Digital Transformation as a Philosophical Phenomenon: Reinterpreting Ethics and Aesthetics in the Context of Artificial Intelligence Dominance

UDC: 001.8:17.022.2 DOI: https://doi.org/10.15421/172616

Mnozhynska Ruslana [1]
Ph.D., Assoc. Prof., https://orcid.org/0000-0001-8459-3496, ruslanamnozhynska@gmail.com
Dubovyk Nataliia [2]
Ph.D., Assoc. Prof., https://orcid.org/0000-0003-0151-9480, ndubovyk@bel.ua
Kravtsov Yuriy [3]
Dr.Sc., Full Prof., https://orcid.org/0000-0003-4968-9112, Kravtsovyuri0@gmail.com
[1]Kyiv National University of Technology and Design (Kyiv, Ukraine),
[2]State University of Information and Communication Technologies (Kyiv, Ukraine)
[3]Dniprovsky State Technical University (Dnipro, Ukraine)

Abstract.
Relevance. Artificial intelligence has ceased to be just a computational tool and increasingly functions as a cultural and normative environment that forms modes of visibility, distributes attention and influences ethical judgments. It has been found that in such environments, the form of information presentation – the interface, the language of prompts, the composition of the screen – is not a «shell» but acts as an infrastructure of trust that strengthens or weakens responsibility, transparency and fairness.
The aim of the study is to conceptually and procedurally integrate ethical and aesthetic dimensions in the life cycle of AI systems, in particular the rationale for the E²-Framework (Ethical–Aesthetic Integration) assessment tool, which translates the principles of responsibility, accountability, and fairness into verifiable design, testing, and operation steps, while at the same time setting requirements for the form of uncertainty presentation, synthetics, and risk semiotics.
Results. Ten "nodes" of procedural suitability of AI ethics are characterized: data hygiene and relevance of samples; multivariate assessment of accuracy and fairness; full-fledged model/system cards and validation logs; interface as a carrier of explainability; real human-in-the-loop; independent audit and red-teaming; portioned consent and risk communication; post-marketing monitoring and rollback thresholds; energy savings; cultural localization of aesthetics. The impact of the integration of ethical and aesthetic criteria is evaluated on behavioral metrics: the share of source verification increases, the frequency of errors after the second step of consent decreases, and the recognition of generalization boundaries increases.
Conclusions. It has been proven that responsible technology is not reduced to model accuracy or "beautiful" design alone; it is crucial to synchronize them in a single architecture of responsibility. The proposed E²-Framework forms a culture of development in which each pixel has normative weight, and each ethical principle finds a formal embodiment in the language, rhythm and iconography of the interface.
Keywords: Digital Ethics, Digital Aesthetics, Artificial Intelligence, Explainability, Interface and Trust, Uncertainty, E²-Framework

Цифрова трансформація як філософський феномен: переосмислення етики та естетики в умовах домінування штучного інтелекту

Множинська Руслана[1], Дубовик Наталія[2], Кравцов Юрій[3]
[1]Київський національний університет технологій та дизайну (Київ, Україна)
[2]Державний університет інформаційно-комунікаційних технологій (Київ, Україна)
[3]Дніпровський державний технічний університет (Дніпро, Україна)

Анотація
Актуальність. Штучний інтелект перестав бути лише інструментом обчислення і дедалі більше функціонує як культурно-нормативне середовище, що формує режими видимості, розподіляє увагу та впливає на етичні судження. Виявлено, що у таких середовищах форма подання інформації – інтерфейс, мова підказок, композиція екрану – не є «оболонкою», а діє як інфраструктура довіри, яка підсилює або послаблює відповідальність, прозорість та справедливість.
Метою дослідження є концептуальна та процедурна інтеграція етичних і естетичних вимірів у життєвому циклі систем штучного інтелекту, зокрема обґрунтування інструменту оцінювання E²-Framework (Ethical–Aesthetic Integration), що переводить принципи відповідальності, підзвітності та справедливості у перевірювані кроки проєктування, тестування та експлуатації, водночас задаючи вимоги до форми подання невизначеності, маркування синтетичності та семіотики ризику.
Результати. Охарактеризовано десять «вузлів» процедурної придатності етики AI: гігієна даних і релевантність вибірок; багатовимірна оцінка точності та справедливості; повноцінні model/system cards і журнали валідації; інтерфейс як носій пояснюваності; реальний human-in-the-loop; незалежний аудит і red-teaming; порційна згода та комунікація ризику; постмаркетинговий моніторинг і пороги відкату; енергетична ощадність; культурна локалізація естетики. Оцінено вплив інтеграції етичних і естетичних критеріїв на поведінкові метрики: зростає частка перевірки джерел, падає частота помилок після другого кроку згоди, підвищується розпізнавання меж узагальнення.
Висновки. Доведено, що відповідальна технологія не зводиться до точності моделі чи «красивого» дизайну окремо; вирішальним є їхня синхронізація в єдиній архітектурі відповідальності. Запропонований E²-Framework формує культуру розробки, у якій кожен піксель має нормативну вагу, а кожен етичний принцип знаходить формальне втілення у мові, ритмі та іконографіці інтерфейсу.
Ключові слова: цифрова етика, цифрова естетика, штучний інтелект, пояснюваність, інтерфейс і довіра, невизначеність, E²-Framework
Стаття надійшла / Article arrived: 17.09.2025 Схвалено до друку / Accepted: 28.10.2025

Вступ.

Штучний інтелект стрімко перетворюється з інструмента автоматизації на середовище існування – таке, що непомітно перебудовує способи бачити, відчувати й оцінювати світ, а відтак і підстави наших моральних суджень. У цьому просторі технічні рішення мають антропологічні наслідки: алгоритми не лише оптимізують поведінку, а й кодують норми, визначають горизонти уваги, формують чуттєві канони. Тому проблематика цифрової етики та естетики більше не є суміжжям двох дисциплін; вона постає як єдиний інтелектуальний вузол, у якому питання справедливості, відповідальності, приватності та автономії переплітаються з питаннями смаку, форми, унаочнення й дизайну людського досвіду. Актуальність цього дослідження зумовлена не стільки темпами впровадження алгоритмів, скільки глибиною їхньої вбудованості у нормативні й символічні порядки – від регулятивних режимів і публічної комунікації до мистецьких практик і повсякденного сприйняття.

Метою статті є концептуальна реконструкція й аналітичне обґрунтування взаємодії етичного та естетичного у цифровому середовищі штучного інтелекту, зокрема виявлення того, яким чином естетичні параметри інтерфейсів, медіа-форм і генеративних моделей впливають на етичні оцінки, а етичні рамки, своєю чергою, задають допустимі горизонти естетичної новизни. Ми прагнемо показати, що «нейтральної» технологічної форми не існує: кожна логіка відбору й подання інформації має нормативний вимір, а кожен нормативний вибір виявляє приховану поетику. У практичному плані завданням є вироблення інтегрованого підходу до оцінки AI-систем, який поєднує критерії відповідальності, справедливості та підзвітності з аналізом формальних рішень – візуальних, мовних, композиційних, – що впливають на сприйняття, довіру й агентність користувача.

Методологічно дослідження спирається на герменевтичний аналіз для тлумачення смислових структур цифрових артефактів, на порівняльний підхід для зіставлення регуляторних і інституційних моделей, на елементи критичної теорії та досліджень науки й технологій (STS) для виявлення владних конфігурацій у проєктуванні та впровадженні алгоритмів, а також на нормативний аналіз, що дозволяє оцінити відповідність практик етичним принципам і правам людини. До цього додаються постфеноменологічні інсайти щодо «міжповерхні» людини й техніки, аналіз дизайну інтерфейсів як інструмента поведінкової архітектури, і кейс-стаді з мистецьких та комунікаційних практик генеративних моделей, які дають змогу перевірити теоретичні висновки у конкретних сценаріях використання.

Наявні підходи до визначення предмета коливаються між етикою принципів і етикою відповідальності, між утилітаристськими оцінками наслідків і деонтологічними обмеженнями, між чеснотнісними перспективами й підходом спроможностей; паралельно в естетиці співіснують аналітичні моделі оцінки мистецького статусу алгоритмічних творів, медіатеорії цифрового образу, підходи дизайну-орієнтованої етики та культурологічні читання візуальних режимів, що задаються даними й моделями. Попри різноманітність, цим лініям часто бракує синхронізації: етичні рамки рідко враховують естетичну «роботу форми», тоді як естетичні дослідження інколи залишають поза увагою інституційно-правові наслідки технічних рішень. У цій статті пропонується інтегративне бачення, яке, з одного боку, зводить естетичні параметри до питання про умови можливості довіри, легітимності та включення, а з другого – вписує етичні критерії у логіку символічного виробництва, де вибір форми є завжди вибором норми.

Таким чином, «цифрова етика та естетика» розуміється тут як спільний предмет дослідження, що описує двосторонній обмін між нормативним і чуттєвим у проєктуванні, регулюванні та використанні AI-систем. Запропонована рамка дозволяє одночасно оцінювати відповідальність і виразність, прозорість і читабельність, справедливість і інклюзивність форми – і тим самим формувати більш адекватні критерії якості технологій, які не зводяться ані до сухих процедур відповідності, ані до автономних художніх експериментів, а натомість орієнтовані на гідність користувача, публічну довіру та культурну стійкість у добу алгоритмів.

Аналіз попередніх досліджень і публікацій. Проблемне поле, у якому перетинаються цифрова етика та естетика штучного інтелекту, уже має свої «маяки», хоч дискусія – ще на підйомі. Класичне ядро сучасної етики AI, як відомо, формувалося довкола принципів відповідальності, підзвітності та справедливості, однак за останні п’ять років ці принципи дедалі частіше «приземлюються» на практики дизайну, інтерфейсів і культурні коди візуальності. Втім, цей поворот від норм до форм відбувається нерівномірно: частина авторів воліє мислити в координатах технічних стандартів, частина – в горизонті гуманітарної критики, і саме між ними (часом продуктивно, часом конфліктно) триває напружений діалог.

Лінію нормативно-принципової етики останніх років репрезентує, зокрема, Л. Флоріді, який систематизує дилеми штучного інтелекту не як «абсолютно нові», а як загострені версії давніх етичних проблем, перенесених у цифрове середовище; у цьому підході впорядковується мова опису й пропонуються рамки для політик та інституційних рішень (Floridi, 2023; 2024). Інституційним продовженням цієї тенденції стала поява стандартоутворювальних документів, що

розгортають мову принципів у процедурні (а інколи й метричні) вимоги: ідеться про рамку NIST AI RMF 1.0, яку розроблено для управління ризиками по всьому життєвому циклу систем штучного інтелекту, а також про спеціальну публікацію NIST щодо виявлення та зменшення упереджень (NIST, 2023; NIST, 2022). На рівні публічного права і політики ЄС цей вектор утілено в AI Act, що закріплює ризик-орієнтовану логіку регулювання й утворює, сказати б, «мінімальний канон» належної розробки та впровадження AI в межах внутрішнього ринку (European Commission, 2024). Поруч – етичні рекомендації рівня soft law, у яких робиться акцент на правах людини, справедливості, інклюзії та культурній чутливості (UNESCO, 2021). Сукупно ці документи структурують поле й, що важливо, задають мову, якою говорять і про техніку, і про цінності.

Разом із тим, критична лінія досліджень настійно нагадує, що «принципи на папері» не знімають напруження між алгоритмами та соціальною тканиною. Е. Бендер з колегами ще на початку хвилі великих мовних моделей попереджали про небезпеку «стохастичних папуг» – систем, що масштабують мовні патерни без розуміння, але з дуже реальними наслідками для довіри, безпеки та екології дискурсу (Bender et al., 2021). А. Біргане пропонує реляційно-етичну оптику (чи краще сказати – перспективу), у якій несправедливості постають не як «збій» коду, а як відбиток ієрархій та асиметрій у даних і практиках їхнього збирання; відповідно, виправлення потребує не лише алгоритм, а й соціальні зв'язки навколо нього (Birhane, 2021). Додамо, що емпіричні огляди освоєння AI в організаціях, попри обережний оптимізм, фіксують асиметрії доступу до вигод, «цифрові розриви» і дефіцити компетенцій, які прямо впливають на справедливість результатів (OECD, 2025; McKinsey, 2025). Публічні настрої та уявлення про ризики – не фоновий шум, а змінна, здатна коригувати легітимність технологій; відповідні зрізи громадської думки показують неоднорідність оцінок і чутливість до контексту застосувань (Pew Research Center, 2025; Stanford HAI, 2025).

На протилежному березі – дослідження естетики й культурної динаміки AI. Останні роки відзначилися інтенсивними спробами описати, що саме змінюється у мистецтві та дизайні, коли творення доручається моделям: від великих міждисциплінарних оглядів про «розуміння і створення мистецтва з AI», де кодифікуються жанри, техніки та питання авторства, до полемічних текстів про статус «AI-арту», який, мовляв, водночас продовжує і підриває звичні категорії авторства й оригінальності (Cetinić & She, 2022; Nannicelli, 2025). Є й – якщо дозволите – більш «педагогічні» або «роз'яснювальні» спроби: пояснити, чому результати генеративних моделей так переконливо наслідують смислові й стилістичні шаблони, і що це говорить нам про природу художнього досвіду (Chiodo, 2024). Окремою лінією вийшли праці, що буквально ставлять поряд естетичні й етичні параметри – приміром, оцінюючи вуглецевий слід людського й машинного авторства (Tomlinson et al., 2024), – і тим демонструють, як «форма» невіддільна від матеріальних обмежень та екологічних наслідків.

На межі цих двох дискурсів – етика стандартів і естетика форм – розгортається дослідження інтерфейсів, тобто тих поверхонь, які ми бачимо щодня, але часто не реєструємо їхнього нормативного ефекту. Тут доречно наголосити: інтерфейс не просто «обличчя» системи, він керує увагою, інсценує довіру, нормалізує відбір і ранжування, а отже, має етичну вагу. Звідси інтерес до «читабельності» моделей, прозорості підказок, пояснюваності та до того, як візуальні чи мовні рішення корегують судження користувача про ризик і справедливість (NIST, 2023; European Commission, 2024). Власне, в такому прочитанні естетика це не «про красу», а про форми досвіду, які роблять можливими або неможливими відповідальне застосування AI.

Нарешті, статистичні й оглядові джерела останніх років працюють як своєрідний «барометр»: у них фіксується темп впровадження, секторальні відмінності, очікування та страхи, а подеколи і зміни в енергетичних та інфраструктурних профілях, пов'язані з використанням моделей (OECD, 2025; McKinsey, 2025; Stanford HAI, 2025; IEA, 2025). Ці матеріали не замінюють філософсько-правовий аналіз, але дозволяють запобігти «романтизації» чи «демонізації» технологій, надаючи емпіричні точки опори для порівняння практик і ефектів.

Підсумовуючи, корпус попередніх досліджень пропонує дві головні опори: нормативно-правові та стандартотворчі рамки, що впорядковують етичні очікування від штучного інтелекту (Floridi, 2023; UNESCO, 2021; NIST, 2023; European Commission, 2024), і культурно-естетичні інтерпретації, які описують трансформацію художніх, медійних і дизайнерських практик у добу алгоритмів (Cetinić & She, 2022; Chiodo, 2024; Nannicelli, 2025). Однак синхронізації між ними ще бракує. Саме тому надалі ми запропонуємо інтегровану рамку, де естетичні параметри інтерфейсів і творених моделей мисляться як невід'ємна складова етичної оцінки, а етичні критерії, як межі й можливості естетичної новизни, що разом задають відповідальнішу і (не побоїмося цього слова) людянішу траєкторію розвитку AI.

Результати дослідження.

1. Цифрова етика штучного інтелекту: від принципів до процедурної придатності.

Етичні «золоті слова» справедливість, підзвітність, прозорість, повага до гідності, працюють лише тоді, коли в них є механіка. Інакше кажучи, коли принципи отримують форму процесу, графік відповідальності та точку відсікання. У штучному

інтелекті ця механіка починається не на екрані, а задовго до нього: з моменту задуму продукту, коли формулюється користь і визначається допустимий ризик для конкретних груп. Тут ми пропонуємо «процедурну карту» етики – спосіб перевести декларації у повторювані, перевірювані кроки, що вбудовуються у життєвий цикл системи й не залежать від добрих намірів окремих інженерів чи менеджерів (NIST, 2023; European Commission, 2024).

Перший вузол – дані. Етичність моделі починається з гігієни наборів: від опису джерел і підстав правомірності збору (правова основа, інформована згода, легітимний інтерес) до картування репрезентативності та вразливих категорій. Практика «data sheets» і «dataset statements» не повинна бути красивою декларацією для репозиторію, а контрактом із реальністю: що саме зібрано, як це балансували, які параметри заборонено відтворювати, який «пороговий рівень» токсичності чи асиметрії тригерить повторний відбір (Birhane, 2021). Тут же знаходиться базова чесність щодо невизначеності: відсутність даних не слід маскувати алгоритмічною фантазією; краще «скромне» передбачення з широким інтервалом довіри, ніж точна, але оманлива цифра.

Другий вузол – навчання й оцінювання. Модель потребує не лише метрик точності, а й метрик справедливості, стабільності та відтворюваності. Наш підхід наполягає на «багативимірному заліку»: якщо система не проходить мінімальні пороги для групових показників (наприклад, різниця в false negative rate для субпопуляцій виходить за встановлений діапазон), вона не має рухатися в продакшен, навіть якщо загальна точність блискуча. Це не перфекціонізм, а техніка процедурної справедливості: не карати «середнім» значенням тих, хто систематично програє на краях розподілу (NIST, 2022). І так, тут потрібні не лише формули, а й протоколи: хто підписує реліз із пограничними метриками, хто ініціює повторне навчання, хто має право сказати «стоп».

Третій вузол – документація. «Model cards» і «system cards» повинні перестати бути PR-артефактами. У них мають жити сфера застосування, заборонені сценарії, межі узагальнення, залежності від контексту, режим відкату й контактний канал для скарг. Не менш важливі журнали валідації: скільки експериментів провели, на яких зрізах даних, які граничні кейси провалені й чому. Коли виникає інцидент, документація стає процесуальним щитом і, водночас, моральним обов'язком перед користувачем.

Четвертий вузол – інтерфейс і мова взаємодії. Здавалося б, «естетика» є поверхневим шаром, але саме тут етика або оживає, або в'яне. Пояснення без читабельності не виконує своєї функції, попередження без таймінгу теж. Ми доводимо, що етичні принципи необхідно втілювати в «поведінкових конструкціях» інтерфейсу: видиме маркування невизначеності (наприклад, діапазони, ймовірнісні смуги), семантично коректні мікротексти без удаваної впевненості, кнопки «перевірити джерела» в один клік, «червоні лінії» для високоризикових сценаріїв, які не можна обійти простим підтвердженням. Іншими словами, прозорість це не лише відкритий код, а й відкрита форма доступу до сумнівів системи. І так, боротьба з «темними патернами» – не питання смаку, а питання справедливості.

П'ятий вузол – людина в циклі. «HITL» має бути реальним, а не декоративним. Людський контроль це не кнопка «approve», яку натискають автоматично, а процедура з правом вето, з часом на переоцінку, з ясним переліком ситуацій, де машина має помилятися «на користь обережності». Ми пропонуємо розподіляти ролі за матрицею RACI ще на старті продукту: хто відповідає за етичні ризики, хто консультує, хто приймає рішення, хто несе відповідальність у разі шкоди. Без цього «підзвітність» розчиняється у колективній безвідповідальності, а «людський нагляд» перетворюється на ритуал.

Шостий вузол – незалежний аудит і «червона команда». Розробник не може бути єдиним суддею у власній справі. Регулярні зовнішні перевірки з доступом до журналів, датасетів і протоколів тестування не лише підвищують якість, а й створюють культуру «етичної дисципліни». Red-teaming – окрема історія: агресивні сценарії з обходом фільтрів, атаки на контекст, спроби підбурювання до шкідливих дій. Після кожної сесії з'являється не просто звіт, а обов'язковий план ремедіації, із датами та відповідальними. Такий підхід особливо резонує з ризик-орієнтованою логікою AI Act і практиками керування ризиками в NIST AI RMF (European Commission, 2024; NIST, 2023).

Сьомий вузол – комунікація ризику та згоди. Інформована згода в цифрових сервісах часто є фікцією, і ми це прямо визнаємо. Але це не привід опускати руки. Потрібно відмовлятися від «загальних» політик на 50 сторінок і переходити до «порційної» згоди: короткі блоки, прив'язані до конкретної дії (наприклад, увімкнення персоналізованого режиму), зі зрозумілою опцією відмови без покарання. Коли йдеться про високий ризик, згода має стати двокроковою і відстроченою (give it a minute): лише так можна мінімізувати імпульсивні погодження, до яких штовхають усталені UX-патерни. Цей, зізнаємося, «повільний» дизайн є частиною етики відповідальності.

Восьмий вузол – експлуатація і супровід. Етика не закінчується релізом. Потрібні пороги автоматичного відкату у відповідь на деградацію метрик або на сигнали з каналу скарг. Потрібна програма «постмаркетингового» моніторингу з чіткими періодами переоцінки, особливо для систем, які «вчаться на льоту». Потрібна, зрештою, бухгалтерія енергоспоживання: обчислювальна ощадність не лише про гроші, а й про екологічний обов'язок, який,

визнаймо, теж є етичним (Tomlinson et al., 2024).

Дев'ятий вузол – соціальний контекст і легітимність. Навіть бездоганна зсередини система може «не пройти» суспільне прийняття, якщо її мета або спосіб упровадження суперечать колективним уявленням про справедливість. Соціологічні зрізи не прикра статистика, а частина етичної оцінки: вони показують, де пульсує недовіра, які групи відчувають непропорційний тиск, які домени потребують додаткових запобіжників (Pew Research Center, 2025; OECD, 2025). Ми, отже, розуміємо етику не як «правильність у вакуумі», а як здатність витримати зустріч із реальною публікою.

Десятий вузол – семантика пояснюваності. Пояснення це не декорація довкола чорної скриньки, а мовленнєвий акт, який змінює поведінку. Погане пояснення небезпечніше за його відсутність: воно вселяє хибну впевненість і підштовхує до передчасної згоди. Ми наполягаємо на чесній мові невизначеності, на явних застереженнях щодо меж узагальнення, на відкритому розведенні «наших припущень» і «наших спостережень». Тут варто пам'ятати і про «стохастичних папуг» – системи, здатні складно формулювати банальності та водночас пропускати очевидні лакуни (Bender et al., 2021). Коротко: пояснюваність має бути вірною за змістом, поміркованою за тоном і корисною за наслідком.

Насамкінець – про культурну оптику. Етичні норми не універсалізуються силоміць: дизайн, що працює в одній культурі, може мати іншу конотацію в іншій. Тому «процедурна придатність» включає етап локалізації: мову, символіку, темп, навіть гумор. Лише так принципи Флоріді про «загострені, а не нові» етичні проблеми набувають життєвості у щоденних практиках взаємодії з технологією (Floridi, 2023; 2024). Процедура, якщо вона добра, завжди має місце для людини – її сумніву, її права відмовитися, її права перепитати.

Якщо звести все до одного речення, то воно буде таким: етика штучного інтелекту не існує поза процесом, а процес поза формою подання. Саме перехід від принципів до процедурної придатності робить системи не лише законно допустимими, а й людяними, бо гарантує не просто «відсутність зла», а присутність турботи про користувача – від вибору даних до останнього пікселя повідомлення про ризик.

2. Естетика цифрової доби: форма як інфраструктура довіри.

Почнемо з простого спостереження, яке, однак, має далекосяжні наслідки: у цифрових середовищах форма не «супроводжує» зміст, вона його організує. Розташування елементів, типографіка, контраст, щільність інформації, швидкість анімацій, ритм появи підказок, усе це не другорядні дрібниці, а механізми керування увагою, тобто механізми впливу на судження. Саме тому естетика інтерфейсів працює як інфраструктура довіри: вона непомітно налаштовує користувача на те, як читати сигнали системи, коли сумніватися, а коли погоджуватися. Там, де пояснюваність моделі обмежена (а в генеративних системах це трапляється часто), дизайн – хочемо ми того чи ні – бере на себе роль «перекладача» невизначеності, компенсуючи дефіцит прозорості стилем і темпом подання. Занадто впевнена мова мікрокопі, заспокійливі кольори, «очевидні» іконки, плавні анімації без пауз – і користувач майже фізично відчуває схильність довіряти. Варто лише додати, що ця довіра може бути як заслуженою, так і оманливою.

Естетика, яка підтримує відповідальну взаємодію з AI, має декілька ключових властивостей (назвемо їх так для зручності, але маємо на увазі узгоджену систему, а не набір прийомів). По-перше, чесна візуалізація невизначеності: замість «однієї відповіді» мають бути діапазони, ймовірнісні смуги, маркування конфіденційності, чіткі індикатори помилок і «м'які» переходи до альтернатив (не «окей/скасувати», а «переглянути джерела», «уточнити запит», «попросити людську перевірку»). По-друге, розумна ритмізація: невеликі затримки на критичних кроках (на кшталт «дай собі секунду, це важливе рішення») зменшують імпульсивність і водночас сигналізують, що система «знає про ризик». По-третє, відкритий родовід контенту: маркування синтетичності, посилання на джерела навчання або принаймні на логіку добору матеріалу, зрозумілі підказки щодо меж узагальнення. Вся ця «семіотика походження» підсилює легітимність у момент сприйняття (пор. European Commission, 2024; NIST, 2023). Коли ж ці властивості відсутні, форму легко перетворити на інструмент нав'язування: темні патерни, «липкі» CTA-кнопки, дизайнерські трюки, що розмивають кордони згоди, створюють естетичну режисуру, яка практично підмінює етику.

У генеративному мистецтві та медіа естетичне питання постає ще гостріше. Адже тут форма є не тільки оболонкою, а й результатом. Модель пропонує схеми композиції, колористику, стилістичні цитати, риторику образу; користувач часто лише коригує параметри або, як кажуть, «підкручує» стиль. Це породжує подвійний ефект. З одного боку, відбувається демократизація творення (поріг входу знижується, а репертуар стилів стає доступним усім). З іншого, посилюється серійність і «сліпа пляма» щодо джерел: те, що виглядає «новим», може бути лише вправним переукладанням знайдених у даних патернів (Cetinić & She, 2022; Chiodo, 2024). Питання авторства не тільки юридичне (хоч і це важливо), воно також етичне: кого ми уявляємо автором, коли «виразність» створена статистикою уподобань? Яким чином ми розрізняємо «естетичну переконливість» і «пізнавальну цінність» зображення? Цей скепсис не означає заперечення, а радше вимогу до прозорої гри: вкажіть джерела, покажіть межі, назвіть невідоме.

Власне, у такому світлі стає зрозуміло, чому

питання енергоефективності та матеріальних витрат також естетичне. Не лише в тому сенсі, що «ощадні» інтерфейси дисциплінують надмірну візуальність, а й тому, що естетика економії формує культуру помірності: менше беззмістовних анімацій, менше декоративного «шуму», більше семантичної ваги кожного екрану. Коли на горизонті мільйони звернень до моделей щодня, мікро-рішення дизайну конвертуються у макро-наслідки для енергетичного сліду систем (Tomlinson et al., 2024; IEA, 2025). Чи не є це, зрештою, ще одним виміром довіри: ми демонструємо повагу до користувача і до довкілля тим, що не змушуємо його очі та процесори працювати зайве?

Окрему тему представляє семіотика ризику. Колір небезпеки, форма попереджувальних ікон, гучність звукових сигналів, розміри шрифтів для критичних повідомлень – усе це вже давно досліджено у галузях авіації й медицини, але в цифрових продуктах нового покоління ці стандарти часто ігнорують, надаючи перевагу «красі» над ієрархією важливості. Етична естетика наполягає на зворотному: якщо система не здатна чесно показати межі власної компетентності (червоним, великим шрифтом і без евфемізмів) вона не заслуговує на довіру, якою б витонченою не була її візуальна мова. На практиці це означає відмову від риторики всезнання у мікротекстах («я впевнений», «точна відповідь», «безпомилково») на користь епістемічно обережних формулювань та явних посилань на джерела (Bender et al., 2021).

До цього додається культурна чутливість, без якої жодна «інфраструктура довіри» не працює однаково в різних спільнотах.

Образотворчі коди, які у Північній Європі сприймаються як «мінімалістично-надійні», на Близькому Сході можуть уособлювати «холодну відстороненість»; кольори підтримки та небезпеки, жестові метафори, навіть швидкість переходів між екранами – усе потребує локалізації, а не механічного перенесення. Тут естетика стає суспільною етикою в дії: вона враховує історичну пам'ять, символічні ландшафти, усталені моделі ввічливості й відмови. Іншими словами, ми не проектуємо «голий» інтерфейс, ми конструюємо ситуацію довіри, і ця ситуація щоразу має локальну граматику.

І, нарешті, питання часу – майже забуте в дискусії про форму. Швидкість відповіді, довжина «обмірковування», видимість процесу (той самий «typing indicator» чи «generating…»), наявність паузи перед ризиковою дією – це теж естетика, але естетика темпу. У повільніших конфігураціях користувач отримує простір на рефлексію; у гіпершвидкихвін радше підтверджує, ніж обирає. Спроба «приховати» час обчислення вбиває відчуття процесуальності; натомість скромне візуальне визнання зусилля («обчислюю на неповних даних», «перевіряю джерела», «звіряю з політиками») формує чесний очікуваний горизонт і робить помилку не катастрофою, а приводом до уточнення.

Підсумовуючи, можемо сформулювати практичний критерій: естетика, що гідна довіри, це така організація форми, яка збільшує видимість невизначеності, підсилює здатність користувача зупинитися, відмовитися, перевірити й оскаржити, і водночас не паралізує взаємодію надміром тривожних сигналів. Вона грає «за команду етики», а не проти неї, знімаючи спокусу прикривати дизайнерськими ефектами слабкі місця алгоритму. Саме так розуміється «форма як інфраструктура довіри»: не фасад для прикраси, а несуча конструкція, яка тримає вагу відповідальності від першого погляду на екран до останнього натискання «підтвердити». У цій конструкції кожен піксель – політичний, кожна пауза – моральна, а кожна рима кольорів – аргумент за (або проти) людяності штучного інтелекту (Cetinić & She, 2022; Nannicelli, 2025; European Commission, 2024; NIST, 2023; Tomlinson et al., 2024; IEA, 2025).

3. Перетин етичного та естетичного: інтегрована рамка оцінювання (E^2-Framework).

Етична придатність системи штучного інтелекту не існує окремо від способу її подання, так само як естетична форма не є невинною оболонкою. Тому E^2-Framework (Ethical–Aesthetic Integration) пропонує дивитися на систему водночас «у глибину» алгоритмів і «по поверхні» інтерфейсів, оцінюючи їх як єдину архітектуру відповідальності. У цьому підході немає другорядних елементів: вибір датасету так само впливає на довіру, як і вибір типографіки; поріг ризику так само визначає поведінку користувача, як і мова мікрокопі. Інтеграція відбувається не в деклараціях, а у спільному полі рішень, де етичні критерії накладаються на форму подання й коригують її, а форма, своєю чергою, виводить етику з абстракції в досвід.

Щоб зробити цей рух практичним, рамка працює у двох координатах. Перша це етична достатність: наскільки справедливо розподіляються помилки між групами; чи є прозора схема підзвітності; чи встановлено пороги зупинки й механізми відновлення прав суб'єкта; чи задокументовано межі узагальнення; чи пройдено зовнішній аудит. Друга – естетична відповідальність: чи допомагає форма коректно сприймати невизначеність; чи відсутні «темні патерни», що підштовхують до імпульсивної згоди; чи відрізняється тональність повідомлень у критичних сценаріях від повсякденних; чи видно «родовід» контенту й маркери синтетичності. Важливо, що координати не складаються арифметично (високий бал за алгоритм не «перекриває» маніпулятивний інтерфейс, і навпаки); радше вони утворюють поле напруги, у якому рішення «допустити до застосування» можливе лише за наявності мінімуму на обох осях.

На рівні процесу E^2-Framework прив'язується

до життєвого циклу системи й змінює ритм роботи команд. На етапі задуму формулюється не лише корисність, а й допустимий ризик для конкретних спільнот; одразу ж визначаються «точки зупинки» – ті етичні та естетичні пороги, після яких продукт не рухається далі, як би не спокушали бізнес-метрики (European Commission, 2024). Під час проєктування даних описуються джерела, підстави правомірності збору й вразливі категорії, але паралельно прототипується інтерфейс подання невизначеності: діапазони, ймовірності, альтернативні дії. У навчанні моделі вимірюють не лише точність, а й стабільність і справедливість на субвибірках; у дизайні не лише «красу», а й читабельність пояснень, видимість ризику, опір «слизьким» сценаріям. На валідації алгоритм проходить тестування за групами, а інтерфейс – A/B-експерименти: чи користувач розрізняє припущення і факти; чи встигає відмовитись; чи зростає ймовірність перевірки джерел. У релізі з'являється не тільки model card, а й interface card: сфери застосування, табуйовані сценарії, тональність повідомлень, правила маркування синтетичності, енергетичний профіль взаємодії (і так, це теж частина етики). Після запуску моніторинг метрик справедливості й журнал скарг з'єднуються з «естетичним журналом»: фіксуються випадки, коли форма помилково підштовхнула до згоди або замаскувала невизначеність; за перевищенням порогів автоматично спрацьовує відкат, а не «ще трохи потерпімо».

Щоб уникнути зручної, але хибної «оцінки середнім», E²-Framework оперує тепловими мапами, а не фінальними балами. Умовно кажучи, система може бути «зеленою» за точністю, «жовтою» за справедливістю і «червоною» за читабельністю пояснень – і саме ця «червоність» блокує випуск, бо користувач не має інструментів усвідомленого рішення. Такий підхід дисциплінує інженерів і дизайнерів однаково: якщо інтерфейс знімає сумнів замість того, щоби його показати, він зменшує етичну вагомість згоди; якщо алгоритм підміняє брак даних надмірною впевненістю, то жодна витончена типографіка не виправить дефект. У практиці публічного сектору це особливо помітно: система рекомендацій соціальної допомоги може мати чудові ROC-криві й водночас провалювати естетичний іспит, скажімо, мікротекстами, які соромлять користувача або приховують право на апеляцію; у такій конфігурації E²-Framework каже не «майже добре», а «неприйнятно».

Показовим є й набір показників, що «протягують» етику через форму. Водночас із розривами помилок між групами вимірюються час читання пояснень, частка користувачів, які обрали «перевірити джерела», рівень розпізнавання невизначеності після взаємодії, частота усвідомлених відмов від ризикової дії, індекс когнітивного навантаження (так, ми не боїмося психометрії), відсоток виправлених рішень після другого кроку згоди, а також енергетична вартість типової сесії. Коли дизайн «працює чесно», ці показники ведуть себе передбачувано: користувач повільнішає там, де справді ризиковано; частіше відкриває джерела, коли модель сумнівається; рідше помиляється у повторних завданнях. Коли ж інтерфейс занадто «гладкий», успіхи точності перекладаються у поспіх і сліпу довіру – і це видно по метриках, аж до стрибків зворотного зв'язку у каналі скарг.

Управлінська частина рамки не менш важлива за технічну. Команда отримує чітку матрицю ролей: дизайнер відповідає не лише за колір кнопки, а й за епістемічний тон повідомлення; інженер – не лише за втрати, а й за алгоритмічну скромність; юрист – не лише за відповідність текстам політик, а й за процедурність згоди у конкретних сценаріях. Паралельно інституціалізується «естетичне рев'ю» регулярна зустріч, на якій проходять через інтерфейс так само прискіпливо, як через тест-сет: де ми приховали невизначеність; де підмінили альтернативу «липкою» кнопкою; де забули про локальну семіотику (колір небезпеки, жест відмови, правила ввічливості). І якщо щось іде всупереч цим вимогам, рішення не косметичне, а архітектурне: змінюємо послідовність дій, переосмислюємо повідомлення, повертаємося до порогів.

Можна запитати: чи не перетворює такий підхід розробку на нескінченний ритуал? Лише якщо його імітувати. Там, де E²-Framework приймають всерйоз, він економить ресурси: помилки зникають раніше, ніж стають публічними кризами; аудити минають без паніки, бо документація ведеться «по дорозі», а не «заднім числом»; довіра накопичується як капітал, а не як обіцянка. І, що важливо, рамка не забороняє творчості, вона просто не дозволяє естетиці прикривати етичні порожнини, а етиці залишатися у стані чистої моральної риторики. У цьому сенсі E² – не ще один чек-лист, а культура: здатність бачити в кожному пікселі норму, у кожній нормі іншу форму, і в кожній взаємодії людину, якій ми зобов'язані чесністю, навіть тоді (особливо тоді), коли система виглядає бездоганно переконливою.

На політико-правовому рівні це означає, що вимоги відповідності мають прямо включати параметри представлення (presentation requirements): стандарти для позначення невизначеності, для візуальної диференціації рекомендації й рішення, для маркування синтетичного контенту та його походження. На інституційному рівні це регулярні «естетичні рев'ю» на додачу до безпекових: міждисциплінарні засідання, де дизайнер, юрист і інженер однаково відповідальні за зміст підказки, іконографіку ризику, мову попереджень. На ринку перевагу отримують ті системи, які не приховують межі власної компетентності; користувач, що бачить правдиву картину невизначеності, рідше помиляється і частіше довіряє (не сліпо, а відповідально). І, так,

екологічний вимір теж тут: естетика ощадливих інтерфейсів корелює з ощадливими обчисленнями, а тому знижує енергетичний слід моделей, що для масового застосування вже не дрібниця (Tomlinson et al., 2024; IEA, 2025).

Висновки.

Підсумовуючи здійснене дослідження, слід визнати очевидне, хоча часто приховане за технократичною лексикою: штучний інтелект не лише оперує даними та моделями, він переписує режими видимості й чутливості, а отже торкається підвалин морального судження. Ми спробували показати, що етика й естетика в цифровому середовищі не утворюють двох паралельних потоків, які випадково збігаються у точках застосування; навпаки, йдеться про взаємну залежність, де форма несе норму, а норма визначає межі та можливості форми. Саме тому запропонована інтегрована рамка оцінювання E²-Framework (Ethical–Aesthetic Integration) не додає «краси» до «корисності», а з'єднує відповідальність із читабельністю, справедливість зі зрозумілістю, прозорість із чесною репрезентацією невизначеності. У цій оптиці етична якість системи не може бути високою там, де інтерфейс примушує, вводить в оману або маскує межі компетентності; так само естетична перевага не варта багато, якщо вона здобута ціною «темних патернів» і поведінкових пасток.

Новизна дослідження полягає в тому, що ми відмовляємося від інерційного поділу «принципи тут, дизайн там» і наполягаємо на процедурній придатності етичних вимог через мовні та візуальні рішення, якими користувач взаємодіє з системою. Іншими словами, етична декларація набуває сили норми лише тоді, коли отримує відповідник у мові підказок, у маркуванні невизначеності, у способі подання пояснень, у ясних «порогах зупинки», після яких модель не експлуатується. Цей зсув, від абстракції до архітектури, водночас розширює і сферу відповідальності. Дизайнер стає співносієм етичного змісту, інженер – співавтором естетичної граматики, юрист – куратором сценаріїв використання, де гідність суб'єкта й довіра спільноти не можуть бути зведені до галочки «compliant».

Практичний сенс такого підходу відчутний одразу у трьох площинах. По-перше, в публічній політиці: вимоги відповідності мають прямо включати параметри представлення, від візуальних маркерів синтетичності до стандартизованої подачі ризиків і альтернативних виборів користувача. По-друге, в організаційній культурі: команди мають інституціалізувати «естетичні рев'ю» нарівні з безпековими аудитами, адже саме у мікродеталях інтерфейсу часто вирішується доля етичного вибору. По-третє, в продуктах і сервісах: системи, які чесно показують власну невизначеність і не симулюють всемогутність, у довгій перспективі виграють. Вони зменшують помилки користувачів, підсилюють контрольованість процесів і накопичують капітал довіри.

Звісно, межі нашого дослідження окреслені. Ми працювали зі змішаними методами й інтерпретативними аналізами, що потребують подальшої кількісної верифікації: від A/B-тестів пояснюваності до польових експериментів з маркуванням невизначеності у критичних доменах (медицина, правосуддя, освіта). Окремого опрацювання вимагає екологічний аспект естетики інтерфейсів і моделей: енергоспоживання, вуглецевий слід, «ощадна форма» як етичний обов'язок у масових застосуваннях. Не менш важливо продовжити міжкультурні студії, щоби не універсалізувати локальні естетичні норми та не переносити їх механічно в інші соціальні контексти, де коди довіри та способи взаємодії з технологіями суттєво відрізняються.

Перспективи подальших досліджень вбачається щонайменше у трьох напрямах.

1. Розробка метричної частини E²-Framework, яка дозволить порівнювати системи між собою не лише за точністю чи збалансованістю даних, а й за «естетичною відповідальністю» форми: читабельністю пояснень, коректністю маркування ризику, відсутністю маніпулятивних сценаріїв.

2. Інтеграція прав людини й етичних принципів у процеси дизайну на ранніх етапах, коли архітектурні рішення ще пластичні й найдешевші для виправлення.

3. Створення відкритих кейс-банків, де поєднуються юридичні, технічні й дизайнерські артефакти, що дозволять спільноті вчитись на успіхах і помилках, а не винаходити велосипед окремо в кожній організації.

Отже, якщо підвести риску: у добу алгоритмів гуманність технологій визначається не лише тим, що саме вони обчислюють, а й тим, як саме вони стають видимими у слові, у жесті інтерфейсу, у чесності до власних меж. У цій подвійній оптиці, етичній та естетичній, і полягає наш головний висновок: відповідальний штучний інтелект це не сума процедур і не феєрверк візуальних трюків, а налаштований на гідність суб'єкта, прозорий щодо невизначеності й уважний до форм досвіду інструмент співжиття, який розвивається разом із культурою, котру обслуговує (і, звісно, формує).

REFERENCES

Aris, S., Moosavand, M., & Nosrati, S. (2023). A Digital Aesthetics? Artificial Intelligence and the Future of the Art. *Journal of Cyberspace Studies, 7*(2), 219–236. https://doi.org/10.22059/jcss.2023.366256.1097

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns, 2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Cetinić, E., & She, J. (2022). Understanding and creating art with AI: A comprehensive survey. *ACM Computing Surveys*. https://doi.org/10.1145/3475799

Chiodo, S. (2024). What AI "art" can teach us about art. *Journal of Aesthetics & Culture, 16*(1), 2395511. https://doi.org/10.1080/20004214.2024.2395511

European Commission. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on Artificial Intelligence (AI Act). *Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press. https://doi.org/10.1093/oso/9780198883098.001.0001

Floridi, L. (2024). *The ethics of artificial intelligence: Exacerbated problems rather than novel ones?* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4801799 SSRN

International Energy Agency (IEA). (2025). *Energy and AI*. Paris: IEA. https://www.iea.org/reports/energy-and-aiIEA

McKinsey & Company. (2025). *The state of AI: How organizations are rewiring to capture value*. https://www.mckinsey.com/~/media/mckinsey/business%20 functions/quantumblack/our%20insights/the%20state%20of%20ai/2025/the-state-of-ai-how-organizations-are-rewiring-to-capture-value_final.pdf

Nannicelli, T. (2025). Mass AI-art: A moderately skeptical perspective. *The Journal of Aesthetics and Art Criticism*. https://doi.org/10.1093/jaac/kpaf026 PhilPapers

National Institute of Standards and Technology (NIST). (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf NIST Technical Series Publications

National Institute of Standards and Technology (NIST)**.** (2022). *Towards a standard for identifying and managing bias in AI* (Special Publication 1270). https://doi.org/10.6028/NIST.SP.1270

Organisation for Economic Co-operation and Development (OECD**).** (2025). *Emerging divides in the transition to artificial intelligence*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/emerging-divides-in-the-transition-to-artificial-intelligence_eeb5e120/7376c776-en.pdf

Organisation for Economic Co-operation and Development (OECD). (2025). *The adoption of artificial intelligence in firms*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/05/the-adoption-of-artificial-intelligence-in-firms_8fab986b/f9ef33c3-en.pdf

Pew Research Center. (2025, April 3). *How the U.S. public and AI experts view artificial intelligence*. https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/

Scientific Reports (Tomlinson, B., Black, R. W., Patterson, D. J., & Torrance, A. W.). (2024). The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports, 14*, 3732. https://doi.org/10.1038/s41598-024-54271-x

Special Eurobarometer 554. (2024). *Artificial intelligence and the future of work*. European Commission. http://dx.doi.org/10.2767/8591026

Stanford Institute for Human-Centered AI (Maslej, N., et al.). (2025). *AI Index Report 2025*. https://arxiv.org/abs/2504.07139

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455