

ОГЛЯД ДОСЛІДЖЕННЯ

The Development of Generative Artificial Intelligence from a Copyright Perspective

Розвиток генеративного штучного
інтелекту у контексті авторського права



Реквізити дослідження EUIPO

Параметр	Відомості
Оригінальна назва дослідження	The Development of Generative Artificial Intelligence from a Copyright Perspective
Український переклад назви	Розвиток генеративного штучного інтелекту у контексті авторського права
Видавець	Відомство з інтелектуальної власності Європейського Союзу (EUIPO)
Дата публікації	Травень 2025 року
Автори	University of Turin Law School та Nexa Center for Internet & Society of the Polytechnic of Turin
Обсяг	436 сторінок, 23 додатки
Методологія	Кабінетне дослідження, інтерв'ю зі стейкхолдерами, експертні обговорення
Доступ	Відкритий (CC BY 4.0)
ISBN	978-92-9156-369-2
DOI	10.2814/3893780
Електронний доступ	https://doi.org/10.2814/3893780

Зміст

1. Вступ	4
Актуальність проблематики GenAI та авторського права	
2. Загальна характеристика	4
GenAI Input і GenAI Output як два ключові напрями дослідження	
3. Структура та методологія	5
Методи дослідження (кабінетне, інтерв'ю, експертні обговорення) та межі аналізу	
4. Технічний контекст розвитку генеративного ШІ	6
Етапи створення й роботи ШІ-систем та їх значення для авторського права	
5. Правове регулювання генеративного ШІ у сфері авторського права в ЄС	7
Роль CDSM Directive та AI Act	
6. Використання контенту для навчання моделей генеративного ШІ	8
Конфлікт інтересів ШІ-розробників і правовласників у ланцюгу даних	
7. Резервування прав та механізми opt-out	9
Юридичні й технічні засоби резервування, межі Robots Exclusion Protocol	
8. Додаткові технічні засоби захисту контенту	10
Non-reservation solutions: protective perturbations, crawler blockers, fingerprinting	
9. Ліцензування контенту для систем генеративного ШІ	11
Формування ринку ліцензій, попит на якісний контент	
10. Походження даних та ризику зміни контексту їх використання (data laundering)	11
Передача даних та перевірка його правового статусу	
11. Прозорість та ідентифікація результатів генеративного ШІ	12
Provenance tracking, watermarking, fingerprinting, виявлення ШІ-контенту та їх обмеження	
12. Ризики відтворення охоронюваного контенту у результатах генеративного ШІ	13
Memorisation, regurgitation, plagiaristic output та механізми запобігання	
13. Технічні та договірні механізми управління ризиками	14
Model unlearning / model editing, contractual indemnification	
14. Судові спори та формування практики	15
Судова практика США та ЄС	
15. Роль публічних інституцій та IP офісів	16
Підтримка діалогу та міжвідомчої взаємодії	
16. Значення дослідження для України та можливість проведення національного дослідження	16
17. Оцінка дослідження	18
18. Висновки	18
Примітка щодо характеру огляду	19
Застереження про інформаційно-аналітичний статус матеріалу	
Глосарій ключових термінів	20
Терміни за тематичними блоками з українськими відповідниками та поясненнями	

1. Вступ

Розвиток генеративного штучного інтелекту (Generative Artificial Intelligence, GenAI) став одним із ключових викликів для сучасної системи авторського права, зокрема через використання охоронюваного контенту для навчання моделей GenAI та питання прозорості результатів, згенерованих за допомогою таких систем.

Саме цій проблематиці присвячене дослідження Відомства з інтелектуальної власності Європейського Союзу “The Development of Generative Artificial Intelligence from a Copyright Perspective”, опубліковане у травні 2025 року.

Дослідження має комплексний характер і поєднує технічний, правовий та економічний аналіз взаємодії генеративного штучного інтелекту з авторським правом Європейського Союзу.

Актуальність дослідження зумовлена тим, що генеративний штучний інтелект уже істотно впливає на способи створення, поширення та використання контенту. Системи GenAI здатні генерувати тексти, зображення, аудіо, відео, програмний код та інші результати, які можуть бути використані у творчій, комерційній, освітній, науковій, інформаційній та професійній діяльності. Водночас для розробки таких систем використовуються великі масиви даних, серед яких можуть бути твори, об'єкти суміжних прав, бази даних та інший охоронюваний контент.

Саме тому розвиток GenAI актуалізує низку питань для системи авторського права: як охоронюваний контент використовується для навчання моделей; як правовласники можуть резервувати свої права; чи можуть і як мають ліцензуватися дані для ШІ-навчання; які технічні рішення дають змогу ідентифікувати контент, згенерований за допомогою ШІ; яким чином запобігати генеруванню результатів, що можуть порушувати права третіх осіб; яку роль у цьому процесі можуть відігравати публічні інституції та IP офіси.

Дослідження EUIPO не є офіційним тлумаченням положень права ЄС і не спрямоване на підтримку конкретного технічного рішення. Його цінність полягає в тому, що воно системно описує проблематику GenAI та авторського права, а також аналізує практичні правові, технічні, договірні й інституційні механізми, які вже застосовуються або перебувають у процесі розвитку.

Цей огляд має інформаційно-аналітичний характер і підготовлений на підставі відкритих матеріалів EUIPO. Наведені висновки не є офіційним тлумаченням положень права Європейського Союзу.

2. Загальна характеристика

Дослідження EUIPO побудоване навколо аналізу того, як генеративний штучний інтелект взаємодіє з авторським правом на різних етапах функціонування ШІ-систем. У центрі уваги перебувають два ключові напрями: використання контенту як вхідних даних для розробки та навчання моделей, а також генерування результатів за допомогою GenAI.

Перший напрям умовно можна позначити як GenAI Input. Він охоплює питання збору, обробки, структурування та використання даних для навчання ШІ-моделей. На цьому етапі особливе значення мають правовий статус навчальних даних, можливість застосування винятків для інтелектуального аналізу тексту і даних (text and data mining, TDM), наявність або відсутність застереження правовласника щодо використання контенту (резервування прав/opt-out), договірні обмеження, ліцензування та управління ризиками у ланцюгу розробки ШІ-систем.

Другий напрям – GenAI Output. Він стосується результатів, згенерованих або змінених за допомогою генеративного штучного інтелекту. У цьому контексті досліджуються питання прозорості такого контенту, його маркування, відстеження походження контенту (provenance tracking), цифрового маркування (watermarking), цифрових відбитків (fingerprinting), виявлення ШІ-контенту (AI-generated content detection), а також ризики відтворення охоронюваного контенту у результатах GenAI. Особливою перевагою дослідження є його міждисциплінарність. Авторське право у сфері GenAI неможливо розглядати ізольовано від технічної архітектури моделей, способів збору даних, особливостей вебскрепінгу (web scraping), функціонування RAG-систем, ринку ліцензування та договірних практик ШІ-провайдерів. Саме тому документ поєднує юридичний аналіз із технічним і економічним контекстом.

Важливо, що дослідження має практично орієнтований підхід. Тобто воно не лише констатує наявність проблем, а й систематизує можливі рішення: механізми opt-out та додаткові технічні засоби захисту контенту, ліцензійні моделі, інструменти прозорості ШІ-контенту, заходи протидії запам'ятовуванню навчальних даних (memorisation) та їх відтворенню в результатах (regurgitation), договірні гарантії, а також можливу роль публічних інституцій.

Водночас у документі простежується обережний підхід: EUIPO не подає жоден механізм як універсальне рішення для всіх ситуацій. Навпаки, дослідження показує, що взаємодія GenAI та авторського права потребує поєднання правових, технічних, договірних, організаційних та інформаційних інструментів.

3. Структура та методологія

Методологія дослідження включає кілька взаємопов'язаних елементів. По-перше, було проведено кабінетне дослідження. Воно охоплювало аналіз наукових публікацій, галузевих документів, експертно-аналітичних документів (white papers), звітів міжнародних та національних інституцій, матеріалів технологічних компаній, судових матеріалів, прес-матеріалів та інших джерел. З огляду на швидкий розвиток GenAI, до уваги також бралися джерела, які відображають актуальні ринкові й технологічні практики.

По-друге, дослідження включало інтерв'ю із стейкхолдерами. До них належали представники ШІ-компаній, постачальників технічних рішень, правовласників з різних секторів, а також публічних організацій. Такий підхід дозволив врахувати не лише теоретичний аналіз, а й практичні потреби учасників ринку.

По-третє, попередні результати дослідження обговорювалися у межах експертних воркшопів. Це дозволило уточнити висновки, перевірити їх із практичної точки зору та краще оцінити стан розвитку технічних і правових рішень.

Водночас важливо враховувати межі дослідження. Дослідження не слід розуміти як конкретне юридичне тлумачення положень права ЄС. Багато понять у сфері GenAI та авторського права залишаються новими, а їх остаточне тлумачення належить до компетенції судів та відповідних органів.

Також дослідження не є схваленням конкретного технічного рішення. Технічні механізми аналізуються з точки зору їхніх функцій, переваг, обмежень та потенційної ролі в екосистемі GenAI. Відсутність певного рішення у дослідженні не означає його нерелевантності.

4. Технічний контекст розвитку генеративного ШІ

Для розуміння питань авторського права у сфері GenAI важливо враховувати технічну логіку роботи таких систем. Генеративний штучний інтелект не генерує контент “із нічого”. Його здатність генерувати нові результати ґрунтується на складному процесі навчання моделей на великих масивах даних.

Загалом процес розробки та функціонування GenAI-систем можна описати через кілька етапів.

Етап	Зміст	Значення для авторського права
Data collection	Збір даних із різних джерел, включно з web scraping, API, датасетами, ліцензованими базами або синтетичними даними	Може охоплювати твори, об’єкти суміжних прав або бази даних
Data pre-processing	Очищення, структурування, фільтрація, токенизація, анотування даних	Можуть створюватися технічні копії або змінюватися формат даних
Model pre-training	Первинне навчання моделі на великих масивах даних	Виникає питання правового режиму використання навчального контенту
Fine-tuning	Донавчання моделі для конкретних завдань, секторів або сценаріїв використання	Може використовувати спеціалізовані або ліцензовані датасети
Reinforcement learning	Додаткове налаштування поведінки моделі, зокрема з використанням людської оцінки	Може передбачати використання додаткових input-даних
RAG	Поєднання генеративної моделі з пошуком або отриманням інформації із зовнішніх джерел	Контент може використовуватися не лише для навчання, а й під час генерації результату
Output generation	Генерування тексту, зображення, аудіо, відео, коду або іншого результату	Можливі питання щодо подібності до охоронюваного контенту та прозорості AI-generated content

Особливо важливим є етап збору даних. Саме він часто є початковою точкою виникнення ризиків у сфері авторського права, оскільки дані для навчання можуть збиратися з відкритого інтернету, включати охоронюваний контент або походити з джерел із невизначеним правовим статусом.

Також важливим є RAG – Retrieval-Augmented Generation (доповнене пошуком генерування). Такий підхід дозволяє ШІ-системам використовувати актуальну інформацію із зовнішніх джерел без постійного повного перенавчання моделі. З погляду авторського права це означає, що контент може бути релевантним не лише на етапі навчання моделі, а й під час отримання інформації із зовнішніх джерел або генерації результату.

Отже, авторсько-правова оцінка GenAI залежить від конкретного технічного етапу: чи йдеться про збір даних, навчання моделі, донавчання, RAG, генерацію результату або його подальше використання.

5. Правове регулювання генеративного ШІ у сфері авторського права в ЄС

Одним із центральних блоків дослідження EUIPO є аналіз правового регулювання Європейського Союзу. У цьому контексті особливе значення мають два акти права ЄС: [Директива \(ЄС\) 2019/790 про авторське право і суміжні права на Єдиному цифровому ринку \(CDSM Directive\)](#) та [Регламент \(ЄС\) 2024/1689 про встановлення гармонізованих правил щодо штучного інтелекту \(Artificial Intelligence Act, AI Act\)](#).

CDSM Directive передбачає положення щодо інтелектуального аналізу тексту і даних (text and data mining, TDM). TDM має особливе значення для GenAI, оскільки передбачає автоматизований аналіз великих масивів текстів, зображень, аудіо, відео та інших даних. У процесі TDM можуть здійснюватися технічні копії або інші дії з контентом, що потенційно зачіпають виключне право на відтворення, а також права власників баз даних.

AI Act, зі свого боку, встановлює регуляторні вимоги для ШІ-систем у ЄС і передбачає окремі обов'язки для провайдерів моделей штучного інтелекту загального призначення (General Purpose AI Models, GPAI), зокрема у частині дотримання авторського права та прозорості.

Акт	Основне значення для GenAI та авторського права
CDSM Directive	Встановлює винятки для інтелектуального аналізу тексту і даних (TDM); стаття 3 стосується TDM для наукових досліджень, стаття 4 – ширшого TDM з можливістю opt-out
AI Act	Передбачає обов'язки GPAI-провайдерів щодо дотримання opt-out, публікації достатньо детальних резюме навчальних даних і забезпечення технічної можливості виявлення результатів, згенерованих за допомогою ШІ-систем

Стаття 3 CDSM Directive передбачає виняток для TDM, що здійснюється дослідницькими організаціями та установами культурної спадщини для цілей наукових досліджень. Цей виняток має спеціальне спрямування і покликаний підтримувати наукову діяльність та доступ до знань.

Стаття 4 CDSM Directive передбачає ширший виняток для TDM, який може застосовуватися різними суб'єктами, включно з комерційними користувачами. Водночас цей виняток залежить від того, чи правовласник зарезервував свої права. Для контенту, доступного онлайн, таке резервування має бути здійснене прямо, належним чином, у тому числі за допомогою засобів, придатних для автоматизованого зчитування. Якщо правовласник здійснив opt-out, використання відповідного контенту для TDM потребує авторизації правовласника, зокрема шляхом ліцензування.

Ці положення є одним із ключових для GenAI. Якщо правовласники мають ефективні способи резервування прав, це може створювати передумови для розвитку ліцензійного ринку. Якщо ж opt-out механізми залишаються фрагментованими, технічно складними або недостатньо зрозумілими, це ускладнює як захист правовласників, так і правову визначеність для ШІ-розробників.

AI Act доповнює відповідне регулювання у частині вимог до ШІ-систем. Провайдери моделей загального призначення мають дотримуватися політики щодо авторського права, зокрема

враховувати opt-out резервування, здійснені відповідно до статті 4 CDSM Directive. Також передбачено обов'язок публікувати достатньо детальні резюме щодо контенту, використаного для навчання моделей. Така прозорість має допомагати правовласникам оцінювати можливе використання їхнього контенту та за потреби захищати свої права.

Крім того, AI Act передбачає вимоги щодо виявлення AI-generated content. Результати, згенеровані або змінені за допомогою ШІ-систем, мають бути технічно виявлюваними у машиночитаному форматі, наскільки це можливо з огляду на стан технологій. Такі вимоги пов'язані не лише з авторським правом, а й із ширшими питаннями довіри до цифрового середовища, протидії дезінформації та прозорості.

6. Використання контенту для навчання моделей генеративного ШІ

На етапі GenAI Input центральним є питання використання контенту для навчання моделей. Саме тут виникає один із найбільш чутливих конфліктів між інтересами ШІ-розробників і правовласників.

Для ШІ-розробників доступ до великих обсягів даних є необхідною умовою створення якісних моделей. Чим ширший, якісніший і різноманітніший датасет, тим більше можливостей для навчання моделі. Водночас правовласники звертають увагу на те, що їхні твори, виконання, фонограми, публікації, фотографії, ілюстрації, тексти, бази даних або інші об'єкти можуть використовуватися без дозволу, без винагороди та без достатньої прозорості.

Значна частина навчальних даних збирається з відкрито доступного онлайн-контенту. Проте відкритість доступу не означає автоматичної свободи використання для будь-яких цілей. Контент може бути доступним на сайті, але водночас залишатися охоронюваним авторським правом або суміжними правами. Крім того, умови використання сайту можуть прямо обмежувати scraping, автоматизований доступ або використання матеріалів для ШІ-навчання.

Дослідження EUIPO підкреслює, що процес збору даних має здійснюватися з дотриманням вимог законодавства про авторське право. Залежно від ситуації це може означати необхідність дотримання opt-out резервувань, отримання ліцензій, перевірку правового статусу датасетів, врахування умов доступу до контенту та застосування інструментів вебскрейпінгу у спосіб, що не порушує права правовласників.

Окремою проблемою є правовий статус великих публічних датасетів. Деякі з них можуть містити матеріали з невизначеним або некоректно зазначеним ліцензійним статусом. Інші можуть включати контент, який був зібраний без належного дозволу або походить із сумнівних джерел. У таких випадках ризик не завжди залишається лише на рівні того, хто створив датасет. Він може передаватися далі по ланцюгу – до розробника моделі, постачальника ШІ-сервісу або суб'єкта, який розгортає відповідну систему.

Для бізнесу та публічних установ це означає, що використання GenAI-сервісів потребує уважного аналізу умов конкретного інструменту. Важливо розуміти, які гарантії надає провайдер, чи є політика дотримання авторського права, чи передбачені договірні гарантії захисту від претензій третіх осіб (indemnification clauses), які обмеження встановлені в Terms & Conditions і чи дозволяється комерційне використання результатів.

Таким чином, GenAI Input – це не лише технічний етап, а складна правова зона, де перетинаються авторське право, договірне право, регулювання ШІ, питання прозорості, відповідальності та управління ризиками.

7. Резервування прав та механізми opt-out

Одним із практично важливих блоків дослідження є аналіз механізмів резервування прав, або opt-out. Для правовласників це один із ключових інструментів контролю за використанням їхнього контенту в межах TDM та потенційно для навчання ШІ-моделей.

EUIPO розмежовує юридичні та технічні механізми резервування прав.

Тип механізму	Приклади	Переваги	Обмеження
Юридичні механізми	Односторонні заяви правовласників, ліцензійні обмеження, умови використання вебсайтів	Дозволяють прямо зафіксувати позицію правовласника щодо використання контенту	Не завжди є зручними для автоматизованого зчитування ШІ-розробниками
Механізми, прив'язані до місця розміщення контенту (location-based)	robots.txt, TDM Reservation Protocol, robots meta tags	Відносно прості для впровадження на рівні вебсайту	Залежать від адміністратора сайту та не завжди “супроводжують” сам контент
Механізми, прив'язані до цифрового об'єкта (asset-based)	C2PA, JPEG Trust, рішення на основі метаданих і відстеження походження контенту	Можуть бути пов'язані безпосередньо з цифровим файлом або об'єктом	Потребують стандартизації, сумісності та технічного впровадження
Інфраструктурні рішення	Реєстри, платформи управління правами, федеровані бази даних	Потенційно можуть полегшити масштабне управління opt-out резервуваннями	Перебувають на різних стадіях розвитку і потребують довіри між учасниками

До юридичних механізмів належать односторонні заяви правовласників, положення ліцензійних договорів, умови використання вебсайтів та інші формулювання, за допомогою яких правовласник прямо зазначає, що його контент не може використовуватися для певних цілей без відповідної авторизації. Такі механізми можуть бути особливо важливими для видавців, музичної індустрії, фотостоків, медіа, наукових видавництв, платформ і власників великих каталогів контенту.

До технічних механізмів належать інструменти, які дозволяють передавати інформацію про резервування прав у форматі, придатному для автоматизованого зчитування. Серед них дослідження розглядає Robots Exclusion Protocol, TDM Reservation Protocol, robots meta tags, C2PA, JPEG Trust та інші рішення, що використовують метадані, відстеження походження контенту, а також location-based або asset-based підходи.

Location-based механізми пов'язані з конкретним місцем розміщення контенту, наприклад із певним сайтом або URL. Вони можуть бути простішими для впровадження на рівні вебресурсу, але залежать від адміністратора сайту і не завжди “супроводжують” контент, якщо він поширюється в інших місцях.

Asset-based механізми пов'язані безпосередньо із самим цифровим об'єктом. Їхня перевага полягає в тому, що інформація про права може бути прив'язана до файлу або іншого

цифрового об'єкта незалежно від конкретного місця його розміщення. Водночас такі рішення можуть бути складнішими для впровадження, потребують стандартизації, сумісності та надійної перевірки правового статусу особи, яка здійснює таке резервування.

Важливий висновок EUIPO полягає в тому, що наразі не існує одного універсального opt-out механізму. Правовласники використовують комбінацію юридичних і технічних заходів, а ефективність кожного з них залежить від типу контенту, сектору, технічної інфраструктури, ринкової зрілості рішення та готовності ШІ-розробників враховувати такі сигнали.

Robots Exclusion Protocol наразі виконує роль де-факто стандарту для управління вебкраулінгом і вебскрейпінгом, однак дослідження обережно підкреслює, що він не є оптимальним інструментом TDM opt-out. Його обмеження пов'язані з недостатньою гранулярністю, добровільним характером виконання, залежністю від адміністраторів сайтів і необхідністю ідентифікації автоматизованих систем. Крім того, REP не завжди дозволяє чітко розмежувати різні цілі використання: індексацію пошуковими системами, навчання ШІ-моделей, RAG, архівування або інші автоматизовані дії.

Тому більш реалістичним є поступовий розвиток комбінованих і стандартизованих підходів. У майбутньому важливу роль можуть відігравати відкриті стандарти, сумісні технічні рішення, реєстри opt-out резервувань, а також зрозумілі рекомендації для правовласників і ШІ-компаній.

8. Додаткові технічні засоби захисту контенту

Повне дослідження окремо аналізує не лише механізми резервування прав, а й так звані non-reservation solutions – рішення, які не завжди прямо виражають opt-out, але можуть допомагати правовласникам зменшувати ризики несанкціонованого використання контенту або контролювати доступ до нього.

До таких рішень належать, зокрема, protective perturbations, crawler blockers та digital fingerprinting.

Рішення	Загальна характеристика	Потенційне значення
Protective perturbations	Технічні зміни у файлах, які можуть ускладнювати використання контенту для навчання моделей	Можуть бути актуальними для візуальних творів і художників
Crawler blockers	Технічне обмеження доступу автоматизованих збирачів даних до вебресурсів	Може зменшувати ризик масового scraping
Digital fingerprinting	Формування цифрових відбитків творів або іншого контенту	Може допомагати виявляти копії або випадки використання контенту

Такі інструменти не слід сприймати як повну заміну правовим механізмам. Їхня роль радше допоміжна: вони можуть посилювати контроль правовласника, ускладнювати небажане використання контенту або допомагати у виявленні можливих порушень. Водночас вони також мають технічні обмеження і не завжди можуть бути застосовані до всіх типів контенту.

Включення таких рішень у повне дослідження є важливим, оскільки показує: проблема GenAI та авторського права не зводиться лише до питання “дозволити або заборонити”. На практиці правовласники можуть потребувати ширшого набору інструментів – від договорів і

метаданих до технічних бар'єрів, маркування, моніторингу та виявлення використання контенту.

9. Ліцензування контенту для систем генеративного ШІ

Паралельно з розвитком механізмів резервування прав формується ринок ліцензування контенту для потреб систем генеративного ШІ. Це один із найбільш перспективних, але водночас складних напрямів, який дослідження EUIPO розглядає особливо уважно.

Ліцензування може бути способом збалансувати інтереси правовласників і розробників систем ШІ. Для правовласників воно створює потенційні нові джерела доходів. Для ШІ-компаній – зменшує правові ризики та забезпечує доступ до якісного, структурованого, перевіреного контенту. Для ринку загалом – сприяє більш прозорій і передбачуваній моделі використання охоронюваних матеріалів.

Дослідження звертає увагу, що попит на якісний контент може зростати через кілька факторів. По-перше, розробники систем ШІ потребують великих масивів даних, але не всі дані однаково корисні. Високоякісний контент із надійними метаданими, структурою та правовою визначеністю може мати більшу цінність, ніж масовий неструктурований вебскрейпінг. По-друге, із розвитком ринку та зростанням кількості судових спорів ШІ-компанії можуть бути більш зацікавлені у зменшенні правових ризиків. По-третє, для окремих застосувань, зокрема генерації з пошуковим доповненням (RAG), важливою є не лише кількість, а й актуальність, достовірність і авторитетність контенту.

Особливе місце можуть посідати преса, медіа та наукові видавництва. Їхній контент часто є актуальним, структурованим, професійно створеним і важливим для інформаційних сервісів. У контексті RAG та сервісів відповідей такий контент може бути цінним не лише для навчання моделей, а й для надання користувачам актуальних відповідей.

Водночас ринок ліцензування ще не можна вважати повністю сформованим. Залишаються відкритими питання ціноутворення, обсягу дозволеного використання, строків, території, можливості використання контенту для попереднього навчання моделей, донавчання або RAG, пов'язання винагороди з доходами ШІ-сервісів, прозорості використання контенту та доступу малих правовласників до таких ринків.

Окремим викликом є нерівність переговорної сили. Великі видавці, музичні каталоги, фотобанки або медіагрупи мають більше можливостей укласти угоди з ШІ-компаніями. Натомість малі правовласники, незалежні автори, фотографи, ілюстратори або невеликі видавці часто не мають достатніх ресурсів для моніторингу використання свого контенту, технічного впровадження резервування прав або переговорів із великими ШІ-компаніями.

Саме тому особливого значення можуть набувати колективні механізми управління правами, платформи-агрегатори та інституційні ініціативи. Вони можуть допомогти меншим правовласникам ефективніше заявляти про свої права, пропонувати контент для ліцензування, отримувати винагороду та брати участь у формуванні ринку даних для навчання ШІ-моделей.

10. Походження даних та ризики зміни контексту їх використання (data laundering)

Окреме місце у дослідженні займає питання так званого “data laundering” – ризику, коли дані або датасети, створені чи зібрані в межах однієї правової підстави, надалі використовуються в іншому контексті, який може вимагати інших умов або дозволів.

Наприклад, дослідницька організація може здійснювати TDM у межах винятку для наукових досліджень. Однак якщо створений нею датасет надалі використовується комерційним ШІ-розробником, виникає питання, чи може такий суб'єкт покладатися на ту саму правову підставу. Дослідження показує, що між створенням датасетів на попередніх етапах і їх подальшим комерційним використанням для ШІ-навчання потрібно проводити обережне розмежування.

Це питання особливо важливе тому, що датасети часто циркулюють між різними учасниками ланцюга розробки та використання ШІ-систем. Один суб'єкт збирає дані, інший структурує датасет, третій навчає модель, четвертий надає сервіс, п'ятий інтегрує цей сервіс у власний продукт. У такому складному ланцюгу не завжди очевидно, хто і на якому етапі має перевіряти правовий статус даних.

З практичної точки зору це означає, що ШІ-компаніям і користувачам ШІ-сервісів дедалі важливіше мати належну документацію щодо походження даних, умов їх використання, наявності ліцензій, opt-out резервувань і політик дотримання авторського права. Для правовласників важливими стають інструменти фіксації та передачі інформації про права, зокрема засоби, придатні для автоматизованого зчитування, метадані та інші технічні інструменти.

Для публічних інституцій це також створює новий виклик: необхідно не лише інформувати про права, а й сприяти розвитку зрозумілих стандартів, практик належної перевірки, прозорості ланцюгів використання даних і механізмів діалогу між різними групами учасників.

11. Прозорість та ідентифікація результатів генеративного ШІ

Другий великий блок дослідження присвячений GenAI Output – результатам, згенерованим або зміненим за допомогою систем генеративного штучного інтелекту.

У цій частині виникають два основні питання. Перше – як ідентифікувати контент, згенерований або змінений за допомогою ШІ. Друге – як зменшити ризик генерування результатів, які можуть порушувати права третіх осіб.

Питання прозорості AI-generated content має значення не лише для авторського права. Воно також пов'язане з довірою до інформації, протидією дезінформації, боротьбою з deepfakes, захистом споживачів і загальною прозорістю цифрового середовища. Однак у сфері авторського права прозорість має окреме значення, оскільки допомагає відрізнити людську творчість від контенту, згенерованого або істотно зміненого ШІ-системою, а також може бути корисною для виявлення можливих порушень.

Напрямок	Приклади рішень	Значення
Provenance tracking	C2PA, JPEG Trust, TRACE4EU	Допомагає фіксувати походження та історію змін цифрового активу
Watermarking	Цифрове маркування AI-generated content	Сприяє виявленню штучно згенерованого або зміненого контенту
Fingerprinting	Цифрові відбитки контенту	Дозволяє порівнювати контент із базами даних

AI-generated content detection	Детектори ШІ-контенту, зокрема для зображень, аудіо або відео	Може допомагати у прозорості, але не гарантує повної точності результатів
Membership inference	Спроби визначити, чи використовувався конкретний об'єкт у навчанні моделі	Має потенційне значення для перевірки навчальних даних, але є технічно складним підходом
Output safeguards	Фільтри запитів, фільтри результатів, інструменти порівняння	Зменшують ризик генерування результатів, що можуть порушувати права третіх осіб

Відстеження походження контенту (provenance tracking) спрямоване на фіксацію походження контенту, історії його формування або змін. Такі рішення можуть допомагати встановити, чи був контент створений людиною, згенерований за допомогою ШІ або змінений із використанням певного інструменту.

Цифрове маркування (watermarking) передбачає внесення у контент спеціального сигналу або маркера, який дозволяє згодом визначити його ШІ-походження. Цифрові відбитки контенту (fingerprinting) використовуються для порівняння контенту з іншими файлами або базами даних. Виявлення контенту, згенерованого за допомогою ШІ (AI-generated content detection), ґрунтується на використанні алгоритмів для визначення ймовірності того, що контент був згенерований або змінений за допомогою ШІ.

Водночас дослідження підкреслює, що жоден із цих інструментів не є ідеальним. Метадані можуть бути видалені, цифрове маркування може бути вразливим до технічних змін, інструменти виявлення ШІ-контенту можуть давати помилкові результати, а цифрові відбитки залежать від наявності відповідної бази для порівняння. Крім того, різні типи контенту – текст, зображення, аудіо, відео, код – потребують різних технічних підходів.

Тому найбільш перспективним виглядає не одне універсальне рішення, а комбінація різних інструментів, стандартизація, сумісність між платформами та розвиток практик.

12. Ризики відтворення охоронюваного контенту у результатах генеративного ШІ

Одним із найбільш складних питань, пов'язаних із результатами, згенерованими за допомогою GenAI, є ризик того, що модель може згенерувати результат, занадто близький до охоронюваного контенту. Це пов'язано з явищами запам'ятовування навчальних даних (memorisation), відтворення навчального контенту (regurgitation) та генерування плагіатоподібних результатів (plagiaristic output).

Запам'ятовування навчальних даних (memorisation) означає, що модель може зберігати у своїх параметрах певні елементи навчальних даних. Це не означає, що будь-яке навчання є “запам'ятовуванням” у правовому сенсі. Більшість ШІ-моделей виявляє закономірності та узагальнює інформацію з великих масивів даних. Однак у певних випадках модель може відтворювати фрагменти навчального контенту настільки близько, що це створює правові ризики.

Відтворення навчального контенту (regurgitation) – це ситуація, коли модель генерує фрагмент контенту, який був у навчальних даних. Такий результат може бути проблемним,

якщо він становить відтворення охоронюваного твору або його суттєвої частини. Особливо чутливими можуть бути тексти, програмний код, ілюстрації, персонажі, музичні твори або інші виразні елементи.

Дослідження також звертає увагу на те, що ризики можуть виникати не лише через технічні особливості моделі, а й через дії користувачів. Користувач може намагатися спеціально отримати результат, подібний до відомого твору, персонажа, стилю або конкретного фрагмента. Тому провайдери GenAI-систем застосовують не лише технічні рішення всередині моделі, а й зовнішні запобіжники: фільтри запитів (prompt filters), фільтри результатів (output filters), інструменти автоматизованого порівняння (automated comparison tools), механізми відмови у виконанні запиту (refusal mechanisms) та інші засоби.

Водночас такі засоби мають обмеження. Фільтри можна обійти, подібність може бути складно оцінити автоматично, а надмірно жорсткі обмеження можуть впливати на законні способи використання ШІ-систем. Тому завдання полягає не в повному усуненні всіх ризиків, а у формуванні розумної системи управління ними.

13. Технічні та договірні механізми управління ризиками

Окремо дослідження розглядає технічні підходи до вилучення або нейтралізації інформації в моделі (model unlearning) та редагування моделі (model editing). Вилучення або нейтралізація інформації в моделі спрямовані на усунення впливу певної інформації, закодованої у моделі. Редагування моделі передбачає коригування її поведінки щодо конкретних фактів, відповідей або шаблонів.

Ці напрями активно досліджуються, однак їх не варто сприймати як повністю зрілі універсальні рішення для великих базових моделей (foundation models). Вони можуть бути перспективними, проте залишаються технічно складними, ресурсомісткими і не завжди гарантують точне видалення конкретного контенту без впливу на інші властивості моделі.

Окремим практичним інструментом є договірне зобов'язання щодо захисту від правових претензій або відшкодування (contractual indemnification). Йдеться про умови, за якими провайдер ШІ-сервісу може брати на себе частину ризиків, пов'язаних із використанням результатів GenAI.

Такі механізми особливо важливі для корпоративних користувачів. Якщо компанія використовує ШІ-сервіс для генерування контенту для комерційного використання, програмного коду, маркетингових матеріалів або інших результатів, вона може бути зацікавлена в тому, щоб провайдер взяв на себе частину відповідальності за можливі претензії.

Водночас такі договірні гарантії зазвичай мають умови та обмеження. Вони можуть не застосовуватися, якщо користувач порушив правила сервісу, обійшов фільтри, свідомо намагався отримати результат, який може порушувати права третіх осіб, використав власні вхідні дані (input) без належних прав або змінив результат (output) у спосіб, що спричинив порушення.

Тому такі договірні гарантії не замінюють належної правової оцінки. Вони є лише одним з елементів управління ризиками. Користувачам ШІ-сервісів важливо аналізувати умови використання сервісу (Terms & Conditions), політики комерційного використання, умови щодо прав на результати, обмеження щодо вхідних даних, наявність або відсутність гарантій, а також винятки з таких гарантій.

14. Судові спори та формування практики

Дослідження EUIPO звертає увагу на зростання кількості спорів між правовласниками та провайдерами GenAI. Найбільша кількість таких спорів наразі спостерігається у США, де одним із ключових питань є застосування доктрини добросовісного використання (fair use) до використання охоронюваного контенту для ШІ-навчання. Водночас відповідні справи також виникають у ЄС та інших юрисдикціях.

У європейському контексті особливу увагу привертає справа Kneschke vs. LAION: рішення Регіонального суду Гамбурга від 27 вересня 2024 року (LG Hamburg, Urteil vom 27.09.2024 – 310 O 227/23).¹ У дослідженні EUIPO ця справа описується як одне з перших помітних судових рішень у ЄС у приватному спорі щодо авторського права та ШІ-навчання.

У справі розглядалося використання фотографії у датасеті LAION. Суд дійшов висновку, що LAION могла посилатися на виняток для інтелектуального аналізу тексту і даних для цілей наукових досліджень відповідно до статті 3 CDSM Directive. Водночас окремі міркування суду, наведені як obiter dicta (побіжні зауваження, що не є частиною обов'язкової основи рішення), можуть бути важливими для подальшого розуміння того, як суди можуть підходити до оцінки резервування прав відповідно до статті 4 CDSM Directive.

Загалом судова практика у сфері GenAI лише формується. Різні юрисдикції мають різні підходи до винятків і обмежень авторського права, fair use, добросовісної практики (fair dealing), TDM, договірних обмежень, прав на бази даних та відповідальності платформ. Тому висновки, зроблені в одній юрисдикції, не завжди можуть бути прямо перенесені на іншу.

Напрямок спорів	Типові питання
Використання контенту для ШІ-навчання	Чи було використання дозволено; чи застосовуються винятки; чи було здійснено opt-out резервування
Відтворення контенту у результатах GenAI	Чи генерує модель результат, подібний до охоронюваного твору або його суттєвої частини
Дотримання умов вебсайтів	Чи порушує вебскрейпінг умови використання відповідного ресурсу
Прозорість навчальних даних	Чи може правовласник встановити, чи використовувався його контент для навчання моделі
Договірні та ліцензійні питання	Чи були потрібні ліцензії; який обсяг дозволеного використання

Водночас сам факт зростання кількості спорів має практичне значення. Він стимулює ШІ-компанії уважніше ставитися до походження даних, правовласників – активніше заявляти про свої права, а публічні інституції – шукати збалансовані механізми, які можуть зменшити конфліктність ринку.

Паралельно із судовими спорами розвиваються ліцензійні угоди. Це свідчить про те, що ринок не рухається виключно шляхом конфлікту: поряд із судовими процесами формується договірна практика, яка може стати більш гнучким механізмом врегулювання відносин між правовласниками та ШІ-розробниками.

¹ EUIPO, *The Development of Generative Artificial Intelligence from a Copyright Perspective*, May 2025, Section 2.3.1.1 “Kneschke vs. LAION (Germany)”; LG Hamburg, Urteil vom 27.09.2024 - 310 O 227/23. див. також: WIPO Lex, Robert Kneschke v. LAION e.V., Case No. 310 O 227/23.

15. Роль публічних інституцій та IP офісів

Однією з важливих частин дослідження є аналіз потенційної ролі публічних інституцій, зокрема IP офісів. У складній та швидкозмінній сфері GenAI такі інституції можуть виконувати не лише реєстраційні або інформаційні функції, а й ширшу роль – сприяти діалогу, підвищенню обізнаності та розвитку практичних інструментів.

Щодо використання контенту для навчання моделей, публічні інституції можуть допомагати правовласникам краще розуміти доступні механізми резервування прав. Це може включати інформаційні матеріали про opt-out, пояснення різниці між юридичними та технічними механізмами, огляди robots.txt, TDM Reservation Protocol, рішень на основі метаданих, C2PA, JPEG Trust та інших інструментів. Важливо, щоб така інформація була нейтральною і не сприймалася як підтримка або схвалення конкретного приватного рішення.

Також IP офіси можуть сприяти документуванню ринкових тенденцій: які технічні рішення використовуються, як розвиваються ліцензійні практики, які проблеми виникають у правовласників, які потреби мають ШІ-компанії. Це може допомогти формувати більш обґрунтовану політику у сфері ШІ та інтелектуальної власності.

Щодо результатів, згенерованих за допомогою GenAI, публічні інституції можуть інформувати про механізми ідентифікації ШІ-контенту, належні практики використання ШІ, ризики deepfakes, відповідальне формулювання запитів, а також перевірку результатів перед комерційним використанням. Також вони можуть сприяти підвищенню обізнаності користувачів щодо умов використання GenAI-сервісів.

Окрему роль може відігравати підтримка діалогу між стейкхолдерами. Автори, видавці, медіа, ШІ-розробники, платформи, технічні провайдери, наукові установи, юристи та державні органи часто бачать проблему з різних позицій. Без такого діалогу складно виробити рішення, яке одночасно підтримуватиме інновації та повагу до прав творців.

У дослідженні EUIPO також зазначалося про намір запустити Copyright Knowledge Centre до кінця 2025 року². Очікувалось, що він надаватиме практичну інформацію правовласникам, стане платформою для діалогу між стейкхолдерами та сприятиме обговоренню ефективних механізмів управління правами у сфері GenAI.

16. Значення дослідження для України та можливість проведення національного дослідження

Дослідження може бути корисним орієнтиром для України у контексті подальшого розвитку підходів до захисту авторського права та суміжних прав у цифровому середовищі. Генеративний штучний інтелект уже впливає на діяльність авторів, видавців, медіа, освітніх установ, IT-компаній, креативних індустрій, користувачів цифрових сервісів та публічних інституцій.

У зв'язку з цим питання використання охоронюваного контенту для навчання ШІ-систем, прозорості AI-generated content, ліцензування, відповідального використання GenAI та запобігання порушенням прав набувають практичного значення і для українського правового та економічного середовища.

З урахуванням євроінтеграційного курсу України, аналіз підходів ЄС до GenAI та авторського права може сприяти кращому розумінню тенденцій, які формуються на рівні Європейського Союзу. Зокрема, важливими для національного контексту є питання text and data mining,

² На момент підготовки цього огляду Copyright Knowledge Centre вже функціонує як окремий інформаційний ресурс EUIPO. Сторінка Центру доступна за посиланням: <https://www.euipo.europa.eu/en/copyright-knowledge-centre>

механізмів резервування прав, договірних моделей ліцензування контенту для ШІ-систем, прозорості навчальних даних, ідентифікації AI-generated content, а також ролі публічних інституцій у підвищенні обізнаності правовласників і користувачів.

У цьому контексті доцільним може бути проведення окремого національного дослідження щодо впливу генеративного штучного інтелекту на сферу авторського права та суміжних прав в Україні.

Предметом такого дослідження могли б стати, зокрема:

Напрямок дослідження	Можливі питання для аналізу
Правове регулювання	Наскільки чинне законодавство України дозволяє реагувати на виклики GenAI; які питання потребують додаткового роз'яснення або аналізу
Використання охоронюваного контенту для навчання ШІ-систем	Як українські автори, видавці, медіа, музична індустрія, фотобанки та інші правовласники оцінюють ризики використання їхнього контенту для ШІ-навчання
Обізнаність правовласників	Чи знають правовласники про opt-out, ліцензування, технічні засоби маркування та управління правами, право особливого роду (sui generis)
Ліцензування контенту для ШІ-систем	Чи існує потенціал для формування українських практик ліцензування контенту для ШІ-систем
Контент, згенерований за допомогою ШІ (AI-generated content)	Які ризики виникають щодо маркування, використання, комерціалізації та правового статусу контенту, згенерованого за допомогою ШІ
Потреби бізнесу та користувачів	Яких роз'яснень потребують компанії, що використовують GenAI у маркетингу, дизайні, освіті, медіа, ІТ та інших сферах
Роль публічних інституцій	Які інформаційні, аналітичні або просвітницькі інструменти могли б бути корисними для правовласників, користувачів і ШІ-розробників

Таке дослідження могло б мати практичну цінність одразу для кількох груп. Для правовласників воно допомогло б краще зрозуміти можливі способи управління своїми правами у середовищі GenAI. Для бізнесу – оцінити ризики використання ШІ-сервісів, зокрема щодо input, output, договірних умов і комерційного використання результатів. Для публічних інституцій – визначити напрями інформаційної, аналітичної та просвітницької роботи. Для законодавчої та експертної спільноти – окреслити питання, які можуть потребувати подальшого обговорення у контексті гармонізації з європейськими підходами.

Окремим результатом такого дослідження могла б стати підготовка практичних рекомендацій для українських правовласників щодо використання договорів, умов вебсайтів, метаданих, технічних засобів маркування, застережень щодо ШІ-навчання та інших механізмів управління правами. Також доцільним виглядає напрацювання інформаційних матеріалів для користувачів GenAI-сервісів щодо відповідального використання AI-generated content, перевірки умов сервісів, комерційного використання результатів та уникнення порушення прав третіх осіб.

Таким чином, дослідження EUIPO може слугувати не лише джерелом інформації про європейські підходи, а й додатковим орієнтиром для продовження формування українського аналітичних підходів у сфері генеративного штучного інтелекту та авторського права .

17. Оцінка дослідження

Дослідження поєднує технічний, правовий та економічний аналіз, що є необхідним для розуміння GenAI. Питання авторського права неможливо оцінити без розуміння того, як працюють моделі, як збираються дані, як формуються результати та які технічні інструменти використовуються для управління ризиками.

Водночас дослідження має практичну спрямованість. Воно не обмежується загальним описом ризиків, а аналізує конкретні категорії рішень: механізми opt-out, заходи резервування прав, ліцензування, інструменти прозорості, технічні підходи до вилучення або коригування інформації в моделі, договірні гарантії та можливу підтримку з боку публічних інституцій. Це робить дослідження корисним не лише для науковців, а й для правників, правовласників, ШІ-компаній і публічних інституцій.

У документі підкреслюється, що правова практика у сфері GenAI лише формується, а технічні рішення мають власні обмеження. Жоден із розглянутих механізмів не подається як універсальне рішення для всіх ситуацій. Такий підхід є методологічно коректним для сфери, яка швидко змінюється.

Водночас дослідження має певні обмеження. Його значний обсяг може бути складним для широкої аудиторії, а багато технічних розділів потребують спеціальних знань. Крім того, сфера GenAI розвивається дуже швидко, тому окремі ринкові практики, технічні інструменти або судові справи можуть змінюватися вже після публікації.

Також важливо, що дослідження зосереджене передусім на праві ЄС. Для інших юрисдикцій його висновки можуть бути корисними як орієнтир, але не завжди можуть застосовуватися безпосередньо. Це особливо важливо для України, де необхідно враховувати власне законодавство, практику, інституційний контекст та євроінтеграційні процеси.

18. Висновки

Дослідження є важливим аналітичним документом, який системно розглядає взаємодію генеративного штучного інтелекту з авторським правом. Його цінність полягає в тому, що воно поєднує аналіз технологій, права, ринку та інституційних механізмів.

Ключовий висновок полягає в тому, що розвиток GenAI потребує переосмислення практичних підходів до застосування авторського права з урахуванням нових технологічних способів створення, використання та поширення контенту. Основні виклики пов'язані з використанням охоронюваного контенту для навчання моделей, ефективністю opt-out механізмів, розвитком ліцензійних ринків, прозорістю AI-generated content та запобіганням результатам, які можуть порушувати права третіх осіб.

Водночас дослідження не пропонує одного універсального рішення. Навпаки, воно демонструє, що у сфері GenAI потрібне поєднання різних механізмів: правових, технічних, договірних, організаційних та інформаційних. Саме така багаторівнева система може сприяти балансу між розвитком інновацій і належною охороною прав авторів та інших правовласників.

Для України це дослідження може бути корисним як орієнтир для подальшого формування підходів до захисту авторського права у цифровому середовищі, розвитку практик ліцензування контенту для ШІ-систем, підвищення обізнаності правовласників і користувачів,

а також визначення ролі публічних інституцій у сфері штучного інтелекту та інтелектуальної власності.

Таким чином, дослідження демонструє, що ефективне реагування на виклики GenAI потребує не ізольованого правового або технічного рішення, а комплексного підходу, який поєднує регулювання, ліцензування, прозорість, технічні стандарти та інституційну підтримку.

Примітка щодо характеру огляду

*Цей матеріал є інформаційно-аналітичним оглядом дослідження EUIPO *The Development of Generative Artificial Intelligence from a Copyright Perspective* та підготовлений з метою узагальнення його основних положень для кращого розуміння проблематики генеративного ШІ та авторського права.*

Огляд не є офіційним перекладом дослідження EUIPO, офіційним тлумаченням положень права ЄС або України, а також не становить юридичної консультації щодо конкретної ситуації. Наведені формулювання, висновки та українські відповідники окремих термінів мають інформаційно-пояснювальний характер і мають оцінюватися з урахуванням контексту оригінального англомовного дослідження.

Інформація, наведена в огляді, відображає зміст дослідження EUIPO та стан відповідних джерел на момент підготовки матеріалу.

Глосарій ключових термінів

Терміни у глосарії наведено відповідно до логіки та термінології основного дослідження. Англomовні терміни та аббревіатури подано для зручності пошуку й зв'язки з оригінальним текстом. Для найбільш коректного розуміння змісту окремих термінів доцільно враховувати їх використання в англomовному тексті дослідження.

1. Загальні поняття у сфері штучного інтелекту

Термін в оригіналі	Український відповідник	Пояснення
Generative AI / GenAI	Генеративний штучний інтелект	Різновид штучного інтелекту, здатний генерувати новий контент, зокрема текст, зображення, аудіо, відео, програмний код або інші результати.
General Purpose AI Model / GPAI	Модель ШІ загального призначення	ШІ-модель, яка може виконувати широкий спектр завдань і бути інтегрована в різні системи, сервіси або продукти.
Large Language Model / LLM	Велика мовна модель	ШІ-модель, навчена на великих масивах текстових даних і здатна генерувати, аналізувати, узагальнювати або трансформувати текст.
Generative Pre-trained Transformer / GPT	Генеративний попередньо навчений трансформер	Тип великої мовної моделі, побудований на архітектурі трансформера та попередньо навчений на великих масивах даних.
Artificial Neural Network / ANN	Нейронна мережа	Архітектура ШІ-моделі, що складається з пов'язаних між собою елементів, які обробляють інформацію на основі вхідних даних.
Transformer	Трансформер	Архітектура ШІ-моделі, яка використовується для обробки послідовностей даних, зокрема тексту, і лежить в основі багатьох сучасних великих мовних моделей.
Attention Mechanism	Механізм уваги	Технічний механізм, який дає моделі змогу зосереджуватися на найбільш релевантних частинах вхідних даних під час формування результату.
Generative Adversarial Network / GAN	Генеративно-змагальна мережа	Тип генеративної моделі, у якій один компонент генерує нові зразки, а інший оцінює їхню схожість із реальними даними.
Variational Autoencoder / VAE	Варіаційний автокодувальник (автоенкодер)	Тип генеративної моделі, що стискає дані у внутрішнє представлення та використовує його для генерування нових зразків.
Diffusion Model	Дифузійна модель	Генеративна модель, яка генерує зображення або інший контент шляхом поступового перетворення випадкового шуму на структурований результат.

Graphics Processing Unit / GPU	Графічний процесор	Обчислювальний процесор, який часто використовується для навчання ШІ-моделей завдяки здатності швидко обробляти великі масиви даних.
--------------------------------	--------------------	--

2. Дані та етапи навчання ШІ-моделей

Термін в оригіналі	Український відповідник	Пояснення
Training Data	Навчальні дані	Дані, які використовуються для навчання ШІ-моделі та формування її здатності генерувати результати.
Dataset	Датасет / набір даних	Структурований або неструктурований набір даних, який може використовуватися для навчання, тестування, донавчання або оцінки ШІ-моделі.
Data Collection	Збір даних	Етап, на якому дані отримуються з різних джерел, зокрема вебсайтів, баз даних, API, ліцензованих ресурсів або спеціально створених наборів.
Data Pre-processing	Попередня обробка даних	Технічна підготовка даних до навчання моделі, зокрема очищення, структурування, фільтрація, токенізація або анотування.
Tokenisation	Токенізація	Поділ тексту або інших даних на менші одиниці – токени, які модель може обробляти.
Pre-training	Попереднє навчання моделі	Початковий етап навчання ШІ-моделі на великих масивах даних для формування загальних закономірностей.
Fine-tuning	Донавчання моделі	Додаткове навчання вже розробленої моделі на спеціалізованих даних для виконання конкретних завдань.
Reinforcement Learning	Навчання з підкріпленням	Метод навчання, за якого модель коригує поведінку на основі отриманих оцінок, сигналів або винагород.
Reinforcement Learning with Human Feedback / RLHF	Навчання з підкріпленням на основі людського зворотного зв'язку	Підхід, за якого поведінка моделі додатково налаштовується з урахуванням оцінок або переваг людей.
Retrieval Augmented Generation / RAG	Генерація з пошуковим доповненням	Підхід, за якого генеративна модель поєднується з пошуком або отриманням інформації із зовнішніх джерел для формування відповіді чи результату.

3. Правові поняття та регулювання у сфері авторського права

Термін в оригіналі	Український відповідник	Пояснення
Collective Management Organisation / CMO	Організація колективного управління	Організація, яка управляє майновими правами правовласників на колективній основі.
Text and Data Mining / TDM	Інтелектуальний аналіз тексту і даних	Автоматизований аналіз тексту і даних у цифровій формі для виявлення закономірностей, зв'язків, тенденцій або нової інформації.
CDSM Directive	Директива (ЄС) 2019/790 про авторське право і суміжні права на Єдиному цифровому ринку	Акт права ЄС, який, зокрема, встановлює правила щодо text and data mining та можливості резервування прав правовласниками.
AI Act	Регламент (ЄС) 2024/1689 про встановлення гармонізованих правил щодо штучного інтелекту	Регламент ЄС, який встановлює правила для ШІ-систем, зокрема обов'язки провайдерів моделей загального призначення щодо прозорості та дотримання авторського права.
Opt-out / Rights Reservation	Резервування прав	Механізм, за допомогою якого правовласник прямо зазначає, що його контент не може використовуватися для TDM або ШІ-навчання без відповідної авторизації.
Authorisation	Авторизація / дозвіл правовласника	Надання правовласником дозволу на використання охоронюваного контенту, зокрема шляхом ліцензування.
Licence / Licensing	Ліцензія / ліцензування	Дозвіл правовласника на використання об'єкта права інтелектуальної власності на визначених умовах.
Terms of Service	Умови використання сервісу	Умови використання вебсайту або сервісу, які можуть містити обмеження щодо scraping, ШІ-навчання, комерційного використання або відповідальності користувача.
Technological Protection Measures / TPM	Технічні засоби захисту	Технічні інструменти, які можуть обмежувати доступ до охоронюваного контенту або його використання.

4. Доступ до онлайн-контенту та резервування прав

Термін в оригіналі	Український відповідник	Пояснення
Web Scraping	Вебскрейпінг	Автоматизований збір інформації з вебсайтів шляхом отримання та аналізу їхнього вмісту.
Web Crawling	Вебкраулінг	Автоматизований обхід вебсторінок, зокрема для пошуку, індексації або збору інформації.

Robots Exclusion Protocol / REP	Протокол виключення роботів / управління доступом роботів	Технічний протокол, який використовується для управління доступом пошукових та інших автоматизованих систем до вебсайту
robots.txt	Файл robots.txt	Текстовий файл на вебсайті, за допомогою якого адміністратор може вказати, які частини сайту не слід сканувати автоматизованим системам.
TDM Reservation Protocol / TDMRep	Протокол резервування прав для TDM	Технічний підхід до вираження резервування прав щодо інтелектуального аналізу тексту і даних у форматі, придатному для автоматизованого зчитування.
Robots Meta Tags	Метатеги для роботів	HTML-елементи, які можуть містити інструкції для автоматизованих систем щодо індексації або обробки вебсторінки.
Location-based Mechanism	Механізм, прив'язаний до місця розміщення контенту	Технічний або організаційний підхід, пов'язаний із конкретним вебсайтом, сторінкою або URL, де розміщено контент.
Asset-based Mechanism	Механізм, прив'язаний до цифрового об'єкта	Підхід, за якого інформація про права або походження пов'язується безпосередньо з цифровим файлом чи об'єктом, незалежно від місця його розміщення.
Metadata	Метадані	Інформація про дані, зокрема про файл, автора, дату створення, права, джерело, технічні характеристики або історію змін.
Data Hash	Хеш даних	Результат криптографічного перетворення даних, який допомагає перевіряти цілісність файлу або виявляти його зміну.
Digital Signature	Цифровий підпис	Технічний механізм, який використовується для перевірки автентичності та цілісності цифрового документа або повідомлення.
International Standard Content Code / ISCC	Міжнародний стандартний код контенту	Ідентифікатор цифрового контенту, який може використовуватися для розпізнавання, опису або управління цифровими об'єктами.
International Press Telecommunications Council / IPTC	Міжнародна рада з преси та телекомунікацій	Організація, пов'язана зі стандартами метаданих та обміну інформацією у сфері медіа й новинного контенту.

5. Прозорість, походження та маркування ШІ-контенту

Термін в оригіналі	Український відповідник	Пояснення
AI-generated Content	Контент, згенерований штучним інтелектом	Контент, згенерований або істотно змінений за допомогою системи штучного інтелекту.
Synthetic Content	Синтетичний контент	Ширше поняття для штучно згенерованого цифрового контенту, зокрема контенту, згенерованого або зміненого за допомогою ШІ. (<i>AI-generated content – це один із різновидів synthetic content, але synthetic content може охоплювати й інші штучно згенеровані або змінені цифрові матеріали</i>)
Provenance Tracking	Відстеження походження контенту	Технічний підхід до фіксації походження цифрового контенту та історії його створення або змін.
C2PA Content Authenticity Initiative	Ініціатива C2PA щодо автентичності контенту	Технічна ініціатива, спрямована на фіксацію походження цифрового контенту, історії його створення та змін.
JPEG Trust	JPEG Trust	Стандарт / ініціатива, пов'язана з довірою до цифрових медіа, зокрема з фіксацією інформації про походження, автентичність і права щодо контенту.
Watermarking	Цифрове маркування	Технічне маркування контенту, яке може допомагати визначити, що він був згенерований або змінений за допомогою ШІ.
Fingerprinting	Цифровий відбиток	Формування унікального цифрового відбитка контенту для його подальшого розпізнавання, порівняння або виявлення використання.
AI-generated Content Detection	Виявлення контенту, згенерованого ШІ	Технічні засоби або алгоритми, спрямовані на визначення ймовірності того, що контент був згенерований або змінений за допомогою ШІ.
Deepfake	Deepfake / діпфейк	Штучно згенерований або змінений за допомогою ШІ - контент, який може створювати враження, що певна особа сказала або зробила те, чого насправді не було.
Prompt	Промпт / запит користувача	Інструкція, текст або інший вхідний запит, який користувач надає ШІ-системі для отримання результату.
Prompt Filtering	Фільтрація промπτів / запитів	Механізм, який обмежує або блокує запити, що можуть призвести до небажаних або потенційно порушуючих результатів.
Output Filtering	Фільтрація результатів	Перевірка або блокування результатів, які можуть бути незаконними, порушувати права третіх осіб або не відповідати правилам сервісу.

6. Ризики відтворення контенту та механізми управління ними

Термін в оригіналі	Український відповідник	Пояснення
Memorisation	Запам'ятовування навчальних даних	Технічне явище, за якого ШІ-модель може зберігати елементи навчальних даних і в окремих випадках відтворювати їх у результатах.
Regurgitation	Відтворення навчального контенту у результатах	Ситуація, коли ШІ-модель фактично відтворює фрагмент контенту, який був у навчальних даних.
Plagiaristic Output	Плагіатоподібний результат	Результат GenAI, який може бути надто подібним до охоронюваного твору або відтворювати його суттєві елементи.
Content Regurgitation	Відтворення контенту у вихідних результатах	Явище, коли ШІ-система генерує результат, який може прямо або дуже близько відтворювати контент із навчальних даних.
Membership Inference Attack	Атака на встановлення участі в навчальних даних	Технічний підхід, спрямований на з'ясування, чи використовувався конкретний об'єкт або запис у навчанні ШІ-моделі.
Model Unlearning	Вилучення / нейтралізація інформації в моделі	Технічний підхід, спрямований на видалення або нейтралізацію певної інформації, закодованої у моделі.
Model Editing	Редагування моделі	Технічний підхід до коригування поведінки моделі або оновлення певної інформації без повного перенавчання моделі.
Data Laundering	Відмивання даних / неправомірна зміна контексту використання даних	Ризик, коли дані або датасети, сформовані в межах однієї правової підстави, надалі використовуються в іншому контексті, який може вимагати інших умов або дозволів.