

DOI: <https://doi.org/10.18372/2412-2157.42.21006>

УДК 17:004.8(045)

## РОЗПОДІЛЕНА АГЕНТНІСТЬ І РОЗПОДІЛЕНА ВІДПОВІДАЛЬНІСТЬ У СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ: ІНФОЕТИЧНИЙ АНАЛІЗ

Тетяна Шоріна

Державний університет «Київський авіаційний інститут»

<https://orcid.org/0000-0001-9281-7840>

**Анотація.** досліджена проблема розподіленої агентності та розподіленої відповідальності в системах штучного інтелекту в контексті сучасної інформаційної етики. Розглянуто класичні філософсько-етичні уявлення про моральну дію, агентність і відповідальність, а також виявлено обмеженість їх безпосереднього застосування до сучасних систем штучного інтелекту як складних соціотехнічних конфігурацій. Особливу увагу приділено інфоетичному підходу Л. Флоріді, а також поняттям штучної агентності, прогалини у відповідальності, підзвітності та управлінні. Уточнено їх значення для етичного аналізу штучного інтелекту.

**Ключові слова:** штучний інтелект, інфоетика, розподілена агентність, розподілена відповідальність, управління ШІ, підзвітність, інфосфера, соціотехнічні системи, надійність ШІ, етика технологій.

### Вступ

Стрімкий розвиток систем штучного інтелекту (ШІ), їхнє впровадження в комунікаційні, управлінські, освітні, медичні, безпекові та економічні практики істотно посилює увагу до етичних і соціальних наслідків цифрових технологій. Сьогодні системи ШІ дедалі активніше впливають на процеси прийняття рішень, організацію інформаційних потоків, умови доступу до знання, форми людської взаємодії та механізми суспільного контролю. З огляду на це актуалізуються питання про межі застосування ШІ, характер його дії та нормативні засади використання. Крім очевидно-прикладного, ці питання мають і світоглядно-філософське значення.

Складність цієї проблематики зумовлена багатовимірною природою штучного інтелекту. Його осмислення виходить далеко за межі суто технічного або інструментального підходу. Штучний інтелект дедалі частіше розглядають як соціотехнічний феномен, який перебуває на перетині філософії техніки, інформаційної етики, соціальної епістемології, права, регуляторної політики, медіадосліджень та філософської антропології. Дискусії навколо ШІ охоплюють широке коло питань: від автономії, контролю та прозорості до упередженості, справедливості, надійності, пояснюваності та підзвітності. У цьому контексті особливого значення набуває питання про те, як саме слід осмислювати дію систем ШІ і визначати відповідальність тоді, коли наслідки технологічних рішень зачіпають людину, соціальні інститути або публічний простір.

Проблематика агентності та відповідальності у сфері штучного інтелекту активно розробляється в сучасних дослідженнях представниками різних напрямів. Цю проблематику аналізують фахівці з інформаційної етики, філософії техніки, етики технологій, управління ШІ та публічної політики. Оскільки саме в межах західної дослідницької традиції було сформульовано ключові концептуальні підходи – від концепції прогалини у відповідальності (responsibility gap), запропонованої А. Маттіасом, до людиноцентричного підходу ЄС та теорії інформаційної етики Л. Флоріді, – особливої уваги заслуговують праці таких авторів, як А. Маттіаса, Д. Дж. Джонсон, М. Вердіккіо, П.-П. Вербека, К. Новеллі, М. Таддео, Я. Мьокандера та інших

дослідників, які в різний спосіб осмислюють співвідношення між технічною агентністю, людським контролем, нормативною оцінкою та інституційними механізмами відповідальності.

Перспективність розроблення цієї теми визначається не лише її теоретичною новизною, а й незаперечною суспільною значущістю. Осмислення дії та відповідальності в системах ШІ важливе для формування етичних стандартів проектування й використання технологій, для розроблення належних моделей регулювання, а також для захисту людини в умовах зростаючої ролі алгоритмічних систем у повсякденному житті. Тому проблема розподіленої агентності та розподіленої відповідальності є сьогодні однією з ключових у сучасному філософському аналізі штучного інтелекту.

### Мета та завдання дослідження

Мета статті – філософсько-етичний аналіз проблеми розподіленої агентності та розподіленої відповідальності в системах штучного інтелекту в контексті сучасної інформаційної етики.

Для досягнення поставленої мети передбачено виконання таких завдань: проаналізувати класичні філософсько-етичні уявлення про (моральну) дію, агентність і відповідальність, а також виявити межі суб'єктоцентричної моделі дії в контексті сучасних технологічних трансформацій; з'ясувати трансформацію розуміння агентності в умовах розвитку ШІ та соціотехнічних систем у сучасному філософському дискурсі; розкрити особливості дії систем ШІ як розподіленої соціотехнічної конфігурації та уточнити зміст поняття штучної агентності (artificial agency) в інфоетичному підході Л. Флоріді; дослідити теоретичний зміст проблеми прогалини у відповідальності й показати підстави переходу від індивідуалізованої моделі відповідальності до її розподілених і процедурних форм; окреслити значення понять підзвітності, управління та аудиторності для сучасного етичного аналізу систем ШІ і для осмислення розподіленої відповідальності.

### Методологія дослідження

Методологічну основу статті становить міждисциплінарний підхід, що поєднує інструменти інформаційної етики, філософії техніки, філософської антропології та сучасних досліджень соціотехнічних систем. Такий ракурс дає змогу розглядати штучний інтелект з перспективи феномену, що змінює

уявлення про дію, контроль, причинність і відповідальність у цифровому середовищі.

Провідним концептуальним орієнтиром дослідження є інфоетичний підхід, розроблений у працях Л. Флоріді, який дає змогу аналізувати ШІ в ширшому контексті інфосфери та виявляти етичні наслідки функціонування штучних агентів у середовищі людських і нелюдських взаємодій. Водночас у роботі використано соціально-технічний підхід, у межах якого системи ШІ розглядаються як багаторівневі конфігурації, де технічні, людські, організаційні та нормативні компоненти утворюють єдину структуру дії.

Основним методом дослідження є концептуальний аналіз, спрямований на уточнення змісту понять агентності, штучної агентності, відповідальності, розподіленої відповідальності та підзвітності в контексті етичного аналізу систем ШІ. З метою зіставлення різних підходів до проблеми застосовано порівняльний аналіз праць сучасних дослідників ШІ, зокрема у царині інформаційної етики, етики технологій та управління ШІ. Історико-філософська реконструкція використовується для виявлення меж класичної суб'єктоцентричної моделі відповідальності та з'ясування того, чому вона виявляється недостатньою для опису сучасних систем ШІ.

Окрім того, у статті застосовано елементи герменевтичного підходу до інтерпретації ключових теоретичних позицій, а також узагальнення як способу переходу від аналізу окремих концепцій до формулювання цілісного бачення розподіленої агентності та розподіленої відповідальності в системах ШІ.

### Результати дослідження

Однією з ключових передумов аналізу розподіленої відповідальності в системах штучного інтелекту є переосмислення самого поняття агентності. У класичній європейській етичній традиції (моральна) дія розглядалася як акт, що має визначеного суб'єкта, намір та наслідок. Така модель формується вже в межах аристотелівської практичної філософії, де моральна оцінка безпосередньо пов'язується з добровільністю й самохітністю дії. Як зазначає Арістотель у Нікомаховій етиці: «Якщо несамохітне здійснюється підневільно і по невіданню, то самохітним мало б вважатися те, джерело чого – у самому діячі, свідомому тих обставин, за яких вчинок має місце» (Арістотель 2002, III. 3). Хвалити чи засуджувати вчинки людини можна лише в разі їх самохітного характеру, коли сама людина є продуктивним джерелом своїх діянь. Для того, аби відбулася дія такого роду, має здійснитися розумовий акт, вибір або рішення, що тягне за собою дію. Як міркує далі Арістотель: «Услід за розмежуванням самохітного і несамохітного йде виклад питання про свідомий вибір (περί προαίρεσως), [тобто про навмисність], адже він найтіснішим чином пов'язаний з доброчесністю (οἰκειότατον εἶναι τῆ ἀρετῆ) і ще більшою мірою, ніж вчинки, дає можливість судити про вдачі (τα ἠθῆ κριτεῖν)» (Арістотель 2002, III. 4 (II)). Моральна дія, як бачимо, у Арістотеля має складну природу, вона об'єднує свідомий вибір (prohairesis) агента щодо вчинку на підставі розуму і мислення, та сам вчинок

як реалізацію прийнятого рішення. Звідси дія (praxis) у класичній практичній філософії мислиться як результат раціонального рішення агента, який усвідомлює підстави своєї поведінки.

У моральній філософії Модерну ця структура отримує нормативне завершення в етиці І. Канта. Моральна дія розглядається як результат автономної волі раціонального суб'єкта. У трактаті «Основи метафізики моралі» І. Кант формулює принцип автономії волі: автономія волі – це властивість волі, завдяки якій вона є законом сама собі (незалежно від будь-якої властивості об'єктів воління). ... воля кожної розумної істоти необхідно пов'язана з нею як умова» (Kant 1998, 4:440, 47). У цьому контексті воля – це особливий тип причинності, властивий раціональним істотам: «Воля є різновидом причинності живих істот тією мірою, якою вони є раціональними, а свобода була б тією властивістю такої причинності, що вона може бути ефективною незалежно від сторонніх причин, що її визначають» (Kant 1998, 4:446, 52). У І. Канта поняття свободи є ключем до пояснення автономії волі. Він далі пояснює: «кожна істота, яка не може діяти інакше, ніж під ідеєю свободи, є саме завдяки цьому справді вільною в практичному відношенні» (Kant 1998, 4:448, 53). Відповідальність у цій моделі має персональний і інтенціональний характер. Вона приписується суб'єктові, здатному діяти відповідно до принципу, який він може визнати універсальним законом.

Навіть у більш пізній етиці відповідальності Г. Йонаса, де береться до уваги безпрецедентне зростання технічної могутності людства, зберігається суб'єктоцентрична структура моральної дії. Г. Йонас переформулює кантіанський імператив у відповідь на новий тип діяльності людини й звернення до нового типу дій, якими вони керуються. Його новий імператив для технологічної цивілізації репрезентований так: «Дій так, щоб наслідки твоєї дії були сумісні з триванням справжнього людського життя на Землі» (Jonas 1984, 11). Хоча ця концепція розширює горизонт відповідальності до майбутніх поколінь і глобальних технологічних наслідків, вона все ж зберігає фундаментальне припущення класичної етики, оскільки моральна дія передбачає визначеного агента, який є носієм наміру, здатен контролювати засоби дії та несе відповідальність за її наслідки.

Саме ця модель – дія як результат наміру автономного суб'єкта – становить теоретичне підґрунтя класичного розуміння відповідальності.

Розвиток сучасних технологій поступово ускладнює класичну суб'єктоцентричну модель дії. В етиці відповідальності Г. Йонас вже визнавав, що сучасні технології «запровадили дії такого масштабу, об'єкти та наслідки, що рамки колишньої етики більше не можуть їх вмістити» (Jonas 1984, 6). Тобто у нього техніка вже розглядається як чинник, що радикально розширює наслідки людських дій і відтак вимагає нової етики відповідальності. Сучасна філософія техніки іде далі й переносить увагу на те, у який спосіб технології беруть участь у формуванні самої структури людської діяльності.

У концепції технологічної медіації, розробленої П.-П. Вербеком у межах постфеноменологічної

традиції, доводиться, що технічні артефакти не бувають ціннісно нейтральними, вони не лише інструменти волі суб'єкта. У технологій є своє моральне значення і сенс, й взаємодія людини з технологією може призводити до суперечливої моральної оцінки: «Коли ми розробляємо технології на основі конкретних систем цінностей, ми ще не знаємо їхніх соціальних наслідків, але, щойно ми дізнаємося про ці наслідки, технології, можливо, вже змінять системи цінностей для оцінки цих наслідків» (Kudina and Verbeek 2019). Саме на цій активній стороні технологічного опосередковування наголошує П.-П. Вербек. Якщо технології допомагають формувати людські дії та те, як ми живемо, тоді технології «активно» беруть участь в етиці». Водночас філософ застерігає не плутати цей феномен морального опосередкування з формою моральної суб'єктності. Він не стверджує, що технології володіють моральною агентністю, його підхід технологічного опосередковування (медіації) стверджує, що моральна агентність – це гібридний процес, що передбачає як людей, так і технології (Verbeek 2014).

За таких змін у структурі взаємодії людини й технологій дія вже не може бути описана як лінійна схема «суб'єкт – засіб – результат». Технологічні артефакти змінюють поле можливих рішень, структурують сприйняття ситуації та опосередковують сам процес вибору. Так, наводить приклади П.-П. Вербек, технології ультразвукової візуалізації опосередковують моральні питання та рішення щодо абортів, дрони опосередковують моральний досвід солдатів, а смартфони – етикет поведінки в ресторані та класі. (Kudina & Verbeek 2019). Водночас у межах концепції технологічної медіації (Verbeek 2016) техніка все ще не розглядається як самостійний моральний агент. Вона виступає радше співучасником людської дії, який змінює її умови, але не замінює суб'єкта як носія інтенції та відповідальності.

Утім, із появою систем ШІ ситуація ускладнюється ще більше, бо сучасні системи штучного інтелекту демонструють інший тип дієвості. Вони здатні автономно функціонувати в межах заданих параметрів, генерувати нові інформаційні об'єкти та взаємодіяти з іншими елементами інформаційного середовища без безпосереднього людського супроводу на кожному етапі. Дія в таких умовах дедалі важче описується як наслідок наміру одного визначеного суб'єкта.

У межах інфоетичного підходу Л. Флоріді ця трансформація осмислюється через поняття штучної агентності. Ще в ранній роботі, присвяченій моральності штучних агентів, Л. Флоріді та Дж. Сандерс зазначали, що штучні агенти «розширюють клас сутностей, які можуть бути залучені до моральних ситуацій», їх можна розглядати як моральних агентів, тобто як сутностей, що здатні «здійснювати дію задля добра чи зла» (Floridi and Sanders, 2004). Ідеться не про ототожнення штучних систем із людськими моральними суб'єктами, а про визнання того, що вони здатні бути дієвими елементами інфосфери, тобто здатні здійснювати операції, які мають

реальні наслідки для інформаційного середовища та для людей, залучених до нього. Звідси Л. Флоріді пропонує відмовитися від антропоморфного тлумачення ШІ. У пізнішій розробці цієї теми він пише, що штучний інтелект доцільніше розуміти не як розширення поняття інтелекту, а як розширення поняття агентності, залучаючи в нього штучні форми, «які не вимагають пізнання, інтелекту, намірів або психічних станів» (Floridi 2025). У такій перспективі ШІ розглядається не як людський розум в іншому носії, а як специфічна форма ефективно, але неінтенційальної дієвості.

Цю думку Л. Флоріді формулює ще стисліше у своїй етиці штучного інтелекту: «Штучний інтелект краще розуміти як нову форму самостійної дії, а не як інтелект»; він став можливим завдяки «роз'єднанню агентивності та інтелекту» (Floridi 2023). Із цього й слідує принципова новизна сучасних систем ШІ. Вони можуть бути операційно успішними, не будучи втім ні свідомими, ні морально автономними у класичному сенсі. Унаслідок цього дія дедалі менше постає як локалізований акт окремого суб'єкта і дедалі більше – як процесуальна та розосереджена конфігурація, що вимагає нового понятійного апарату для її етичного осмислення.

Отже, поява систем ШІ має наслідком зміну самої структури морально значущої дії. Класична модель, у якій інтенція безпосередньо пов'язується з дією, а дія – з персональною відповідальністю, тепер поступається місцем іншій конфігурації: соціотехнічна система → розподілена дія → розподілена відповідальність. Розподілена дія (distributed action) – це процес, що виникає не в одному центрі волі, а в полі взаємодії багатьох елементів – людей, моделей, даних, інституцій, процедур та інфраструктур. Сучасні дослідження ШІ дедалі частіше зміщують увагу від пошуку автора дії до аналізу системи, у межах якої ця дія стає можливою. Якщо дія формується в межах багаторівневої соціотехнічної конфігурації, то і відповідальність повинна бути розподіленою між тими рівнями та учасниками, які беруть участь у проєктуванні, впровадженні та використанні системи.

Найбільш виразно розподілений характер дії поданий у великих мовних моделях (LLM). У статті Я. Мьокандера, Й. Шюетта, Г. Р. Кірк і Л. Флоріді запропоновано трирівневу модель аудиту LLM, що охоплює рівень управління, рівень моделі та прикладний рівень. Цінність цієї моделі – в її практичній придатності для аудиту, а також у тому, що вона робить аналітично видимою багаторівневу структуру дії в системах такого типу. Кінцевий результат, який продукує LLM, не може бути адекватно пояснений на основі лише одного рівня, оскільки на нього впливають і організаційні рішення розробника, і архітектура та передтренування моделі, і конкретні способи її інтеграції в прикладні сценарії. Тому LLM не можна вважати автономним моральним суб'єктом. Радше вона постає як певний «вузол» у розподіленій соціотехнічній конфігурації, де причинність, контроль і наслідки функціонально розосереджені. Звідси випливає, що питання про відповідальність має уникати пошуку єдиного винуватця, натомість передбачати аналіз ролей,

рівнів і процедур, які брали участь у продукуванні морально значущого результату.

У цьому контексті по-новому постає й класична проблема прогаліни у відповідальності. Її вихідне формулювання належить А. Маттіасу, який у статті 2004 року визначив, що стосовно навчальних автоматів (learning automata) виникають ситуації, коли традиційні способи приписування відповідальності перестають відповідати нашому уявленню про справедливість, оскільки ніхто не має достатнього контролю над конкретною дією машини, щоб повною мірою взяти її на себе (Matthias 2004). Відтоді поняття прогаліни у відповідальності набуло статусу одного з ключових діагностичних понять сучасної етики ШІ.

Для сучасних систем ШІ ця проблема не втратила значення, однак сьогодні вона дедалі менше згадується як підстава для зняття відповідальності загалом. Радше вона вказує на необхідність переходу від вузько персоналістської моделі до процедурного розуміння відповідальності. Тут необхідно розрізнати кілька нормативних рівнів аналізу. Моральна відповідальність у класичному сенсі пов'язана з приписуванням провини або заслуги суб'єктові дії. Натомість підзвітність означає обов'язок пояснення, обґрунтування та документування рішень і процедур, у межах яких функціонує система. Управління позначає інституційні рамки розподілу ролей, повноважень і механізмів контролю, тоді як аудиторність стосується можливості перевірки системи, її даних, рішень і процесів. Отже, в системах ШІ підзвітність не скасовує моральної відповідальності, але переводить її в ширший процедурний та інституційний контекст.

У спільній статті К. Новеллі, М. Таддео та Л. Флоріді (Novelli, Taddeo, and Floridi 2024) відповідальність описується через поняття підзвітності, яке означає не моральну провину як таку, а обов'язок пояснити, обґрунтувати й документувати прийняті рішення. У межах класичної етики відповідальність має переважно ретроспективний характер – вона пов'язана з оцінкою вже здійсненої дії. У системах ШІ відповідальність набуває переважно проспективного характеру, тому що націлена на запобігання шкоді, управління ризиками та забезпечення надійності системи. Вважається, що з організацією цих умов функціонування ШІ може бути підконтрольним і нормативно впорядкованим. Власне, ідеться про застосування підходу так званої «процедурної відповідальності» (procedural responsibility). Цей підхід змінює фокус морально значущої дії. Етичне питання більше не зводиться до того, хто саме «є винним», а переноситься на те, чи були у належний спосіб організовані умови проектування, впровадження, моніторингу й використання ШІ. Інакше кажучи, відповідальність пов'язується здебільшого не з одноразовим моральним актом індивіда, а з якістю соціотехнічної екосистеми, у межах якої діє цей індивід (Novelli, Taddeo, and Floridi 2024).

Зміщення відповідальності від індивідуальної провини до організації умов дії знайшло інституційне оформлення в європейському дискурсі ШІ. Групою

експертів високого рівня зі штучного інтелекту (AI HLEG) при Європейській Комісії був розроблений відповідний документ – «Етичні рекомендації для надійного штучного інтелекту» (European Commission 2019), у якому запропонували людиноцентричну концепцію надійності ШІ, а також сформулювали сім ключових вимог до надійних систем ШІ. У Рекомендаціях підзвітність визначається як одна з семи ключових вимог, поряд із людським наглядом і дією (human agency and oversight); технічною надійністю та безпекою (technical robustness and safety); прозорістю (transparency); недискримінацією та справедливістю (diversity, non-discrimination and fairness); приватністю та управлінням даними (privacy and data governance), а також суспільним та екологічним добробутом (societal and environmental well-being). Згідно з цими рекомендаціями, будь-яка система ШІ, щоб вважатися надійною, має відповідати таким трьом ключовим критеріям, як законність – дотримання всіх чинних законів і норм; етичність – дотримання етичних принципів і цінностей та надійність – стійкість як у технічному, так і в соціальному вимірах, зокрема й у непередбачуваних ситуаціях. Зазначимо, що вимога підзвітності в Рекомендаціях прямо пов'язується з аудиторністю, оцінкою алгоритмів, даних і процесів проектування, а також із забезпеченням належних механізмів відшкодування. Отже, ідеться знову про інституційно організовану підзвітність упродовж усього життєвого циклу системи ШІ, з одного боку, і про нормативно-проектований характер відповідальності, з іншого боку. Завдяки цьому ШІ може функціонувати без руйнування довіри, справедливості та можливості людського контролю. Принагідно додати, що у середині 2020-х років зазначений етичний підхід дістав подальшу правову та інституційну конкретизацію в регуляторній політиці Європейського Союзу, насамперед у межах AI Act (Regulation (EU) 2024/1689). Рекомендації AI HLEG окреслювали нормативний горизонт надійного ШІ переважно в етико-політичних термінах, а в AI Act частина цих вимог переводиться у площину формалізованих обов'язків, процедур відповідності, розподілу ролей між провайдерами та користувачами, а також управління ризиками. Це засвідчує, що сучасне осмислення відповідальності щодо ШІ дедалі більше рухається від загального морального заклику до інституційно опосередкованих і перевірюваних механізмів підзвітності.

Зазначений проспективний і процедурний характер відповідальності особливо виразно проявляється тоді, коли до проблеми розподіленої дії додається ще одна обставина – епістемічна обмеженість сучасних систем ШІ. Річ у тім, що наслідки дії, крім проблеми розподіленості між багатьма рівнями системи, не можуть також бути з достатньою повнотою передбачені. Через це відповідальність неминуче набуває форми управління ризиком. Цей аспект стає центральним у подальших роботах Л. Флоріді та експертного кола, коли вони звертаються до питань визначеності, широти охоплення (score) та меж передбачуваності ШІ. Зокрема, у статті «Гіпотеза про фундаментальний компроміс між визначеністю та широтою в

символьному та генеративному штучному інтелекті» (Floridi 2025) філософ обґрунтовує тезу про фундаментальний компроміс (trade-off) між широтою функціонального охоплення системи та можливістю досягти майже повної достовірності її результатів.

Зміст тези у такому: що ширший клас завдань і багатший простір даних, із якими працює ШІ, то менш реалістичними стають очікування безпомилковості; натомість системи, які наближаються до дедуктивно гарантованої надійності, мусять діяти в значно вужчому, попередньо структурованому середовищі. Для генеративного ШІ це означає, що його застосування завжди пов'язане з незвідним епістемічним ризиком.

Для етики відповідальності цей висновок має принципове значення. Якщо в певному класі систем ШІ, особливо у генеративному ШІ та LLM, повний контроль над окремим результатом принципово недосяжний, то відповідальність уже не може визначитися як гарантія безпомилковості. Вона має бути переосмислена як відповідальність за проектування умов, у яких ризики функціонування системи стають зрозумілими, керованими та мінімізованими. В сучасному дискурсі управління ШІ відповідальність стає здебільшого вмінням передбачити можливі зони збою, встановлювати межі застосування системи, визначити механізми людського нагляду, верифікації та корекції рішень. Отже, відповідальність дедалі більше набуває форми управління ризиком, а самі управління та ризик-менеджмент стають не зовнішнім додатком до відповідальності, а її сучасною формою в умовах розподіленої дії (Floridi 2025).

### Обговорення

Отримані результати загалом узгоджуються із сучасним філософсько-етичним дискурсом про штучний інтелект, у межах якого класична модель моральної дії, заснована на зв'язку між інтенцією індивіда, його дією та персональною відповідальністю, стає обмеженою для аналізу складних соціотехнічних систем. У цьому сенсі дослідження збігається з підходом Л. Флоріді, який інтерпретує ШІ як специфічну форму штучної агентності, а також із позицією А. Маттіаса щодо проблеми «прогалини у відповідальності». Водночас запропонований у нашому дослідженні підхід дає змогу уточнити, що сама «прогалина у відповідальності» не є доказом неможливості відповідальності щодо ШІ, вона радше є показником обмеженості моделі, що шукає єдиного суб'єкта дії там, де сама дія розподілена.

На відміну від підходів, у яких відповідальність мислиться переважно як ретроспективне приписування провини, у статті обґрунтовується, що для систем ШІ визначальними стають проспективні й процедурні форми відповідальності, реалізовані через підзвітність, управління, аудитуваність і механізми контролю ризику. У цьому аспекті дослідження не ототожнює штучну агентність із людською моральною суб'єктністю і не приписує системам ШІ автономного морального статусу в сильному сенсі. Наголошуємо на тому, що ШІ бере участь у структурі дії як дієвий елемент соціотехнічної конфігурації, а тому вимагає

відповідного переосмислення того, як розподіляються контроль, причинність і нормативне навантаження.

Разом із тим запропонований підхід має і свої межі. Він не усуває потреби у персональній відповідальності в тих випадках, де конкретні дії розробників, замовників, операторів або користувачів можуть бути достатньо чітко ідентифіковані. Крім того, сама ідея розподіленої відповідальності потребує подальшого уточнення на перетині етики, права та регуляторної політики, оскільки нормативні режими відповідальності, підзвітності та відшкодування шкоди не є повністю тотожними. Проте саме такий підхід дає змогу точніше описати етичну специфіку сучасних систем ШІ, ніж моделі, що зводять проблему або лише до провини окремого індивіда, або лише до абстрактної технічної несправності.

### Висновки

Поява й поширення систем штучного інтелекту, крім нової технічної реальності, зумовлює зміну структури моральної дії в цифрову епоху. У її межах переосмислюються підстави відповідальності, довіри, автономії та людської участі у світі технологічно опосередкованих рішень. Класична етика виходила з відносної цілісності суб'єкта, прозорості наміру та локалізованості вчинку, а в умовах ШІ морально значуща дія частіше стає ефектом розподілених соціотехнічних процесів, де людське, машинне й інституційне вже не існують як цілком відокремлені площини. Через це трансформується і саме розуміння відповідальності, яка вже не вичерпується приписуванням провини після факту, а переважно пов'язується з етичним упорядкуванням тих умов, у яких стають можливими рішення, оцінки й наслідки. У такій перспективі довіра до ШІ також перестає бути просто психологічною готовністю покладатися на систему або визнанням її ефективності; вона стає питанням морально й нормативно обґрунтованого ладу цифрового середовища. Сьогодні філософський аналіз штучного інтелекту збігається водночас з аналізом того, як змінюється сама конфігурація морального досвіду в культурі, де межі між дією, посередництвом, контролем і відповідальністю стають рухомішими.

### Список літератури

1. Арістотель. Нікомахова етика ; пер. з давньогрец. Київ : Аквілон-Плюс, 2002. 480 с.
2. Kant I. *Groundwork of the Metaphysics of Morals* / trans. and ed. by M. Gregor ; introd. by C. M. Korsgaard. Cambridge : Cambridge University Press, 1998.
3. Jonas H. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago : University of Chicago Press, 1984.
4. Kudina O., Verbeek P.-P. Ethics from Within: Google Glass, the Collingridge Dilemma, and the Mediated Value of Privacy. *Science, Technology, & Human Values*. 2019. Vol. 44, no. 2. P. 291–314. DOI: <https://doi.org/10.1177/0162243918793711>.
5. Verbeek P.-P. Some Misunderstandings about the Moral Significance of Technology. *The Moral Status of Technical Artefacts* / ed. by P. Kroes, P.-P. Verbeek. Dordrecht : Springer, 2014. P. 75–88. DOI: [https://doi.org/10.1007/978-94-007-7914-3\\_5](https://doi.org/10.1007/978-94-007-7914-3_5).
6. Verbeek P.-P. *Toward a Theory of Technological Mediation: A Program for Postphenomenological Research*. -

*Technoscience and Postphenomenology: The Manhattan Papers* : collective monograph. Lexington, 2016. P. 189–204. DOI: <https://doi.org/10.5040/9781978731929.ch-0013>.

7. Floridi L., Sanders J. W. On the Morality of Artificial Agents. *Minds and Machines*. 2004. Vol. 14, no. 3. P. 349–379. DOI: <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>

8. Floridi L. AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. *Philosophy & Technology*. 2025. Vol. 38, art. № 30. DOI: <https://doi.org/10.1007/s13347-025-00858-9>.

9. Floridi L. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford : Oxford University Press, 2023. 272 p. DOI: 10.1093/oso/9780198883098.001.0001.

10. Mökander J., Schuett J., Kirk H., Floridi L. Auditing Large Language Models: A Three-Layered Approach. *SSRN Electronic Journal*. 2023. DOI: <https://doi.org/10.2139/ssrn.4361607>.

11. Matthias, Andreas. 2004. The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology* 6 (3): 175–183.

12. Novelli C., Taddeo M., Floridi L. Accountability in artificial intelligence: what it is and how it works. *AI & Society*. 2024. Vol. 39. P. 1871–1882. DOI: <https://doi.org/10.1007/s00146-023-01635-y>.

13. European Commission. *Ethics Guidelines for Trustworthy AI*. Brussels : European Commission, 2019. URL: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

14. Floridi L. A Conjecture on a Fundamental Trade-Off Between Certainty and Scope in Symbolic and Generative AI. *Philosophy & Technology*. 2025. Vol. 38. Art. 93. DOI: 10.1007/s13347-025-00927-z.

15. Díaz-Rodríguez N., Del Ser J., Coeckelbergh M., López de Prado M., Herrera-Viedma E., Herrera F. Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*. 2023. Vol. 99. DOI: <https://doi.org/10.1016/j.inffus.2023.101896>.

16. Johnson D. G., Verdicchio M. Reframing AI Discourse. *Minds and Machines*. 2017. Vol. 27, no. 4. P. 575–590.

## References

1. Aristotle. 2002. *Nikomakhova etyka* [Nicomachean Ethics]. Kyiv: Akvilon-Plus.

2. Kant, Immanuel. 1998. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Introduction by Christine M. Korsgaard. Cambridge: Cambridge University Press.

3. Jonas, Hans. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press.

4. Kudina, Olya, and Peter-Paul Verbeek. 2019. "Ethics from Within: Google Glass, the Collingridge Dilemma, and the

Дата першого надходження: 02.02.2026.

Дата прийняття до друку: 25.02.2026.

Дата публікації: 21.03.2026

Tetiana Shorina

## DISTRIBUTED AGENCY AND DISTRIBUTED RESPONSIBILITY IN AI SYSTEMS: AN INFOETHICAL ANALYSIS

**Introduction.** The article examines distributed agency and distributed responsibility in artificial intelligence systems within contemporary information ethics. AI is considered as a socio-technical phenomenon that reshapes the philosophical understanding of action, responsibility, and human participation in technologically mediated environments. **The aim and tasks.** The article aims to provide a philosophical and ethical analysis of distributed agency and distributed responsibility in AI systems. The study revisits classical views of moral action, agency, and responsibility; examines the transformation of agency in the context of AI; clarifies the notion of artificial agency in Luciano Floridi's information ethics; analyzes the responsibility gap; and outlines the significance of accountability, governance, and auditability for the ethical assessment of AI. **Research methods.** The study is based on an interdisciplinary methodology combining conceptual analysis, information ethics, philosophy of technology, philosophical anthropology, and socio-technical systems analysis. It also employs historical-philosophical reconstruction, comparative analysis, hermeneutic interpretation, and generalization. **Research results.** The article shows that the classical subject-centered model of moral action is insufficient for analyzing contemporary AI systems. In AI environments, action emerges through the interaction of human, artificial, institutional, and infrastructural components. AI is therefore interpreted not as a moral subject in the human sense, but as a form of artificial agency, while

Mediated Value of Privacy." *Science, Technology, & Human Values* 44 (2): 291–314. <https://doi.org/10.1177/0162243918793711>.

5. Verbeek, Peter-Paul. 2014. "Some Misunderstandings about the Moral Significance of Technology." In *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek, 75–88. Dordrecht: Springer. [https://doi.org/10.1007/978-94-007-7914-3\\_5](https://doi.org/10.1007/978-94-007-7914-3_5).

6. Verbeek, Peter-Paul. 2016. "Toward a Theory of Technological Mediation: A Program for Postphenomenological Research." In *Technoscience and Postphenomenology: The Manhattan Papers*, 189–204. London: Lexington

7. Books. <https://doi.org/10.5040/9781978731929.ch-0013>.

8. Floridi, Luciano, and Jeff W. Sanders. 2004. "On the Morality of Artificial Agents." *Minds and Machines* 14 (3): 349–79. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>.

9. Floridi, Luciano. 2025. "AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis." *Philosophy & Technology* 38 (30). <https://doi.org/10.1007/s13347-025-00858-9>.

10. Floridi, Luciano. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198883098.001.0001>.

11. Mökander, Jakob, Jonas Schuett, Hannah Kirk, and Luciano Floridi. 2023. "Auditing Large Language Models: A Three-Layered Approach." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4361607>.

12. Matthias, Andreas. 2004. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology* 6 (3): 175–183.

13. Novelli, Claudio, Mariarosaria Taddeo, and Luciano Floridi. 2024. "Accountability in Artificial Intelligence: What It Is and How It Works." *AI & Society* 39: 1871–82. <https://doi.org/10.1007/s00146-023-01635-y>.

14. European Commission. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

15. Floridi, Luciano. 2025. "A Conjecture on a Fundamental Trade-Off Between Certainty and Scope in Symbolic and Generative AI." *Philosophy & Technology* 38: 93. <https://doi.org/10.1007/s13347-025-00927-z>.

16. Díaz-Rodríguez, Natalia, Javier Del Ser, Mark Coeckelbergh, Marcos López de Prado, Enrique Herrera-Viedma, and Francisco Herrera. 2023. "Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation." *Information Fusion* 99: 101896. <https://doi.org/10.1016/j.inffus.2023.101896>.

17. Johnson, Deborah G., and Mario Verdicchio. 2017. "Reframing AI Discourse." *Minds and Machines* 27 (4): 575–90.

responsibility acquires a distributed and procedural character. **Discussion.** The discussion supports the view that the ethical problem of AI should be framed not only in terms of individual blame, but also in terms of accountability, traceability, governance, and socio-technical conditions of action. **Conclusions.** The spread of AI transforms the normative conditions under which agency, responsibility, and trust are understood. Responsibility is no longer reducible to retrospective blame alone; it increasingly depends on the ethical and institutional organization of the environments in which AI systems are designed, deployed, monitored, and used.

**Keywords:** artificial intelligence, information ethics, distributed agency, distributed responsibility, AI governance, accountability, infosphere, socio-technical systems, trustworthy AI, ethics of technology.

DOI: <https://doi.org/10.18372/2412-2157.42.21008>

УДК 130.123.1:37.014.5-048.35(045)

## МОДЕРНІЗАЦІЯ УКРАЇНСЬКОЇ ОСВІТИ В УМОВАХ ГЛОБАЛЬНИХ ВИКЛИКІВ І ЛОКАЛЬНОЇ КУЛЬТУРНОЇ ВКОРИНЕНОСТІ

Оксана ЯКИМЧУК

Український державний університет імені Михайла Драгоманова

<https://orcid.org/0000-0002-9729-8389>

**Анотація.** Здійснено філософський аналіз модернізації української освіти в умовах глобалізаційних трансформацій та культурної вкоріненості. Освіта інтерпретується як культурно-онтологічний простір, у якому взаємодія глобальних освітніх тенденцій і локального контексту набуває характеру взаємного перетворення. Особливу увагу приділено герменевтичному принципу «злиття горизонтів» і застосуванню понять «нестабільності» та «точки біфуркації» як аналітичних засобів осмислення трансформаційних процесів в освіті. Показано, що сучасна українська освіта перебуває у стані онтологічної нестабільності, який поєднує в собі як ризик втрати культурної специфіки, так і потенціал формування нової освітньої парадигми. Обґрунтовано, що модернізація української освіти має відбуватися через творче поєднання універсальних тенденцій із національною культурною традицією.

**Ключові слова:** глобальне, локальне, точка біфуркації, «злиття горизонтів», невизначеність, модернізація освіти, діалектика глобального та локального, філософія освіти.

### Вступ

Діалектика глобального та локального у стратегіях модернізації української освіти є філософським викликом, що торкається самої природи освітнього буття. Освіта ніколи не існує в порожнечі, оскільки завжди вкорінена в культурі народу, його мові, історичній пам'яті, травматичному досвіді та уявленнях про майбутнє.

Сучасний світ формується під впливом інтенсивних глобальних процесів, пов'язаних із циркуляцією знань, технологій, нормативних моделей та цифрових інфраструктур, що неминуче позначаються і на локальному освітньому просторі. Глобальне, зорієнтоване на універсалістську логіку раціональності, повсякчас прагне стандартизації. Натомість локальне постає як жива традиція, вкорінена в культурній пам'яті народу та сформована його історичним і педагогічним досвідом. Діалектичний зв'язок між глобальним та локальним не зводимий до протиставлення «свого» і «чужого», в ньому зосереджено процес взаємного перетворення, у якому кожен полюс змінюється під впливом іншого.

Модернізація освіти в Україні відбувається в умовах історичної напруги, коли прагнення інтегруватися у європейський академічний простір співіснує з потребою зберегти культурну суб'єктність, обстояти її та виробити власну освітню стратегію. Варто визнати, що глобальне постає викликом, що спонукає до оновлення освітніх практик і стратегій, тоді як локальне зберігає власний смисловий і культурний потенціал, не дозволяючи модернізації перетворитися на механічне запозичення та неефективне копіювання.

Саме в цій точці філософія освіти покликана осмислити цю напругу як продуктивну, здатну в зіткненні універсального й локального віднайти нову якість освітнього простору. У цьому контексті модернізація постає як діалог із традицією, у межах

якого майбутнє освіти формується через творче поєднання інноваційних орієнтирів із національно-культурними засадами у відповідь на сучасні виклики.

### Мета та завдання дослідження

Мета статті – дослідити основні аспекти модернізації української освіти в умовах глобальних викликів і локальної культурної вкоріненості.

Відповідно до поставленої мети основними завданнями статті є: обґрунтування концептуальних засад та стратегічних орієнтирів модернізації української освіти, що забезпечуються творчим поєднанням універсальних глобальних тенденцій з національним культурним контекстом без втрати культурної самобутності; дослідження феномену локальної культурної вкоріненості української освіти та її роль у збереженні національної специфіки.

### Методологія дослідження

Методологія статті базується на комплексному філософсько-освітньому підході, що поєднує діалектичний метод Г. Гегеля (Гегель 2004), герменевтичний аналіз Г. Гадамера (Гадамер 2000) та теорію нелінійної динаміки І. Пригожина та І. Стенгерса (Prigogine, Stengers 1984). Діалектичний метод дає змогу розглядати глобальне та локальне як взаємопов'язані моменти становлення освіти, у межах цього процесу конфлікти, кризи та компромісні рішення постають чинниками формування нової якості. Він забезпечує розуміння «діалектичного неспокою» як рушійної сили розвитку (Гегель 2004, 151), завдяки якій освіта переходить від механічного запозичення стандартів до творчого перетворення власного простору.

Герменевтичний підхід, зокрема уявлення про «злиття горизонтів» дає інструменти для аналізу взаємного збагачення глобального та локального (національного) горизонтів смислу, вказуючи на те, як