прозорості та інформаційного обміну щодо ризиків [10].

Окремим напрямом є участь компаній у багатосторонніх ініціативах, де взаємодіють уряди й приватний сектор. Прикладом є Global Partnership on AI (GPAI) з 2020 року, у межах якого експерти від провідних технологічних корпорацій долучаються до робочих груп з управління ШІ та вироблення рекомендацій для держав [10]. Паралельно компанії підтримують глобальні ціннісні рамки, зокрема принципи OECD і рекомендації ЮНЕСКО, що уможливлює узгодження корпоративних стандартів із міжнародним дискурсом етики й прав людини [10].

Механізми саморегуляції також охоплюють ініціативи, спрямовані на підвищення довіри до цифрового вмісту та протидію дезінформації. Зокрема, участь у C2PA позиціонується як шлях стандартизації походження вмісту та посилення автентичності мультимедійних матеріалів у контексті ризиків синтетичних підробок (дипфейків) [22]. Нормативно-символічну, але важливу роль у формуванні етичного консенсусу відіграє Rome Call for AI Ethics, який у 2020 році підписали Microsoft та IBM, наголошуючи на пріоритеті прав людини, прозорості й інклюзивності [13].

Загалом участь у таких ініціативах формує спільний галузевий рівень стандартів, який доповнює внутрішні корпоративні політики, сприяє зближенню вимог до безпеки та прозорості й знижує ризики фрагментації підходів між компаніями.

Для систематизації корпоративних підходів до відповідального ШІ доцільною є матриця порівняння за трьома вимірами (табл. 1):

ключові принципи (декларації);

практики та внутрішні ініціативи;

участь у зовнішніх ініціативах.

Така структура дає змогу зіставити «що компанія проголошує» з «як саме вона це реалізує» та «у яких механізмах саморегуляції бере участь».

Таблиця 1.

Порівняльна матриця принципів і практик відповідального ШІ у провідних ІТ-компаніях (шкала 0–2)

Компанія	Недискримінація	Прозорість / пояснюваність	Безпечність / стійкість	Приватність / захист даних	Підзвітність / людський нагляд	Інклюзивність / суспільна користь	Корпоративне управління	Моніторинг / звітування	Продуктові обмеження / «червоні лінії»
Google	2	2	2	2	2	2	2	2	2
Microsoft	2	2	2	2	2	2	2	2	2
IBM	2	2	2	2	2	1	2	1	2
Meta	2	2	2	2	2	2	1	2	1
Amazon	2	2	2	2	2	1	2	1	2
Apple	1	1	2	2	1	1	1	1	1
OpenAI	1	2	2	1	2	2	2	2	2

У вимірі декларативних принципів Google демонструє підхід із чітко визначеним набором принципів і «червоних ліній» [6], Microsoft – стабільний набір принципів із акцентом на дотримання вимог і підзвітність [11], IBM – довіру/прозорість і право власності на дані [14], Meta – п’ять складових відповідального ШІ з фокусом на приватність, безпеку та управління [17–19], Amazon – набір пріоритетів, що підкреслюють керованість, стійкість і корпоративне управління [10], Apple – чотири принципи з центром тяжіння на приватності та «обережному проєктуванні» [26–28], OpenAI – підхід через хартію з пріоритетом довгострокової безпеки та суспільної користі [31, 32].

У вимірі практик матриця фіксує типові інструменти практичного втілення: наявність спеціалізованих структур (Responsible Innovation team [7], Office of Responsible AI [11], AI Ethics Board [15]), процедур перегляду ризикових застосувань і тестування на зловживання, інженерних наборів інструментів (Model Cards [8], Fairlearn/InterpretML [11], AIF360 та FactSheets [15], AWS-орієнтовані інструменти відповідального ШІ [23]), а також практик добровільних обмежень технологій у високоризикових сферах (IBM – припинення застосувань розпізнавання облич [16]; Microsoft – обмеження функцій [12]; Amazon – мораторій на Rekognition [12, 24–25]).

У вимірі зовнішніх ініціатив матриця демонструє високу концентрацію участі в PAI та Frontier Model Forum [10], а також приєднання до ширших багатосторонніх рамок (GPAI, підходи OECD/ЮНЕСКО) [10]. Для компаній, що працюють із вмістом та медіа, додатково значущими є ініціативи, такі як C2PA [22], а для формування етичного консенсусу – Rome Call [13].

Таким чином, матриця виконує не лише описову, а й аналітичну функцію: вона виявляє, які принципи мають реальне процедурно-інструментальне підкріплення, а які залишаються здебільшого декларативними, і дає змогу оцінити зрілість управління відповідальним ШІ за наявністю структур корпоративного управління, процесів контролю та технічних засобів.

Отримані результати засвідчують поступове формування збіжної моделі корпоративного відповідального ШІ, у якій спільне «ціннісне ядро» (справедливість, прозорість, безпечність, приватність, підзвітність, інклюзивність) відтворюється в різних компаніях, але набуває специфічних акцентів залежно від продуктового профілю, ризиків сфери застосування та корпоративної філософії [6-11, 14, 17–19, 31]. Збіжність проявляється не лише на рівні декларацій, а й у практиках: наявності етичних органів управління, процедур експертизи чутливих застосувань, інструментів вимірювання упередженості та пояснюваності, а також у зростанні ролі моніторингу після розгортання та політик використання [7, 11, 15].

Першим стійким патерном є інституціоналізація управління ризиками: відповідальний ШІ дедалі частіше оформлюється як внутрішній стандарт і набір контрольних процедур (від експертизи продуктів до технічних тестів), а не як сукупність декларацій [7, 11]. Другим патерном є паспортизація та документування моделей, що підсилює можливість аудиту й прозорість (Model Cards у Google [8], FactSheets у IBM [15], Transparency Notes у Microsoft [11]). Третім патерном є добровільне самообмеження у високоризикових сферах, зокрема в системах розпізнавання облич (IBM [16]; Microsoft [12]; Amazon [12, 24, 25]), що вказує на готовність частини компаній переводити етичні принципи в реальні продуктово-політичні рішення.

Разом з тим відмінності у підходах залишаються суттєвими. Apple найпослідовніше формує відповідальний ШІ довкола приватності та локальної обробки даних на пристрої (on-device AI, диференційна приватність), позиціонуючи приватність як фундаментальне право [27, 28], і декларує невикористання приватного вмісту користувачів для навчання моделей [29, 30]. OpenAI робить акцент на безпеці потужних моделей і балансі відкритості та запобігання зловживанням, що відображено в хартії [31, 32], тестуванні на зловживання та практиках «відповідальної публікації» [34]. IBM зосереджується на довірі бізнес-клієнтів, прозорості та контролі над даними, підкріплюючи це наборами інструментів і рішенням щодо розпізнавання

облич [14–16]. Meta вимушено фокусується на соціальних ризиках платформ, тому її відповідальний ШІ тісно пов'язаний із модерацією вмісту, прозорістю алгоритмічних рішень і правами людини [17–22], водночас її організаційна модель еволюціонувала від централізованої команди до розподіленого впровадження [20]. Amazon підкреслює інженерну строгість контурів корпоративного управління і керованість генеративних моделей через інфраструктурні інструменти та настанови AWS [23], а також демонструє приклад ризик-обмежень у правоохоронному застосуванні [12, 24, 25]. Google і Microsoft позиціонуються як компанії з найбільш формалізованими підходами, що поєднують принципи, інструменти прозорості/справедливості та активну участь у міжкорпоративних ініціативах [6-12].

У підсумку корпоративні практики відповідального ШІ демонструють еволюцію від декларативних формулювань до більш жорстких інженерно-управлінських механізмів, у межах яких інновації узгоджуються з етикою, управлінням ризиками та міжнародною саморегуляцією [35, 36].

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У статті здійснено порівняльний аналіз корпоративного управління у сфері відповідального ШІ в провідних IT-корпораціях (Google, Microsoft, IBM, Meta, Amazon, Apple, OpenAI) шляхом зіставлення принципів, внутрішніх стандартів, процесних процедур та інженерних контрольних механізмів упродовж життєвого циклу систем ШІ та МН. Отримані результати засвідчують, що у провідних технологічних корпораціях формується збіжна модель відповідального ШІ з відносно стабільним «ціннісним ядром» (справедливість і недискримінація, прозорість і пояснюваність, безпечність і стійкість, приватність і захист даних, підзвітність і людський нагляд, інклюзивність), однак ступінь її практичного втілення істотно різниться між компаніями.

Ключовим висновком є те, що зрілість практик відповідального ШІ визначається не декларативністю принципів, а наявністю інституціоналізованого корпоративного управління та інтегрованих механізмів управління ризиками: спеціалізованих органів або функцій (офіси, комітети), формалізованих стандартів і політик, процедур експертизи «чутливих» застосувань (перегляд, аудит, тестування на зловживання), а також технічних інструментів вимірювання і зниження ризиків (набори інструментів для забезпечення справедливості, пояснюваності, стійкості та приватності; технічні запобіжники) і режимів моніторингу після розгортання та звітування.

Виявлено три стійкі закономірності корпоративної практики:

перехід до ризик-орієнтованого управління як операційної норми (перенесення політик у продуктову реалізацію, стандарти та контрольні процедури);

стандартизація артефактів прозорості та можливості аудиту (паспортизація моделей і даних);

застосування добровільних обмежень у високоризикових сферах як індикатора реальної підзвітності (зокрема у сфері розпізнавання облич).

Водночас відмінності зумовлені продуктовою специфікою та корпоративною філософією: Apple системно вибудовує відповідальний ШІ довкола приватності й підходів локальної обробки на пристрої; OpenAI акцентує безпеку передових моделей і контроль зловживань; Meta – соціальні ризики платформ і управління впливом алгоритмів; Amazon – контури корпоративного управління та керованість у хмарній екосистемі; Google і Microsoft демонструють найбільш формалізовані комплексні моделі поєднання принципів, процесів і інженерних контрольних механізмів.

Участь корпорацій у міжнародних ініціативах саморегуляції посилює зближення практик та формує галузевий рівень норм, що доповнює внутрішні політики, знижуючи ризик фрагментації підходів і підтримуючи узгодженість практик безпеки та прозорості. Запропоновані порівняльна матриця та узагальнений контур упровадження відповідального ШІ можуть бути використані як інструмент інтерпретації зрілості корпоративних моделей, а також як основа для адаптації кращих практик в організаціях, що впроваджують системи ШІ/МН в умовах посилення регуляторних і суспільних вимог.

References

1. Papagiannidis E., Mikalef P., Conboy K. Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems. 2025. Т. 34, № 2. С. 101885. URL: https://doi.org/10.1016/j.jsis.2024.101885
2. Ruster L. P., Daniell K. A. How to Operationalize Responsible Use of Artificial Intelligence. MIS Quarterly Executive. 2025. С. 185–202. URL: https://doi.org/10.17705/2msqe.00116
3. Model Cards for Model Reporting / M. Mitchell та ін. FAT* '19: Conference on Fairness, Accountability, and Transparency, м. Atlanta GA USA. New York, NY, USA, 2019. URL: https://doi.org/10.1145/3287560.3287596
4. Datasheets for datasets / T. Gebru та ін. Communications of the ACM. 2021. Т. 64, № 12. С. 86–92. URL: https://doi.org/10.1145/3458723
5. Exploring the impact of responsible AI governance on corporate performance: A quasi-natural experiment / H. Xia та ін. Technological Forecasting and Social Change. 2026. Т. 223. С. 124425. URL: https://doi.org/10.1016/j.techfore.2025.124425
6. Pichai S. AI at Google: our principles. Google. URL: https://blog.google/innovation-and-ai/products/ai-principles/#:~:text=We%20recognize%20that%20such%20powerful,will%20impact%20our%20business%20decisions
7. Responsible AI | Google Cloud. Google Cloud. URL: https://cloud.google.com/responsible-ai

8. Manyika J. Responsible AI: Our 2024 report and ongoing work. Google. URL: https://blog.google/innovation-and-ai/products/responsible-ai-2024-report-ongoing-work/#:~:text=Responsible%20AI:%20Our%202024%20report,throughout%20the%20AI%20development%20lifecycle
9. Responsible AI. Google Research - Explore Our Latest Research in Science and AI. URL: https://research.google/teams/responsible-ai/#:~:text=Responsible%20AI%20,them%20in%20the%20real%20world
10. Responsible AI. Amazon News. URL: https://www.aboutamazon.com/what-we-do/artificial-intelligence-ai/responsible-ai
11. Responsible AI Principles and Approach | Microsoft AI. Your request has been blocked. This could be due to several reasons URL: https://www.microsoft.com/en-us/ai/principles-and-approach#:~:text=Fairness
12. Amazon extends moratorium on police use of facial recognition software | Reuters. URL: https://www.reuters.com/technology/exclusive-amazon-extends-moratorium-police-use-facial-recognition-software-2021-05-18
13. Press release | Rome Call. Rome Call | What is the Matter with AI Ethics?. URL: https://www.romecall.org/press-release/
14. Trust and transparency | IBM. IBM. URL: https://www.ibm.com/policy/trust-transparency
15. Montgomery C. How our commitment to ethics, trust and transparency is differentiating IBM | IBM. IBM. URL: https://www.ibm.com/think/insights/how-our-commitment-to-ethics-trust-and-transparency-is-differentiating-ibm
16. IBM will no longer offer facial recognition technology | World Economic Forum. URL: https://www.weforum.org/stories/2020/06/ibm-facial-recognition-george-floyd/
17. https://humanrights.fb.com/wp-content/uploads/2023/09/2022-Meta-Human-Rights-Report.pdf
18. AI at Meta. URL: https://ai.meta.com/static-resource/responsible-use-guide/
19. Facebook. URL: https://www.facebook.com/AIatMeta/videos/stevie-bergman-and-jacqueline-pan-spoke-on-ai-advancements-and-the-many-difficul/514283333390689/
20. Rahmani I. Building Ethical AI: The Five Pillars of AI Ethics. LinkedIn: Log In or Sign Up. URL: https://www.linkedin.com/pulse/building-ethical-ai-five-pillars-ethics-ibrahim-ibby-rahmani-j1qdc
21. Introducing the AI Alliance. AI at Meta. URL: https://ai.meta.com/blog/ai-alliance/
22. https://about.fb.com/wp-content/uploads/2022/07/Meta_Human-Rights-Report-July-2022.pdf
23. Understanding AWS Responsible AI: Key Concepts and Dimensions. Tutorials Dojo. URL: https://tutorialsdojo.com/understanding-aws-responsible-ai-key-concepts-and-dimensions/
24. ACLU Statement on Extended Amazon Face Recognition Moratorium | American Civil Liberties Union. American Civil Liberties Union. URL: https://www.aclu.org/press-releases/aclu-statement-extended-amazon-face-recognition-moratorium
25. Staff A. We are implementing a one-year moratorium on police use of Rekognition. Amazon News. URL: https://www.aboutamazon.com/news/policy-news-views/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition
26. Introducing Appleâ□□s On-Device and Server Foundation Models. Apple Machine Learning Research. URL: https://machinelearning.apple.com/research/introducing-apple-foundation-models
27. Caswell A. Apple is prioritizing privacy over winning the AI race – here's why. Tom's Guide. URL: https://www.tomsguide.com/ai/apple-intelligence/apple-is-prioritizing-privacy-over-winning-the-ai-race-heres-why
28. Apple's Ethical AI: A Deep Dive into Responsible AI Powering Apple Intelligence. Learn Prompting: Your Guide to Communicating with AI. URL: https://learnprompting.org/blog/apple-intelligence-responsible-ai?srsltid=AfmBOoqePyu50dHlkNU2OMbqrrPxB_mrI7WUxSHuPtbqEUb9QdP1A7Ri
29. Is Apple intelligence trained on data illegaly. URL: https://appleinsider.com/articles/25/07/21/apple-insists-its-ai-training-is-ethical-and-respects-publishers
30. Csathy P. Apple's AI "Intelligence": safe, secure & "ethically sourced" ... or is it?. Peter Csathy's "the brAIn" - real media & AI intelligence | Substack. URL: https://themediabrain.substack.com/p/apples-ai-intelligence-safe-secure
31. What is the OpenAI Charter?. Milvus | High-Performance Vector Database Built for Scale. URL: https://milvus.io/ai-quick-reference/what-is-the-openai-charter
32. Planning for AGI and beyond | OpenAI. URL: Planning for AGI and beyond | OpenAI
33. How we think about safety and alignment | OpenAI. URL: How we think about safety and alignment | OpenAI
34. Facebook. URL: https://www.facebook.com/groups/1077715963163799/posts/1117919729143422/
35. 5 Principles for Responsible AI | SS&C Blue Prism. URL: https://www.blueprism.com/guides/ai/responsible-ai/
36. Meta Responsible AI Framework: Social Media-Focused Approach to AI Governance. Organic Growth Engineering | VerityAI. URL: https://verityai.co/blog/meta-responsible-ai-framework