



Міністерство
цифрової трансформації
України

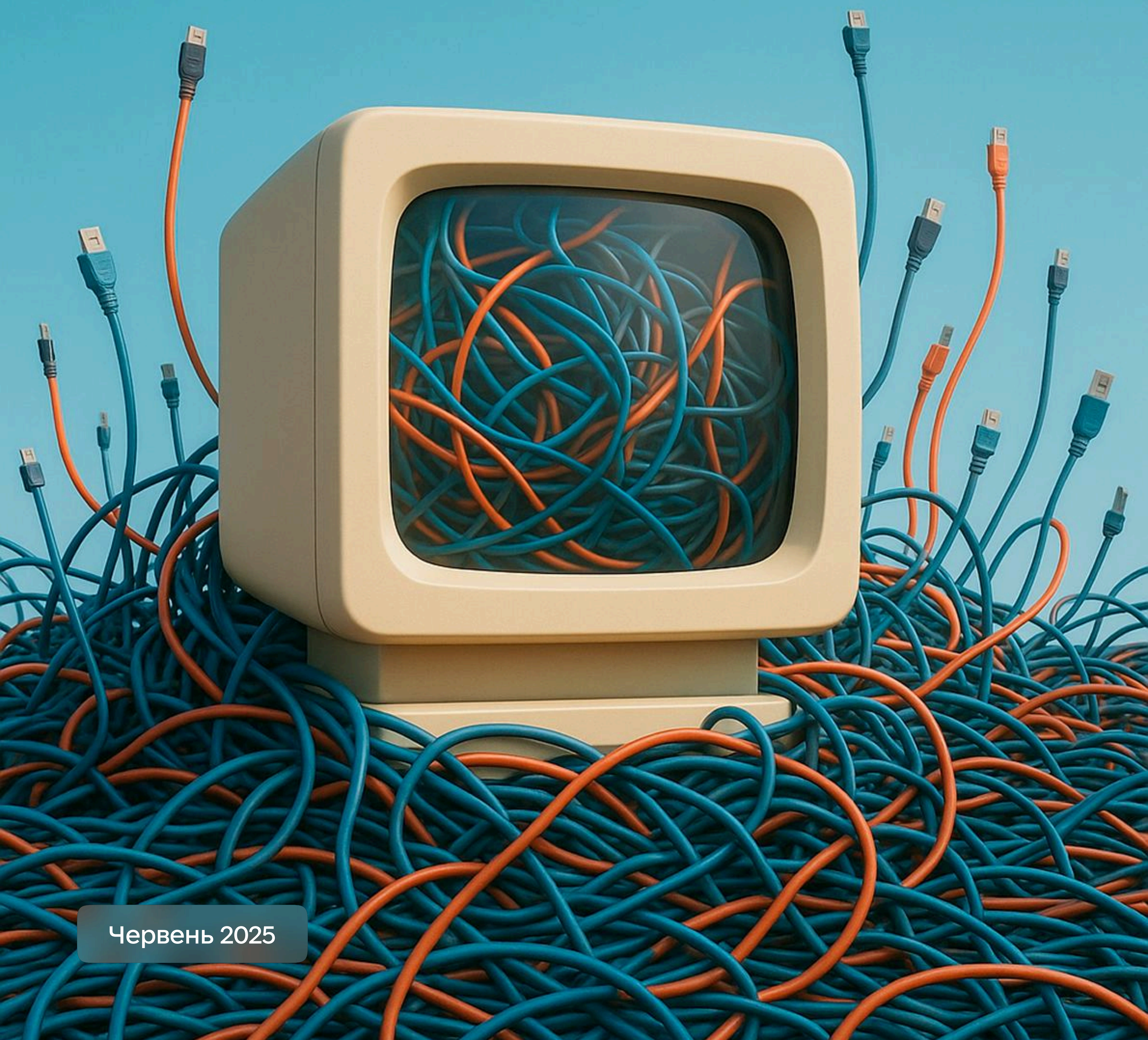


UK International
Development
Partnership | Progress | Prosperity



EURASIA
FOUNDATION

Рекомендації з відповідальної розробки систем із використанням технологій штучного інтелекту



Червень 2025

Зміст

Вступ	4
Термінологія	6
Розділ 1. Дані	10
Розділ 2. Навчання та тонке налаштування моделей	13
Розділ 3. Як обрати велику мовну модель	15
Розділ 4. Вплив штучного інтелекту на права людини	18
4.1. Чому важливо аналізувати вплив штучного інтелекту на права людини	18
4.2. Рекомендації щодо оцінки впливу на права людини	22
4.3. Запобігання дискримінації та упередженням	25
4.4. Питання фахового людського нагляду	27
4.5. Захист прав споживачів	28
4.6. Директива про відповідальність за продукцію (Product Liability Directive)	33
Розділ 5. Авторське право при розробці	34
5.1. Охорона Моделі машинного навчання авторським правом / правом особливого роду	34
5.2. Врегулювання відносин з командою в контексті авторських прав / прав особливого роду	35
5.3. Умови використання об'єктів права інтелектуальної власності та права особливого роду (sui generis) в наборі даних	35
5.4. Використання суспільного надбання	36
5.5. Строк правої охорони майнових авторських прав не є синхронізованим на загальноосвітньому рівні	37
5.6. Copyright Shield й інші приватно-правові ініціативи захисту прав та охоронюваних законом інтересів користувачів систем генеративного ШІ	37

Розділ 6. Приватність

39

6.1. Ключові українські та іноземні регуляторні інструменти, які охоплюють питання обробки ПД з використанням ШІ	39
6.2. Інвентаризація даних та приватність за замовчуванням / проєктуванням	41
6.3. Визначення правових підстав для обробки даних у межах розробки систем ШІ	43
6.4. Оцінка впливу на приватність при розробці систем ШІ	44
6.5. Управління процесами, пов'язаними з реалізацією прав суб'єктами персональних даних	45
6.6. Комунікація щодо обробки ПД шляхом використання ШІ	47
6.7. Технічні та організаційні заходи захисту даних	48
6.8. Технології, які посилюють приватність (PETs), як засіб захисту ПД при обробці ШІ	49
6.9. Хороші Vs погані практики при обробці ПД з використанням ШІ	51

Розділ 7. Антимонопольні ризики

53

Додаток. Інструменти

56

Вступ

Мета, завдання та сфера застосування

Міністерство цифрової трансформації України спільно з Експертно-консультаційним комітетом з розвитку штучного інтелекту при Мінцифри, Асоціацією IT Ukraine, Асоціацією правників України, Центром демократії та верховенства права (ЦЕДЕМ), кафедрою інтелектуальної власності та інформаційного права Навчально-наукового інституту права Київського Національного Університету імені Тараса Шевченка та Науково-дослідним інститутом інтелектуальної власності Національної академії правових наук України, враховуючи світові тенденції, розробили рекомендації із відповідальної розробки систем з використанням технологій штучного інтелекту.

Метою цих рекомендацій є просування етичних принципів та практик відповідальної розробки ШІ, які захищають основні права людини, сприяють соціальній та економічній справедливості, роблять позитивний внесок в інклюзивний суспільний прогрес та зміцнюють довіру суспільства до систем ШІ. Ці рекомендації призначені для розробників системи ШІ та інших зацікавлених сторін і стосуються проведення комплексної оцінки впливу систем ШІ на права людини (HIA). Ці рекомендації ґрунтуються на чинному законодавстві України та актуальних міжнародних практиках і є складовою частиною комплексного підходу до регулювання ШІ в Україні.

Завданнями цих рекомендацій є:

- Скерувати розробників систем ШІ, щодо дотримання етичних принципів розробки та використання ШІ, включно з ризикоорієнтованістю, безпекою та надійністю, захистом приватності та даних, чесністю й справедливістю, прозорістю та пояснювальністю, людським наглядом і моніторингом, лідерством та розвитком поінформованості, захистом прав інтелектуальної власності.
- Заохотити розробників систем ШІ враховувати ширші суспільні впливи їх технологій і прагнути запобігти можливій шкоді.
- Запропонувати структурований підхід для виявлення, оцінки та пом'якшення потенційних ризиків дотримання прав людини, пов'язаних із системами ШІ, включно з упередженням, дискримінацією, порушенням конфіденційності та іншими несприятливими наслідками.
- Допомогти розробникам систем ШІ запровадити стратегії управління ризиками.
- Сприяти безперервному моніторингу та оцінці систем ШІ для виявлення та усунення нових ризиків і впливів у міру їх виникнення.
- Заохотити впровадження ітераційних практик розробки, які містять зворотний зв'язок від користувачів і взаємодію із зацікавленими сторонами.

Обмеження сфери застосування — контекстуально специфічні виклики розробки та використання систем ШІ, наприклад галузеві відмінності, обмеження в доступі до даних, конфіденційність, тощо, які не можуть бути повністю охоплені загальними рекомендаціями. Ці рекомендації не охоплюють питань, що стосуються сектору національної безпеки та оборони.

Актуальність рекомендацій — стрімкий розвиток систем ШІ може призводити до регулярного оновлення цих рекомендацій. Тому переконайтеся, що ви використовуєте актуальну версію рекомендацій. Місяць та рік публікації рекомендацій указаний на їх початку.

Над розробкою цих рекомендацій працювали:

АвдеєваТетяна

Андрієнко Олена

Бистрицька Анастасія

Білик Петро

Дубно Олег

Зеров Костянтин

Кобрін Артем

Краковецький Олександр

Мішина Наталія

Петрів Ольга

Стройко Ірина

Тарасюк Антон

Турута Олексій

Чумаченко Дмитро

Яременко Антон

Ці рекомендації розроблено за сприяння проекту “Цифровізація для зростання, доброчесності та прозорості” (UK DIGIT), що виконується Фондом Євразія та фінансується UK Dev.

Цей матеріал створений за фінансової підтримки Програми допомоги з міжнародного розвитку від Уряду Великої Британії. Зміст є винятковою відповідальністю Міністерства цифрової трансформації України; висловлені погляди не обов’язково відображають офіційну політику Уряду Великої Британії.

¹ Див. [Добровільний Кодекс поведінки з етичного та відповідального використання штучного інтелекту](#)

Термінологія

Відкриті дані

це інформація, яка є вільнодоступною для використання, повторного використання та розповсюдження без обмежень, зазвичай у машиночитаному форматі та з відкритою ліцензією

Дані з обмеженим доступом

Це інформація, доступ до якої обмежений і контролюється певною особою, організацією чи установою. Такі дані не є загальнодоступними й можуть використовуватися лише з дозволу власника або за певних умов.

Верховенство права

принцип, за яким усі особи, установи та організації, включно з державою, підпорядковані законам, які публічно оприлюднені, чіткі та застосовуються на однакових підставах.

Відповідальне використання ШІ

використання систем ШІ з урахуванням їх потенційного впливу на права людини, суспільство та довкілля, з метою запобігання негативним наслідкам.

Гендерна рівність

рівність прав, обов'язків та можливостей усіх людей незалежно від їх статі.

Етичний дизайн ШІ

процес створення систем ШІ, який передбачає етичні міркування на всіх етапах розробки, щоб забезпечити відповідність технологій етичним стандартам і принципам.

Етичний нагляд

процес регулярного аналізу та оцінки систем ШІ з метою забезпечення відповідності етичним стандартам та принципам, зазвичай здійснюється спеціалізованими комітетами або групами експертів.

Етичні принципи

норми та стандарти, що керують поведінкою розробників та користувачів ШІ, зокрема принципи справедливості, рівності, прозорості, підзвітності та відповідальності.

Інклюзивний суспільний прогрес

розвиток суспільства, який забезпечує рівні права, можливості та умови для всіх людей, включно з вразливими групами, щоб вони могли брати участь у житті суспільства та отримувати вигоду від його розвитку.

Ітераційні покращення

процес поступового вдосконалення систем ШІ на основі відгуків користувачів та аналізу впливу, що проводиться регулярно протягом усього життєвого циклу системи.

Механізми зворотного зв'язку

системи або процеси, що дають змогу користувачам або постраждалим особам надавати відгуки або скарги щодо роботи систем ШІ з метою постійного вдосконалення та корекції.

Оцінка впливу на права людини

процес виявлення, оцінки та пом'якшення потенційних ризиків порушення прав людини, пов'язаних із розробкою та використанням систем ШІ.

Приватність

право особи на захист своєї особистої інформації від несанкціонованого доступу, збору, зберігання або використання.

Продукт

система ШІ та/або її невід'ємні елементи, які надаються користувачу незалежно від юридичної моделі такого надання (поставка, користування, ліцензія (зокрема, безоплатна, відкрита тощо), Хаas («Х як послуга») тощо).

Профілювання

форма автоматизованої обробки персональних даних, яка полягає в їх обробці з метою оцінки індивідуальних характеристик, серед яких аналіз та передбачення варіантів поведінки особи (зокрема, у професійній діяльності), її майнового стану, стану здоров'я, особистих уподобань, інтересів, надійності, місцеперебування або пересування;

Соціальна відповідальність

обов'язок організацій і розробників діяти так, щоб їх діяльність сприяла благополуччю суспільства, зокрема через упровадження етичних стандартів та захист прав людини.

Упередженість

наявність систематичних помилок або викривлень в алгоритмах чи наборах даних, що призводять до несправедливих результатів або дискримінації певних груп.

Модель машинного навчання

це конкретний алгоритм або математична модель, яка була навчена на даних і використовується для виконання певних завдань, як-от прогнозування, класифікація чи виявлення аномалій.

Технологія машинного навчання

це сукупність методів, інструментів і підходів, які використовують для створення та застосування моделей машинного навчання. Технологія машинного навчання містить весь процес від збору даних, їх підготовки, вибору й налаштування моделі до її впровадження в реальні системи.

Сімейство великих мовних моделей

як правило, вендори розробляють різні моделі під різні завдання, можуть відрізнятися кількістю параметрів, швидкістю роботи, функціональними можливостями та датою релізу. Наприклад, сімейство моделей GPT-4 містить моделі gpt-4, gpt-4-turbo, gpt-4o, gpt-4o-mini тощо

Інструменти машинного навчання вміщують **бібліотеки** та **фреймворки** (TensorFlow, PyTorch, Scikit-learn), **платформи для збору та обробки даних** (Apache Spark, Pandas), **інструменти для навчання, розгортання й моніторингу моделей** (Kubernetes, TensorFlow Serving, Azure Machine Learning, Azure AI Foundry), а також **автоматизовані платформи для машинного навчання** (Google AutoML, H2O.ai, Azure AutoML).

Інструменти генеративного штучного інтелекту

це програмні засоби, що надають доступ до можливостей генеративного штучного інтелекту через зручні інтерфейси — чатботи, мобільні чи вебзастосунки.

- Сервіси або користувацькі інтерфейси — до таких інструментів належать, наприклад, ChatGPT від OpenAI, Claude від Anthropic, Midjourney, Suno, Microsoft Copilot. Вони надають обмежений доступ (як правило, користувацький) до великих мовних моделей та додаткових інструментів (наприклад, Custom GPT, Canvas, Advanced Data Analysis тощо);
- Playgrounds — інструменти, що допомагають тестувати можливості великих мовних моделей (GitHub Models, Azure AI Foundry Playground, OpenAI Playground, Anthropic Console та інші);
- API — інтерфейс програмного доступу до великих мовних моделей, наприклад, OpenAI API, Azure OpenAI Service API, Anthropic API, Google Vertex AI API тощо.

Підхід у контексті генеративного штучного інтелекту

метод або стратегія, яка використовується для налаштування, адаптації або ефективного застосування великої мовної моделі до конкретних завдань. Ключові підходи:

- інженерія запитів (англ. Prompt Engineering) – створення та оптимізація запитів для отримання бажаних результатів від великої мовної моделі (далі - ВММ)
- ВММ;
- тонке налаштування (англ. Fine-Tuning) – процес адаптації ВММ під конкретні завдання, дані або вимоги без повного донавчання моделі. У процесі тонкого налаштування ВММ зберігає загальні знання, здобуті під час початкового навчання, і налаштовується на специфіку нових даних;
- генерація, доповнена пошуком (англ. Retrieval-Augmented Generation або RAG) – це підхід, що поєднує можливості ВММ з механізмами здобуття знань із зовнішніх джерел, як-от бази даних чи документи.

Часто терміни можуть вживатись у різних контекстах, тому важливо додавати деталі та підказки.

Розглянемо кілька прикладів

ChatGPT



- в узагальненому контексті інструментів генеративного інтелекту (як «ксерокс»)
- як синонім ВММ
- у контексті інструменту chatgpt.com

Midjourney



- у контексті назви компанії
- у контексті назви ВММ для генерування зображень (у такому разі рекомендується вживати конкретну версію ВММ)

Copilot



- у контексті загальної назви цифрового асистента (помічника), що, як правило, вбудовується поверх чинного програмного продукту (наприклад, Zoom AI Assistant, ChatGPT тощо)
- у контексті інструменту від Microsoft (набір інструментів генеративного штучного інтелекту, що складається з вебсервісу <https://copilot.microsoft.com>, мобільних застосунків для iOS та Android, Copilot у Windows, плагіну для веббраузера Microsoft Edge, розширеної версії для персонального користування Microsoft Copilot Pro та корпоративної версії Copilot for Microsoft 365, що працює на основі великих мовних моделей від OpenAI, зокрема, сімейства моделей GPT-4).

З іншими термінами можна ознайомитися, переглянувши
Словник термінів у сфері штучного інтелекту.

Перейти

Розділ 1. Дані

Використовуючи ті чи інші дані для навчання моделей машинного навчання, необхідно переконатися, що:

- у вас є **право їх використовувати** (за однією з відкритих ліцензій, за дозволом власника чи розпорядника даних тощо);
- дані є **збалансованими**, що означає, що різні класи або категорії в наборі даних представлені рівномірно, без значного переважання одних над іншими. Це забезпечує відсутність суттєвих перекосів та упередженості у вибірці;
- дані є **нормалізованими**, що означає, що дані були перетворені так, щоб вони відповідали певному стандартному діапазону або формі розподілу, зазвичай без зміни відносних співвідношень між значеннями, що зменшить або вніможливить помилки та галюцинування в майбутньому;
- дані є **повними**, тобто такими, що описують фактичний стан речей і можуть бути використані для отримання достовірних висновків про реальний світ;
- дані є **знеособленими**, що означає, що видалено або змінено всі елементи, що дають змогу прямо чи опосередковано ідентифікувати конкретну особу, при цьому зберігається цінність даних для аналізу чи досліджень, одночасно захищаючи приватність та конфіденційність осіб, яких ці дані стосуються;
- для даних створені **метадані**, що означає, що дані супроводжуються структурованою інформацією, яка описує різні аспекти самих даних, як-от їх походження, формат, структура, зміст, якість, час створення, авторство та умови використання. Метадані полегшують розуміння, управління та використання основних даних.

Порушення одного чи кількох пунктів часто є причиною помилок роботи моделей (наприклад, у моделях класифікації чи регресійного аналізу) чи отримання ефекту галюцинування (у моделях генеративного штучного інтелекту).

Ось кілька інструментів та методів виявлення проблем з даними:

Візуалізація даних

Графічні інструменти дають змогу швидко виявити аномалії, пропуски та нерівномірність у даних. Для цього використовуються Python-бібліотеки Matplotlib, Seaborn, Plotly, а також бізнес-аналітичні платформи Tableau та Microsoft Power BI.

Статистичні метрики

Кількісні показники допомагають оцінити властивості даних. Наприклад, для аналізу розподілу використовують середнє, медіану, стандартне відхилення, а для дисбалансу класів — співвідношення класів (class ratio), ентропію (entropy), індекс інформаційного приросту (information gain ratio) тощо.

Спеціалізовані бібліотеки та алгоритми

Для виявлення проблем у даних є спеціалізовані інструменти. Наприклад, бібліотека `imbalanced-learn` містить алгоритми для балансування класів у завданнях класифікації, а метод SMOTE (Synthetic Minority Over-sampling Technique) дає змогу створювати синтетичні зразки для покращення балансу.

Автоматизовані інструменти профілювання даних

Інструменти автоматичного аналізу допомагають швидко отримати звіти про якість даних. Наприклад, `ydata-profiling` (раніше Pandas Profiling) генерує детальні звіти про розподіл, кореляції та пропущені значення, а Great Expectations дає змогу створювати правила перевірки якості даних для ETL-процесів

Провідні хмарні платформи, як-от Microsoft Azure, Amazon Web Services (AWS) та Google Cloud, пропонують спеціалізовані інструменти для виявлення проблем із даними перед їх використанням у машинному навчанні

Microsoft Azure

- Azure Data Profiler (у складі Azure Machine Learning Data Wrangler) аналізує якість даних, виявляє викиди, пропущені значення, дублікати та аномалії.
- Azure Monitor for Data дає змогу відстежувати якість даних у реальному часі, включно зі змінами в розподілі та трендах.

Amazon Web Services (AWS):

- Amazon SageMaker Data Wrangler — будований аналізатор, що автоматично визначає проблеми в наборах даних, як-от нестача значень та розподіл змінних.
- Amazon SageMaker Clarify виявляє упередженість у даних (data bias), а також оцінює рівномірність розподілу категорій.
- AWS Glue DataBrew автоматично профілює дані, виявляє аномалії, дублікати та відхилення від очікуваного формату.

Google Cloud

- Dataplex Data Quality оцінює якість даних, шукає пропуски та аномалії, а також забезпечує автоматизований моніторинг.
- Vertex AI Feature Store Monitoring дає змогу відстежувати зміни в розподілі даних і виявляти проблеми зі стабільністю даних у часі (data drift).

Ці інструменти дають змогу автоматично знаходити помилки в даних, як-от аномалії, пропущені значення, незбалансованість класів, нестабільність розподілу та упередженість, що можуть суттєво вплинути на якість моделей машинного навчання. Виявлення та мінімізація цих проблем є частиною наряду відповідального ШІ (Responsible AI), який спрямований на забезпечення прозорості, справедливості та надійності моделей штучного інтелекту. Детальніша інформація щодо цього наведена **в розділі 4.3. Запобігання дискримінації та упередженням.**

Як якість даних впливає на моделі машинного навчання

Якість навчальних даних безпосередньо впливає на **достовірність, об'єктивність і точність** моделей машинного навчання. Використання перевірених, збалансованих і актуальних даних допомагає мінімізувати помилки, упередження та покращити загальну ефективність моделі.

Помилкові або ненадійні дані можуть призвести до того, що модель вважатиме їх достовірними. Навіть якщо вони становлять невелику частину навчального набору, це може впливати на точність прогнозів та рішень моделі. Це особливо критично для великих мовних моделей, які оперують текстовою інформацією та можуть генерувати правдоподібні, але вигадані відповіді або спиратися на дезінформацію, пропаганду чи теорії змов.

Обмеженість та застарілість даних створюють прогалини в знаннях моделі, оскільки вона навчається на інформації, актуальній лише на момент збору. Для великих мовних моделей це може означати відсутність актуальних фактів про останні події, що підвищує ризик конфабуляцій та галюцинацій. У класичних ML-моделях застарілі дані можуть знижувати якість прогнозів та адаптивність до змінних умов.

Дисбаланс та упередженість у даних можуть спричиняти викривлене представлення інформації та необ'єктивні результати. Якщо певні категорії або групи даних надмірно представлені, модель може розвивати упередження. У випадку великих мовних моделей це може призводити до нерівномірного відображення різних поглядів, а в класичних ML-моделях — до поганої узагальненості, коли модель неефективно працює на менш представлених класах.

Якість даних є визначальним фактором для всіх типів моделей машинного навчання. Для **великих мовних моделей** критичними є актуальність, збалансованість і точність текстових даних, щоб зменшити ризик галюцинацій та упередженості. Для **традиційних ML-моделей** важливо забезпечити коректну розмітку, усунути дисбаланс класів і підтримувати оновлюваність даних для стабільної роботи в реальних умовах.

Розділ 2. Навчання та тонке налаштування моделей

Важливо розрізняти три фундаментальні підходи щодо навчання моделей машинного навчання: навчання з учителем, або кероване навчання; навчання без учителя, або некероване навчання; навчання з підкріпленням.

Під час навчання з учителем модель тренується на розмічених даних, де кожному вхідному зразку відповідає конкретний правильний вихід. Мета полягає в тому, щоб навчити модель точно передбачати результати для нових, раніше невідомих даних. Цей метод широко застосовується в таких завданнях, як класифікація зображень, розпізнавання мови чи машинний переклад. Натомість навчання без учителя працює з нерозміченими даними, намагаючись виявити приховані структури чи закономірності. Тут модель самостійно шукає внутрішні взаємозв'язки в даних без попередньо визначених міток. Цей підхід часто використовується для кластеризації, зменшення розмірності даних або виявлення аномалій. Навчання з підкріпленням є окремим підходом, де агент взаємодіє із середовищем, отримуючи винагороди або покарання за свої дії. Мета агента — максимізувати сумарну винагороду через серію проб і помилок. Цей метод особливо ефективний у завданнях, де потрібно приймати послідовність рішень, як-от в іграх, робототехніці чи оптимізації систем управління. На відміну від навчання з учителем, тут немає заздалегідь визначених правильних відповідей, а модель вчиться на основі досвіду взаємодії із середовищем.

На практиці підходи щодо навчання моделей можуть комбінуватися. Наприклад, великі мовні моделі сімейства GPT, як-от GPT-3, спочатку навчалися без вчителя. Цей підхід дав змогу моделям вивчити загальні закономірності мови та набути загальних знань, але без чіткого напрямку застосування. Натомість для моделей ChatGPT 3.5 та 4 був застосований більш складний підхід, що передбачав навчання з підкріпленням на основі зворотного зв'язку від людини (RLHF). Після початкового етапу навчання без учителя ці моделі пройшли процес тонкого налаштування (fine-tuning) з використанням RLHF. У цьому підході модель взаємодіє з людьми, які надають зворотний зв'язок щодо якості відповідей моделі. На основі цього зворотного зв'язку модель навчається генерувати відповіді, які краще відповідають людським очікуванням та вимогам. Такий підхід дав змогу зробити ChatGPT більш адаптованим до діалогу, здатним краще розуміти контекст та генерувати більш релевантні та безпечні відповіді. Однак використання RLHF також породжує певні етичні питання та потенційні проблеми, пов'язані з можливим успадкуванням людських упереджень та тенденцією моделі до надання відповідей, які, на її «думку», сподобаються користувачам, навіть якщо вони не завжди є найбільш точними чи об'єктивними.

Навчання моделей генеративного штучного інтелекту на загальних даних дає можливість побудувати зв'язки між словами та сутностями на їх основі, однак це не гарантує, що модель розв'язуватиме математичні задачі та відповідати на питання, де потрібні логічні міркування. Сучасні моделі вже можуть відповідати на такі типи запитань, однак достовірність цих відповідей все ще перебуває під питанням. Для подолання цих проблем потрібно використовувати більш складні підходи та методи або спеціалізовані моделі, які створені для конкретного типу завдань.

Важливо пам'ятати, що моделі генеративного штучного інтелекту є попередньо навченими (англ. pre-trained), що означає, що в процесі використання моделей вони не навчаються. Інструментом адаптації моделей до специфічних завдань є тонке налаштування, про яке ми вже згадували вище. Однак перш ніж занурюватися в цей процес, варто спробувати оптимізувати результати за допомогою інших методів, наприклад, використати просунуті техніки інженерії запитів, як-от ланцюжок міркувань, запити з одним або кількома прикладами ті інші підходи. Якщо вони не дають бажаних результатів, тоді тонке налаштування може стати вашим наступним кроком.

Тонке налаштування особливо корисне в кількох ключових сценаріях:

- налаштування стилю, тону та формату виводу;
- підвищення надійності отримання бажаних результатів при використанні специфічної термінології, наприклад, у юриспруденції чи медицині;
- за умови використання великих запитів з метою оптимізації кількості вхідних токенів.

Для тонкого налаштування необхідний набір даних у вигляді пар «запит» — «відповідь». Для кращого результату використовуйте формат розмови з ролями (system, user, assistant), забезпечте збалансованість даних. Тонке налаштування краще працюватиме на великій кількості даних, однак якість даних усе ж важливіша за кількість.

Пам'ятайте, що вартість токенів для тонко налаштованих моделей часто є вищою за вартість базових моделей.

Як обрати ML-модель

Вибір класичної моделі машинного навчання залежить від низки ключових факторів, які впливають на її продуктивність, надійність та можливість застосування в конкретному сценарії.

Перший важливий фактор — **репутація постачальника** (розробника). Довіра до постачальника моделі є критичною, особливо у сферах, де помилки можуть мати значні наслідки, наприклад, у фінансах, медицині чи безпеці. Моделі від великих технологічних компаній, як-от Microsoft, Google, OpenAI, Meta та NVIDIA, зазвичай мають кращу підтримку, документацію та перевіреність у реальних сценаріях. Моделі з відкритим кодом, як-от Scikit-Learn, XGBoost або LightGBM, можуть бути ефективними, якщо є експертиза для їх налаштування та валідації.

Другий фактор — **джерело навчальних даних**. Якість і походження навчального набору безпосередньо впливають на точність моделі. Бажано використовувати моделі, що навчалися на відкритих і репрезентативних наборах даних, або мати можливість адаптувати їх під свої потреби. Варто бути обережним із моделями, джерела тренувальних даних яких не розкриті або можуть містити значні упередження.

Третій фактор — **швидкодія та продуктивність**. Важливо враховувати, скільки ресурсів споживає модель та яка її швидкість при виконанні запитів. Глибокі нейронні мережі можуть бути потужнішими, але водночас вимагати значно більше обчислювальних потужностей порівняно з класичними алгоритмами, як-от логістична регресія, випадкові ліси чи градієнтний бустинг. Вибір моделі має враховувати баланс між точністю та швидкістю роботи, особливо якщо модель має працювати в реальному часі.

Четвертий фактор — **вартість використання**. Деякі ML-моделі є безоплатними, але їх упровадження може потребувати значних обчислювальних потужностей. Вендорські рішення, як-от Azure Machine Learning, Google Vertex AI та AWS SageMaker, пропонують зручні інструменти для роботи з моделями, але можуть бути дорогими у високонавантажених сценаріях. Важливо оцінити сукупну вартість володіння, включно з витратами на хмарні обчислення, ліцензіями, підтримкою та супутніми інфраструктурними витратами.

П'ятий фактор — **результати бенчмарків**. Порівняння моделей за допомогою стандартизованих тестів допомагає вибрати найефективніше рішення. Основні метрики для оцінки передбачають точність класифікації (Accuracy / F1-score), баланс між правильним виявленням та уникненням хибнопозитивних результатів (Precision / Recall), час виконання прогнозу (Inference time) та обсяг необхідної пам'яті (Memory footprint). Використання відкритих платформ, як-от Papers with Code, Hugging Face Leaderboard, Chatbot Arena або MLPerf, дає змогу оцінити конкурентні рішення за реальними показниками продуктивності.

Шостий фактор — **можливість адаптації та інтеграції**. Деякі моделі можна легко адаптувати до нових даних через тонке налаштування або трансферне навчання. Важливо враховувати, наскільки модель легко інтегрується у ваш технологічний стек, наприклад, через підтримку ONNX, TensorFlow Serving, PyTorch або MLflow.

Розділ 3. Як обрати велику мовну модель

Найперше варто звернути увагу на джерела даних для навчання. Більшість моделей навчається на інформації, отриманій шляхом парсингу відкритих джерел в інтернеті (вебскрепінг). Ці дані можуть містити не лише достовірну інформацію, а й помилки, чутки та недостовірні факти, що в подальшому впливає на якість генерованого контенту. На жаль, більшість публічних великих мовних моделей не розкриває джерел даних або використовує набори даних сумнівної якості, тому цей фактор важко враховувати на практиці.

Крім того, навчальні набори даних зазвичай обмежені певною датою збору. Це призводить до неповноти інформації про останні події, що особливо критично для завдань, які вимагають актуальних знань.

Ще одним фактором є обмеження контекстного вікна — це максимальна кількість токенів, які модель може обробити одночасно під час аналізу або генерації відповіді. Це важливий параметр, який визначає, скільки попереднього контексту модель може «пам'ятати» й використовувати для формування відповідей або продовження тексту, що може призвести до втрати важливої інформації при обробці довгих текстів, а отже, і до неправильних висновків чи тверджень.

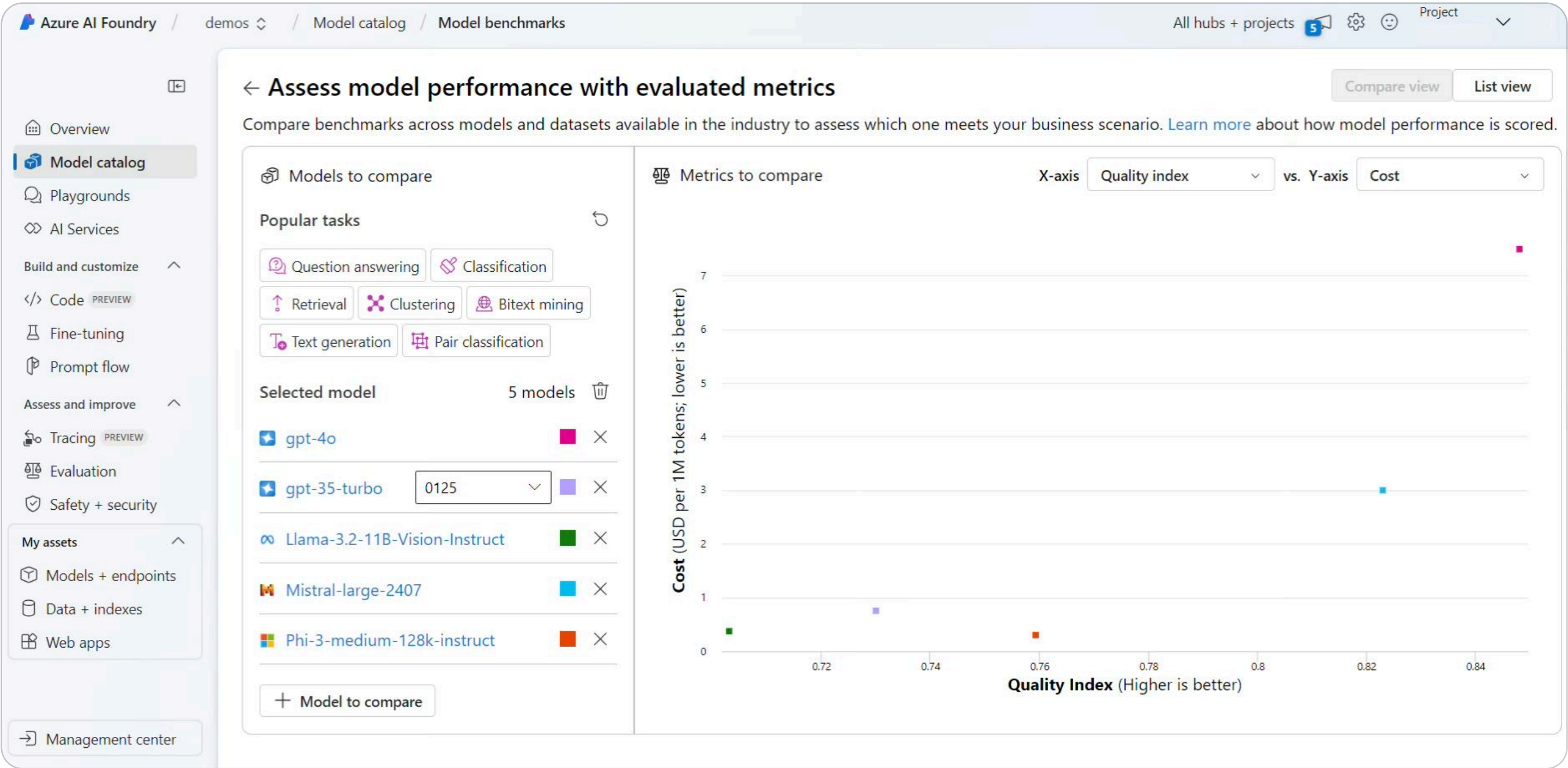
Згадуючи про фактори, що впливають на якість моделей генеративного штучного інтелекту, не можна оминати питання кількості параметрів моделі. Хоча більша кількість параметрів дає змогу моделі вивчати складніші патерни та зберігати більше інформації, це не єдиний визначальний фактор. Якість та різноманітність навчальних даних, архітектура моделі, методи навчання й регуляризації відіграють не менш важливу роль. Збільшення кількості параметрів не гарантує пропорційного покращення якості моделі, але на певному етапі може слугувати швидким ідентифікатором «інтелектуальних» здібностей моделі.

Зараз є три основні типи моделей генеративного штучного інтелекту: **стандартні, з міркуванням та гібридні. Стандартні моделі** (наприклад, GPT-4o) оптимізовані для широкого спектра завдань, включно з генерацією тексту, перекладом й узагальненням. **Моделі з міркуванням** (як-от OpenAI o1, OpenAI o3) мають покращені можливості аналізу та дедуктивного мислення, що робить їх придатними для складних логічних задач. **Гібридні моделі** (наприклад, Claude Sonnet 3.7) поєднують характеристики стандартних і міркувальних моделей, забезпечуючи баланс між швидкістю, контекстним розумінням та складними когнітивними завданнями. Вибір моделі залежить від її призначення: для швидких відповідей у чатах підійдуть стандартні моделі, для глибокого аналізу — моделі з міркуванням, а для універсального використання — гібридні.

Наступний етап — порівняння різних моделей. Для цього можна скористатися бенчмарками — стандартизованими методами оцінки та порівняння продуктивності різних моделей штучного інтелекту, зокрема великих мовних моделей.

Приклади метрик для оцінювання великих мовних моделей

- 1. Точність (Accuracy) вимірює, наскільки часто відповіді моделі точно збігаються з еталонними відповідями.
- 2. Когерентність (Coherence) оцінює, наскільки добре мовна модель створює вихідні дані, які плавно перетікають та читаються природно.
- 3. Плавність (Fluency) оцінює мовну майстерність генеративної моделі ШІ, зокрема граматичну правильність та природність звучання.
- 4. Схожість із GPT (GPT similarity) кількісно визначає подібність між еталонним текстом і текстом, згенерованим моделлю ШІ.
- 5. Обґрунтованість (Groundedness) вимірює, наскільки добре згенеровані відповіді узгоджуються з інформацією з вхідного джерела.
- 6. Релевантність (Relevance) оцінює, наскільки згенеровані відповіді є доречними та безпосередньо пов'язаними з поставленими запитаннями.



Порівняння великих мовних моделей у **Azure AI** Foundry

Альтернативним способом визначити найкращу модель є використання «колективного розуму» через сервіси на зразок Chatbot Arena (lmarena.ai), які дають змогу користувачам тестувати та порівнювати різні великі мовні моделі у реальному часі. Користувачі можуть оцінювати відповіді моделей і голосувати за найкращу, що допомагає формувати рейтинг моделей на основі реальних сценаріїв використання та відгуків спільноти.

Для ручного оцінювання моделей рекомендується використовувати так звані Playgrounds – вебсервіси, що надають доступ до моделей з можливістю експериментувати з різними параметрами та налаштуваннями. Ці платформи дають змогу користувачам взаємодіяти з різними моделями штучного інтелекту в реальному часі, змінювати вхідні дані, налаштування та спостерігати за результатами. Вендори, як правило, надають Playground як окремий вебсервіс, наприклад, OpenAI Playground (<https://platform.openai.com/playground/>) Anthropic Console (<https://console.anthropic.com/>), а також можна скористатися платформами, де зібрана велика кількість моделей від різних вендорів, як-от

Azure AI Foundry

Перейти ↗

Google Vertex AI Studio

Перейти ↗

GitHub Models

Перейти ↗

HuggingFace

Перейти ↗

Використовуючи Playgrounds або доступ до моделей через API, є можливість додатково налаштовувати низку параметрів, найголовнішим з яких є температура (temperature). Цей параметр контролює «креативність» або варіативність відповідей моделі. За низької температури (близькій до 0) модель генерує більш передбачувані та консервативні відповіді, обираючи найбільш імовірні слова. За високої температури (ближче до 1 або вище) модель стає більш «креативною», генеруючи різноманітніші та менш передбачувані відповіді.

Окрім температури, інші важливі параметри, які можна налаштовувати, містять:

1. Top-p (або nucleus sampling) визначає кумулятивну ймовірність, за якою модель обирає наступне слово.
2. Max tokens установлює максимальну кількість токенів (слів або частин слів), які модель може згенерувати у відповіді.
3. Presence penalty та frequency penalty впливають на ймовірність повторення слів або фраз у генерованому тексті, допомагаючи уникнути надмірних повторень.
4. Stop sequences дають змогу вказати послідовність символів, досягнувши яких, модель припинить генерацію.

Налаштування цих параметрів дає можливість точно контролювати поведінку моделі та отримувати результати, які найкраще відповідають конкретним потребам завдання чи застосування. Експериментування з різними комбінаціями цих параметрів може значно вплинути на якість і характер генерованого контенту.

Найпростішим способом розпочати використання моделей генеративного ШІ є використання спеціалізованих інструментів на кшталт GitHub Copilot, Google Duet AI, Amazon CodeWhisperer, Cursor. Їх відмінність від загальних інструментів на кшталт ChatGPT в тому, що вони адаптовані для розробників, інтегровані з відомими IDE та можуть додатково враховувати контекст усього програмного проєкту, а не лише конкретного файлу чи завдання.

Вендори можуть використовувати дані із запитів для навчання нових версій моделей та інструментів, тому з метою захисту інформації потрібно віддавати перевагу корпоративним версіям інструментів, наприклад, ChatGPT Enterprise, Azure OpenAI Service.

Використовуючи моделі, важливо убезпечитися не лише від ін'єкцій (prompt injection), джейлбрейку (jailbreaking) та інших зловмисних технік. Для цього можуть використовуватися додаткові моделі та інструменти, як-от Moderation модель у OpenAI API, Azure Content Safety тощо.

Використовуючи технології генеративного штучного інтелекту, питання авторських прав та прав власності на згенеровані результати є ключовими. Різні платформи мають різні підходи до цього питання, і важливо знати, які умови застосовують у кожному конкретному випадку, тому:

1. Уважно читайте умови використання кожного інструменту. Кожен сервіс може мати свої специфічні умови, які відрізняються залежно від платного або безкоштовного плану.
2. Перевіряйте права на контент. Якщо згенерований контент може містити матеріали третіх сторін (наприклад, текст або зображення), які можуть бути захищені авторським правом, варто бути обережним із його використанням.
3. Враховуйте етичний аспект. Генеративний ШІ може створювати контент, що схожий на реальні твори, тому слід з повагою ставитися до інтелектуальної власності інших.

Це допоможе уникнути непорозумінь та потенційних правових проблем при використанні генеративного штучного інтелекту.

Розділ 4. Вплив штучного інтелекту на права людини

4.1. Чому важливо аналізувати вплив штучного інтелекту на права людини

Розробники штучного інтелекту, безперечно, несуть значну відповідальність за те, щоб їхні рішення на основі ШІ були етичними, не шкодили людям та активно сприяли справедливості, рівності та прогресу суспільства. Ця відповідальність виходить за межі технічних характеристик продуктів і систем та охоплює соціальні й етичні наслідки використання ШІ. Упроваджуючи принципи справедливості та рівності в процеси дизайну та розробки, розробники ШІ можуть запобігти упередженням, дискримінації та іншим шкідливим наслідкам, які непропорційно впливають на вразливі групи населення. Завдяки сумлінним зусиллям у проведенні оцінки впливу на права людини протягом усього життєвого циклу продуктів та систем, розробники ШІ можуть сприяти розширенню людських можливостей, дотриманню прав людини, гендерній рівності та інклюзивному суспільному прогресу.

Розробники ШІ мають усвідомлювати потенційні ризики порушень прав людини, які можуть виникнути на кожному з етапів життєвого циклу систем ШІ. Розуміння цих ризиків є важливим для створення етичних та відповідальних технологій на основі ШІ. Нижче наведено приклади фундаментальних прав людини, які можуть зазнавати впливу від рішень на основі ШІ.

Право на життя, фізичну, психологічну та моральну недоторканність

Оскільки використання ШІ у всіх сферах життя буде тільки посилюватися, важливо розуміти, що системи ШІ можуть прямо або опосередковано порушувати фундаментальне право на життя, фізичну, психологічну та моральну цілісність. Відсутність належного контролю й заходів з оцінки впливу може призвести до серйозних наслідків для людей.

Приклад порушень:

- Системи ШІ, що використовуються в охороні здоров'я для діагностики та рекомендацій щодо лікування, можуть спричинити шкоду, якщо ухвалюють неправильні рішення. Наприклад, алгоритм ШІ, що некоректно діагностує серйозне захворювання або не виявляє критичний стан, може призвести до неадекватного або неправильного лікування, наражаючи пацієнтів на небезпеку.

Право на приватність

Системи ШІ вимагають великої кількості даних (наборів даних) для ефективної роботи. Ці дані можуть містити особисту та конфіденційну інформацію про людей, використання якої може призвести до порушення права на приватність. Застосування ШІ у сфері спостереження, соціальних медіа та аналізу даних може порушувати приватність людей шляхом збирання, зберігання та аналізу їхніх особистих даних без належної згоди або заходів безпеки.

Крім того, розробники ШІ часто співпрацюють з третіми сторонами для збору або завантаження даних через API або інші інтерфейси, що може призвести до витоку даних або їх неналежного використання. У багатьох випадках, особливо в таких чутливих сферах, як фінансовий сектор, залучення сторонніх розробників або аутсорсинг певних рішень може створювати серйозні ризики для безпеки даних.

Приклад порушень:

- Несанкціонований доступ до особистих даних у соціальних медіа на основі ШІ.
- Системи спостереження, які інвазивно обробляють та зберігають біометричні дані людей без їхнього відома чи згоди.
- Витік даних під час передачі між банківськими системами та сторонніми розробниками.

2

Право на недискримінацію, справедливість та рівність

Системи ШІ можуть зберігати та навіть посилювати наявні упередження, що містяться в наборах даних, на яких вони навчаються. Це може призвести до дискримінаційних результатів роботи ШІ на основі статі, віку, наявності інвалідності або інших соціальних характеристик. Такі упередження можуть впливати на рішення у сфері найму, кредитування, навчання та інших важливих галузях, що призводить до несправедливого ставлення до певних груп.

Приклад порушень:

- Алгоритми найму, які віддають перевагу певним демографічним групам на основі упереджених навчальних даних.

3

Право на свободу вираження поглядів

Автоматизовані системи, що видаляють або депріоритезують контент на основі попередньо визначених критеріїв, можуть випадково цензурувати правомірний матеріал. Висока вірогідність таких порушень можлива, коли алгоритми є непрозорими або упередженими, що призводить до надмірної цензури або придушення опозиційних думок. У таких випадках ШІ може стати інструментом ствердження авторитарних режимів та придушення свободи слова.

Приклад порушень:

- Платформи соціальних медіа, які використовують ШІ для видалення контенту, що може бути позначений як неприйнятний, але фактично є формою законного вираження.
- Автоматизовані системи, що блокують або фільтрують політичний контент або контент активістів.

4

Право на справедливі та належні умови праці

Системи ШІ, що використовуються на робочому місці для моніторингу, оцінки ефективності або розподілу завдань, можуть вплинути на реалізацію прав працівників. Надмірне спостереження та несправедлива оцінка ефективності можуть створити ворожу робочу атмосферу, вплинути на безпеку робочого місця й гідність працівників. Рішення на основі ШІ, які ухвалюють без дотримання принципів прозорості або можливості оскарження, можуть призвести до несправедливих умов праці.

Приклад порушень:

- Інструменти ШІ, які безперервно відстежують діяльність працівників, наприклад, чи усміхаються працівники клієнтам закладу, можуть призводити до порушення приватності та підвищення стресу.
- Системи оцінки ефективності, які несправедливо карають працівників на основі упереджених даних або непрозорих критеріїв.

5

Право на доступ до засобів правового захисту

Коли системи ШІ ухвалюють рішення, що негативно впливають на життя людей, важливо мати механізми для того, щоб ці люди могли оскаржувати такі рішення та шукати засоби правового захисту. Відсутність прозорості й підзвітності в системах ШІ може ускладнити доступ до ефективних засобів правового захисту, залишаючи людей без можливості виправити ситуацію.

Приклад порушень:

- Автоматизовані системи кредитного скорингу, що відмовляють у наданні позик без чітких причин або можливості оскарження.
- Алгоритмічні рішення в системі кримінального правосуддя, що впливають на звільнення під заставу або винесення вироків без прозорості чи можливості оскарження.

6

Право на участь у культурному житті, на науку та інтелектуальну власність

Системи ШІ можуть впливати на доступ до культурних та наукових ресурсів, а також на захист прав інтелектуальної власності. Розробники повинні забезпечити, щоб системи ШІ не обмежували несправедливо доступ до культурного контенту або не експлуатували інтелектуальну власність без належного визнання та компенсації.

Приклад порушень:

- Алгоритми ШІ, що визначають доступність культурного або освітнього контенту в інтернеті, що може обмежити доступ для певних груп.
- Роботи, створені на основі ШІ, що порушують питання прав інтелектуальної власності та справедливої компенсації людським творцям.

7

Економічні та соціальні права

Системи ШІ можуть впливати на здійснення права на працю, права на достатній рівень життя та права на соціальне забезпечення тощо.

Приклад порушень:

- Алгоритми штучного інтелекту, які використовують у галузі нерухомості для схвалення іпотеки, можуть посилити наявні упередження, що призведе до дискримінаційної практики щодо певних, часто вразливих, груп.
- Системи ШІ, які використовують для визначення права на отримання соціальних послуг, як-от соціальна допомога, допомога з безробіття або аліменти, можуть бути помилковими або необ'єктивними.

8

Право на свободу думки, совісті та релігії

Системи ШІ іноді здатні відстежувати переконання та впливати на формування чи зміну поглядів. Це може порушувати автономію людини в питаннях світогляду або релігійних переконань, особливо якщо відбувається приховане маніпулювання.

Приклад порушень:

- Рекомендаційні алгоритми, що здатні підштовхувати користувачів до певної політичної чи релігійної позиції без прозорості та збалансованості альтернатив.

9

Право на свободу мирних зібрань і об'єднань

Автоматизовані механізми модерації або обмеження доступу до онлайн-інструментів можуть фактично перешкоджати реалізації права на свободу зібрань.

Приклад порушень:

- Системи ШІ, що помилково позначають заклики до мирного протесту як «екстремістські» або «мова ворожнечі», можуть призвести до блокування груп громадян, які планують мирне зібрання.

10

Право на власність

Системи ШІ можуть впливати на користування майном або визначати, кому надавати кредити / іпотеки.

Приклад порушень:

- Автоматизована система для аналізу іпотечної спроможності може несправедливо занижувати показники певним категоріям позичальників (через упереджені або неповні дані), обмежуючи їх доступ до житла й фактично порушуючи право на володіння власністю.

11

Право на освіту

Якщо системи ШІ в навчанні чи оцінюванні упереджені, може виникати нерівність в освіті та порушення принципу рівного доступу до знань.

Приклад порушень:

- Платформа для дистанційної освіти, що використовує ШІ для формування індивідуальних завдань, виявляється оптимізованою для певної культурної групи чи мови й не враховує потреб учнів з інвалідністю або учасників із віддалених регіонів. Через це можуть занижуватися оцінки та формуватися неправильні рейтинги.

12

Право на свободу пересування

Деякі системи ШІ можуть створювати додаткові бар'єри або накладати обмеження на пересування осіб. Надмірне та непрозоре використання цих технологій здатне ускладнювати або навіть перешкоджати свободі пересування.

Приклад порушень:

- Автоматизована система контролю на кордоні, що хибно визначає когось як «ризикового пасажир», призводить до незаслуженої депортації чи затримання.

13

Перелік прав людини, які можуть зазнати впливу внаслідок застосування штучного інтелекту, є значно ширшим, ніж зазначено в цих рекомендаціях. Він охоплює фундаментальні права від права на людську гідність і життя до свободи самовираження, релігії, пересування, об'єднань, праці, освіти, приватності та багатьох інших, що демонструє, наскільки різноманітними та взаємопов'язаними є ці права й наскільки важливо, щоб розробники систем ШІ враховували весь цей комплекс під час проєктування, впровадження та оцінки ризиків технологій.

Адже усвідомлюючи потенційні порушення прав людини, розробники ШІ можуть вжити проактивних заходів для зменшення ризиків і забезпечення того, щоб їхні технології сприяли справедливості, рівності та повазі до всіх людей, незалежно від статі, віку, соціального статусу тощо. Ця обізнаність є критично важливою для побудови довіри та підзвітності в системах ШІ, що в підсумку сприяє їх етичному та відповідальному розвитку й упровадженню.

4.2. Рекомендації щодо оцінки впливу на права людини

Ці рекомендації створені, щоб допомогти розробникам дотримуватися стандартів етичного та відповідального ШІ, захищати права людини та забезпечувати соціальну відповідальність під час розробки систем ШІ. Ці рекомендації послуговуються найкращими міжнародними практиками здійснення оцінки впливу на права людини, зокрема **методологією HUDERIA**. Водночас форми і способи їх упровадження залишаються на розсуд організацій та розробників.

² Корисні посилання для методології детального опрацювання впливів:

- [HUDERIA](#)
- [Microsoft Responsible AI Impact Assessment Guide](#)
- [Quantiphi, Staying Ahead of the Curve: Why Responsible AI Impact Assessment is a Must for Modern Businesses Security](#)

³ Цей приклад є ознайомлювальним, не охоплює повною мірою всієї методології, а лише ілюструє її основні етапи для загального розуміння.

⁴ Кілька прикладів нижче:

- [Регламент \(ЄС\) 2024/1689 від 13 червня 2024 року про встановлення гармонізованих правил щодо штучного інтелекту та внесення змін до Регламентів \(ЄС\) № 300/2008, \(ЄС\) № 167/2013, \(ЄС\) № 168/2013, \(ЄС\) 2018/858, \(ЄС\) 2018/1139 і \(ЄС\) 2019/2144 та Директив 2014/90/ЄС, \(ЄС\) 2016/797 і \(ЄС\) 2020/1828 \(Закон про штучний інтелект\)](#)
- [NIST Risk Management Framework \(RMF\)](#)
- ISO/IEC 31000:2018 — Настанови з управління ризиками
- ISO/IEC 27005:2018 — Управління ризиками інформаційної безпеки
- [База даних інцидентів ШІ](#)
- [Оцінка впливу ШІ на етику](#)

Контекстуальна оцінка ризиків (COBRA)

- Попередньо визначити сфери, де продукт може впливати на права людини, демократичні процеси чи верховенство права.
- Попередньо визначити фактори ризику та етапи життєвого циклу системи ШІ (проєктування, розробка, розгортання тощо), на яких вони виникають.
- Визначити рівні ризику на підставі аналізу їх масштабу, обсягу, імовірності та можливості відновити попередній стан.
- Перевірити якість даних, які використовують для навчання системи ШІ (чи є ризик упередженості, недостатньої репрезентативності тощо).

³Приклад:

- Компанія планує використовувати штучний інтелект, щоб скоротити витрати на команду модераторів і швидко видаляти «небажаний» контент (образливі пости, мову ворожнечі, наклеп, надмірно відверті матеріали тощо). Однак у вихідних даних та історичних прикладах модерації можуть існувати упередження: певні лінгвістичні або культурні групи можуть бути надмірно заблоковані, а мова ворожнечі іноді залишатися непоміченою.
- Основний ризик полягає в потенційному порушенні права на свободу вираження поглядів та нерівному ставленні до різних культурних або соціальних груп. Ураховуючи значний обсяг даних і високу ймовірність помилок, система вимагає детального аналізу й запровадження дій, пов'язаних з мінімізацією ризиків.

1

Процес залучення зацікавлених сторін (стейкхолдерів) (Stakeholder Engagement Process, SEP)

- Визначити всі потенційно зацікавлені сторони та категоризувати їх на основі рівнів інтересу, впливу, вразливості та потенційної шкоди для них (наприклад, ваші користувачі, особи з особливими потребами, національні меншини, громадяни, які стикаються з дискримінацією, органи державної влади, правозахисники, громадські об'єднання, медіа тощо) – усі вони так чи інакше можуть впливати на роботу вашої системи ШІ.
- Розробити стратегію ефективної взаємодії з кожною зацікавленою стороною / групою зацікавлених сторін.
- Обрати найефективніші способи взаємодії.
- Оцінити, наскільки команда розробників може розбиратися в правах людини, чи здатна вона мати певні обмеження в цих знаннях, що може спричинити появу упереджень у системі ШІ.
- Залучити за потреби правозахисні організації, громадські об'єднання, спілки, асоціації, щоб ідентифікувати потенційні ризики для зацікавлених сторін.
- Забезпечити можливість надати зворотний зв'язок для зацікавлених осіб.

Приклад:

- Серед зацікавлених сторін – користувачі платформи, правозахисні організації, рекламодавці та сама команда розробників. Команді бракує експертизи з прав людини, тому необхідно залучити сторонніх фахівців і поспілкуватися з представниками різних мовних і культурних груп.
- Для уточнення критеріїв модерації та запобігання дискримінації проводять фокус-групи з користувачами та консультації з правозахисними організаціями. Отримані пропозиції і нарікання фіксуються для подальших змін в алгоритмі.

2

Оцінка ризиків і впливів (Risk and Impact Assessment, RIA)

3

- Контекстуалізувати ризики, детально оцінити їх масштаб, обсяг, імовірність та можливість відновити попередній стан.
- Кожному потенційному ризику присвойте категорію (низький, середній, високий).

Приклад:

- Система може блокувати нормальний контент, натомість пропускати дійсно небезпечні пости, залежно від того, як алгоритм навчився розпізнавати мову ворожнечі та нецензурні висловлювання.
- Насамперед розглядають масштаб (скільки людей постраждає), імовірність (чи часто виникатимуть помилки) та відновлюваність (чи можна швидко виправити блокування або видалення). Виявлені високі ризики вимагають додаткових технічних й управлінських механізмів контролю.

План заходів з усунення та/або пом'якшення ризиків (Mitigation Plan)

- Насамперед звернути увагу на ризики, де «потенційна шкода» є особливо критичною, там докладіть найбільших зусиль для її недопущення. Саме розуміння цього дає методологія.
- Визначити способи усунення або пом'якшення ризиків.
- Розробити план упровадження таких заходів, ураховуючи рівень тяжкості ризику та ймовірність його настання, опишіть конкретні заходи, строки виконання, відповідальних осіб, бюджет і ресурси тощо.
- При розробці керуватися принципом пріоритету усунення ризику над компенсацією за порушення постфактум.
- Передбачити різні заходи або їх комбінацію з усунення та/або пом'якшення ризиків, адже не всі ризики ви можете повністю усунути. Для цього ви маєте змогу оцінити, чи можна уникнути цього ризику від початку, скоригувавши функціонал або взагалі відмовившись від нього, щоб виключити негативні наслідки. Або чи можете ви запровадити технічні та організаційні механізми, які допоможуть знизити негативний вплив, щоб пом'якшити ризики. Подумайте над тим, якщо потенційний ризик все ж таки відбудеться, як ви можете швидко відновити порушені права та завдану шкоду, як виправити ситуацію та чи можна надати постраждалим особам компенсацію збитків.
- Передбачити процедури комунікації із зацікавленими сторонами, щоб у них була можливість звернутися до вас, з метою перевірки чи перегляду рішень ШІ.

Приклад порушень:

- Щоб уникнути надмірної цензури, рекомендується переглянути структуру тренувального датасету та алгоритм (уникнення ризику). За потреби залучається контроль реальними модераторами (знизити негативний вплив), а авторам видалених постів дається механізм оскарження (можливість відновити порушені права).
- Компанія документує політику модерації, проводить систематичні перевірки вибірок контенту, налаштовує зручний інструментарій для апеляцій і комунікує процедури публічно, щоб користувачі мали прозоре розуміння та засоби правового захисту.

4

Перегляд оцінки впливу та моніторинг (Iterative Review)

- Здійснювати регулярний моніторинг вашої системи ШІ, перевіряйте фактичні результати, запровадьте моніторингові індикатори, які потребуватимуть перегляду оцінки впливу на права людини в разі збільшення кількості скарг на систему ШІ, збільшення відсотка хибних позитивних або негативних рішень, якості обробки даних тощо.
- Адаптувати систему ШІ до змін у процесах чи середовищі, враховуйте, що зі зміною законодавства, появою нових моделей ШІ, зміною в політиках постачальників API, появою нових типів користувачів, зацікавлених сторін (стейкхолдерів) система може потребувати додаткових налаштувань.
- Ініціювати повторну оцінку (COBRA, SEP, RIA) в разі виявлення суттєвих змін або появи нових ризиків, залежно від типу змін ви можете почати оцінку не з початку, а лише з конкретного етапу, наприклад, ви ідентифікували новий спосіб усунення та/або пом'якшення ризиків, вам достатньо почати з останнього етапу.
- Інформувати зацікавлених сторін про зміни, які ви вносите в систему ШІ, зокрема щодо захисту прав людини, це підвищить довіру до вашого проекту і мінімізує ризики.

Приклад порушень:

- Щоб уникнути надмірної цензури, рекомендується переглянути структуру тренувального датасету та алгоритм (уникнення ризику). За потреби залучається контроль реальними модераторами (знижити негативний вплив), а авторам видалених постів дається механізм оскарження (можливість відновити порушені права).
- Компанія документує політику модерації, проводить систематичні перевірки вибірок контенту, налаштовує зручний інструментарій для апеляцій і комунікує процедури публічно, щоб користувачі мали прозоре розуміння та засоби правового захисту.

5

Ураховуючи, що команда розробників не може мати сильної експертизи в питаннях оцінки впливу на права людини, забезпечте, щоб у команді були фахівці з різних галузей та сфер спеціалізації. Також ви можете провести регулярне навчання, щоб уся команда розуміла не лише технічні аспекти, а й правові, етичні та соціальні виклики, пов'язані із системою ШІ.

Зберігайте всі етапи оцінки ризиків, результати залучення зацікавлених осіб, обґрунтування ухвалених рішень тощо. Це підвищує прозорість процесу розробки.

Розгляньте впровадження зовнішніх аудитів, оцінки упереджень системи, щоб упевнитися в дотриманні принципів прав людини, недискримінації та етичного використання системи ШІ.

4.3. Запобігання дискримінації та упередженням

На всіх етапах життєвого циклу ШІ важливо забезпечувати рівність та запобігати всім проявам дискримінації, що часто потребує активних кроків від розробників. Вони можуть передбачати як упровадження технічних рішень, так і розробку протоколів і процедур для забезпечення рівності на рівні адміністрування. Однією з ключових вимог до розробників є робота з наборами даних. Це передбачає два основні кроки:

- перевірка наборів даних щодо наявності інформації, яка відображає інституційну суспільну дискримінацію чи індивідуальні упередження. Такі дані мають бути видалені із загального набору або спеціально позначені як приклади неприйнятної поведінки. За можливості також краще не використовувати біометричних даних, які найчастіше є категорією даних, на основі якої виникають упереджені результати;
- перевірка збалансованості набору даних та належної репрезентації всіх груп, класів або категорій, на права яких може вплинути діяльність майбутньої системи ШІ. Якщо було виявлено недостатню репрезентацію певної групи, класу або категорії даних, набір варто збалансувати додатковими даними або – якщо неможливо отримати такі дані – доповнити синтетичними даними, щоб фінальний набір містив приблизно рівномірну кількість даних у кожній категорії. По суті, йдеться про впровадження принципу «різноманіття-за-дизайном» (англ. **diversity-by-design**).

Наступним кроком є перевірка наявності небажаних кореляцій між характеристиками (наприклад, чи впливають стать або раса на результати роботи моделі). Для аналізу даних можуть використовувати такі інструменти: Fairness Indicators (Google), AI Fairness 360 (IBM), Fairlearn (Microsoft), What-If Tool (Google), SHAP/ LIME.

Розробникам варто використовувати методи **контрольованого навчання** (англ. supervised learning), **а також постійного фахового людського нагляду** (особливо у високоризикових системах). Зокрема, **дослідження** IBM доводять, що методи навчання без нагляду (як-от навчання з підкріпленням) більш схильні до посилення упереджень, ніж контрольовані види навчання систем ШІ. Аналогічна проблема може виникати з навчанням на вхідних даних від користувачів. У такому разі розробникам варто встановити базові фільтри, що не допускають обробки явно дискримінаційних запитів системою ШІ.

Для уникнення та зменшення упереджень моделі варто **обережно проводити вибір та коригування функціональностей системи ШІ**, зокрема шляхом застосування методів масштабування ознак (англ. feature scaling) та інженерії ознак (англ. feature engineering) у машинному навчанні. Корисним заходом може бути змагальне виявлення упереджень (англ. adversarial de-biasing), у межах якого модель, що тестує стосовно наявності упереджень, тренується паралельно з моделлю, яка напрацьовує захист від або відповідь на упередження.

Оскільки оцінка моделі має проводитися на всіх етапах розробки системи ШІ (і всіх етапах її життєвого циклу загалом), доцільно **розробляти і впроваджувати метрики упередженостей** (англ. metrics of bias). Такі метрики дають змогу встановити «червоні лінії» щодо того, коли кількість і значущість упереджень у моделі є критичними, а також підтримувати один стандарт оцінки та порівнювати динаміку конкретних систем ШІ щодо подолання чи, навпаки, накопичення упереджень. Одним з прикладів таких метрик є **метрика для Vertex AI**.

Високоризикові моделі ШІ мають бути здатні **пояснювати й належно обґрунтовувати** результати своєї роботи (принцип пояснюваності). Це дасть змогу чітко розуміти причини упереджень у моделі й відстежувати, які саме зміни і на якому етапі розробки варто внести для усунення прогалин та несправностей.

Окрім того, розробникам завжди варто **орієнтуватися на контекст ринку**, для якого створюється система ШІ. Зокрема, треба брати до уваги історичні, культурні, мовні та інші особливості, а також дослідити приклади чутливих тем для певної країни чи регіону. Особливо актуально це для систем із підвищеним ризиком дискримінації та упереджень (наприклад, генеративного ШІ, систем, що використовують у сфері судочинства чи працевлаштування тощо). У разі запуску вже готової моделі в нових контекстах варто здійснити додаткову перевірку й адаптувати її до особливостей і вимог конкретного середовища (у разі потреби).

Виявляти упередження та контекстуалізувати моделі допоможе **залучення сторонніх незалежних експертів**, що працюють з темами протидії дискримінації та упереджень. Серед іншого, це передбачає співпрацю з громадянським суспільством, академічною спільнотою, незалежними аудиторами та іншими гравцями у сфері ШІ. Залучення таких гравців може мати різну інтенсивність: від ad hoc звернень до повної співпраці на всіх етапах розробки системи ШІ. Про потребу залучати різних стейкхолдерів **говорить і HUDERIA** — методологія оцінки впливу систем ШІ на права людини.

Зрештою, коли неможливо забезпечити видалення всіх упереджених даних, збалансувати набори даних чи належно промаркувати їх для запобігання дискримінації, розробники мають **робити дисклеймер для користувачів про можливі упереджені результати** роботи системи. Дисклеймер має супроводжуватися посиланням на портал для зворотного зв'язку, за допомогою якого користувачі або треті особи, чиї права були порушені через роботу системи ШІ, можуть звернутися до розробників та повідомити про інцидент.

4.4. Питання фахового людського нагляду

Фаховий людський нагляд є однією з ключових вимог для забезпечення відповідальної розробки, особливо, коли йдеться про високоризикові системи ШІ (**стаття 14 Акта про ШІ**). У разі організації нагляду за навчанням та іншими етапами життєвого циклу системи ШІ ключовим є правильний вибір моделі навчання (такий, що уможлиблює людський нагляд) та впровадження внутрішніх процедур (розробка протоколів дій, призначення відповідальних осіб, упровадження системи підзвітності тощо). Фаховий нагляд є особливо актуальним у випадках, коли ризики для фундаментальних прав неможливо прибрати чи пом'якшити альтернативними способами.

Насамперед варто **чітко розподілити обов'язки щодо нагляду** за системою ШІ. Залежно від масштабу функціональності моделі, кількості у сфері та ризикових питань, які були виявлені під час проведення оцінки впливу на права людини та оцінки ризиків, варто визначити: кількість людей, відповідальних за здійснення фахового нагляду, чіткі сфери їх відповідальності залежно від експертизи й профілю, систему підзвітності, кризовий протокол на випадок інцидентів. Детальні технічні вимоги до здійснення фахового нагляду можна буде знайти, наприклад, у **стандартах Міжнародної організації зі стандартизації (ISO)**, які мають затвердити незабаром.

Також необхідно запровадити **зручний, доступний та зрозумілий інтерфейс для взаємодії** моделі з людиною, що здійснює нагляд. Такий інтерфейс має давати змогу наглядачеві належно і пропорційно виконувати такі дії:

- розуміти можливості й обмеження системи ШІ та мати змогу контролювати її роботу, зокрема в контексті виявлення та усунення аномалій, дисфункцій і неочікуваних результатів;
- залишатися свідомим тенденції надмірного покладання на вихідні дані, надані системою ШІ;
- правильно інтерпретувати результати роботи системи ШІ, враховуючи доступні інструменти та методи інтерпретації;
- ухвалювати рішення в будь-якій конкретній ситуації, не використовувати системи ШІ або інакше ігнорувати, скасовувати чи змінювати її результати;
- втручатися в роботу системи ШІ або переривати її роботу за допомогою кнопки «стоп» або подібної процедури, яка зупиняє систему в безпечному стані.

Необхідно **належно фіксувати всю інформацію**, яка збирається чи виявляється в процесі здійснення нагляду за роботою систем ШІ. Зокрема, варто впроваджувати уніфіковані форми для збору інформації про несправності чи інциденти, які згодом будуть аналізуватися для оновлення моделей та усунення недоліків.

Важливо розробити механізм фахового нагляду так, щоб він відповідав ризикам, рівню автономності системи ШІ та контексту роботи. Це передбачає впровадження таких механізмів нагляду **на ранніх етапах розробки системи**, а також створення **окремих механізмів нагляду для майбутніх користувачів** системи ШІ. Коли йдеться про системи віддаленої біометричної ідентифікації, стаття 14 Акта про ШІ наголошує на необхідності валідувати результати перевірок систем ШІ щонайменше двома фахівцями з відповідними компетенціями, навичками та повноваженнями.

Ефективним механізмом для покращення роботи системи є **широкомасштабний збір зворотного зв'язку за допомогою окремих порталів** для скарг та пропозицій. На додачу до відкритого порталу, де кожен має змогу залишити зворотний зв'язок, розробники можуть акумулювати дані щодо взаємодій користувачів із системою ШІ (з обов'язковим дотриманням правил щодо захисту персональних даних та іншої конфіденційної інформації).

Зрештою, для покращення роботи експертів, що здійснюють нагляд за роботою системи ШІ, варто регулярно проводити **тренінги з підвищення кваліфікації**, що передбачає, зокрема, роботу з місцевими партнерами щодо підвищення обізнаності про зміну контексту на ринку / в середовищі, де оперує система ШІ тощо.

4.5. Захист прав споживачів

Українське законодавство

Якщо розробник надає послуги споживачам — фізичним особам, на відносини між ними поширюються норми законів України «Про цифровий контент та цифрові послуги», «Про захист прав споживачів», «Про електронну комерцію». Зупинимось коротко на найважливіших вимогах указаних законів, застосованих до таких відносин.

Закон «Про цифровий контент та цифрові послуги», який діє з 2 березня 2024 року.

Він містить низку основоположних визначень:

цифрова послуга — послуга, що надає можливість споживачу створювати, обробляти, зберігати та поширювати дані в цифровій формі або отримувати доступ до таких даних, а також здійснювати будь-які інші дії з даними в цифровій формі, що були створені чи завантажені споживачем або іншими користувачами такої послуги. До цифрових послуг належать, зокрема, такі, що дають змогу створювати, обробляти, отримувати доступ або зберігати дані в цифровій формі, включно з хостингом файлів, обробкою текстів або гри, які пропонуються в середовищі хмарних обчислень і соціальних мережах;

товар із цифровими елементами — рухома річ, що містить у собі або пов'язана з використанням цифрового контенту та/або цифрової послуги таким чином, що відсутність цифрового контенту та/або цифрової послуги унеможлиблює виконання таким товаром його функцій

цифровий контент — дані, які створюються і надаються в цифровій формі. До цифрового контенту належать, зокрема, комп'ютерні програми, застосунки, відеофайли, аудіофайли, музичні файли, цифрові ігри та електронні книги

цифрове середовище — апаратне, програмне забезпечення та будь-яке мережеве під'єднання, що використовується з метою отримання доступу до цифрового контенту та/або цифрової послуги й забезпечує можливість їх використання споживачем

інтеграція — поєднання цифрового контенту та/або цифрової послуги з елементами цифрового середовища споживача з метою використання цифрового контенту та/або цифрової послуги згідно з вимогами щодо відповідності, визначеними цим Законом

інтероперабельність — придатність цифрового контенту та/або цифрової послуги до взаємодії з апаратним чи програмним забезпеченням, відмінним від того, що зазвичай використовується із цифровим контентом та/або цифровою послугою того самого типу

модифікація — зміна виконавцем цифрового контенту або цифрової послуги, що надаються споживачу безперервно протягом певного строку (зокрема, оновлення, покращення цифрового контенту або цифрової послуги тощо)

сумісність — придатність цифрового контенту та/або цифрової послуги до взаємодії з апаратним чи програмним забезпеченням, що зазвичай використовується з таким цифровим контентом та/або цифровою послугою, без необхідності їх перетворення

функціональність — придатність цифрового контенту та/або цифрової послуги до виконання своїх функцій відповідно до призначення цифрового контенту та/або цифрової послуги

При цьому розрізняються суб'єктивні та об'єктивні критерії відповідності цифрового контенту та/або цифрової послуги, правові наслідки невідповідності такого контенту чи послуги, строки та способи задоволення вимог споживача, водночас тягар доведення покладається на виконавця.

За певними винятками, якщо договором передбачено разове або кількаразове надання цифрового контенту та/або цифрової послуги, виконавець несе відповідальність за будь-яку невідповідність вимогам, передбаченим Законом, яка існувала на момент надання цифрового контенту та/або цифрової послуги і була виявлена протягом 2 років з моменту надання цифрового контенту та/або цифрової послуги. Якщо ж договором передбачено безперервне надання цифрового контенту та/або цифрової послуги упродовж установленого договором строку, виконавець несе відповідальність за будь-яку невідповідність вимогам закон, яка існувала та/або була виявлена протягом такого строку.

Варто підкреслити, що вказаний закон не застосовується, якщо передання програмного забезпечення або надання виконавцем доступу до програмного забезпечення відбувається **на підставі безоплатної публічної ліцензії з відкритим вихідним кодом, якщо персональні дані, надані споживачем, використовуються виконавцем виключно з метою** вдосконалення надійності, сумісності та інтероперабельності такого програмного забезпечення.

Закон «Про захист прав споживачів». Цей закон застосовується до відносин між споживачем та розробником, які не врегульовані Законом «Про цифровий контент та цифрові послуги». Закон «Про захист прав споживачів» наразі не згадує цифрові послуги, проте є загальною підставою для захисту прав фізичних осіб — споживачів і встановлює вимоги щодо якості послуг, зокрема, наданих цифровими платформами. Він зобов'язує розробників надавати точну інформацію про послуги, забезпечувати їх безпеку та відповідність встановленим стандартам.

Варто звернути увагу, що **після припинення чи скасування воєнного стану набуде чинності нова редакція** вказаного закону (№ 3153-ІХ від 10.06.2023 р.). Така редакція міститиме поняття, близькі до наведених вище визначень із закону «Про цифровий контент та цифрові послуги», зокрема

цифрова послуга — послуга, що дає змогу споживачу створювати, обробляти, зберігати або отримувати доступ до даних у цифровій формі; поширювати або будь-яким іншим чином використовувати дані в цифровій формі, що завантажені чи створені споживачем або іншими користувачами такої послуги

товар із цифровими елементами — будь-яке рухоме майно, яке включає або взаємопов'язане із цифровим вмістом або цифровою послугою таким чином, що відсутність цифрового вмісту або цифрової послуги унеможлиблює використання товару за призначенням;

цифровий вміст — дані, які виробляються, відтворюються, зберігаються і надаються в цифровій формі

інтероперабельність цифрового вмісту — здатність товару із цифровим вмістом функціонувати з апаратними засобами чи програмним забезпеченням, відмінними від тих, з якими зазвичай використовуються товари із цифровим вмістом одного типу

сумісність цифрового вмісту — здатність товару із цифровим вмістом функціонувати з апаратними засобами чи програмним забезпеченням, з якими зазвичай використовуються товари із цифровим вмістом такого самого типу, без необхідності перетворення товарів, апаратних засобів чи програмного забезпечення

вимоги до **Єдиного державного вебпорталу для споживачів у сфері електронної комерції** (Порталу е-покупець), який має бути створений не пізніше ніж через 3 роки з дня набрання чинності цим законом. Після введення в експлуатацію Порталу е-покупець суб'єкти електронної комерції, що реалізують товари, виконують роботи, надають послуги з використанням інформаційно-комунікаційних систем, зобов'язані зареєструватися на Порталі е-покупець протягом 6 місяців з дня введення його в експлуатацію.

Окремо варто зауважити, що закон передбачає надання споживачу інформації про **заходи технічного захисту, наявність та доступність оновлення та про наслідки відмови від його встановлення** (зокрема, при постачанні цифрового вмісту не на матеріальному носії), а також передбачає, що якщо договором про придбання товарів із цифровим вмістом передбачено неодноразове постачання цифрового вмісту або неодноразове надання цифрової послуги, суб'єкт господарювання несе відповідальність за будь-який недолік цифрового вмісту або цифрової послуги, виявлений споживачем протягом 2 років з моменту постачання такого товару, а якщо такий строк постачання перевищує 2 роки, то протягом всього строку постачання цифрового вмісту або надання цифрової послуги.

Закон **«Про електронну комерцію»** регламентує формально-процесуальний бік відносин між розробником та споживачем, оскільки такі відносини реалізуються дистанційно з використанням інформаційно-комунікаційних систем. Зокрема, вказаний закон визначає обсяг інформації, яку розробник та споживач повинні надавати одне одному, вимоги до реклами (комерційних електронних повідомлень), порядок укладання й зберігання електронних договорів (правочинів) тощо.

З огляду на це, розробнику варто створити, оприлюднити та постійно оновлювати документи, які супроводжують увесь життєвий цикл системи ШІ на ринку, як-от рекламних повідомлень, політик використання, оферт, договорів приєднання тощо.

Варто пам'ятати, що в разі порушення прав споживача, він може вдатися до таких способів захисту:

1. засоби захисту з метою отримання відшкодування, як-от звернення з претензією до розробника, з позовом — до суду або ініціювання процедури медіації;
2. засоби захисту з метою покарання розробника (перевірок, накладення штрафів, ініціювання кримінальних справ), зокрема, йдеться про звернення зі скаргою до силових та органів контролю (насамперед, до Державної служби України з питань безпечності харчових продуктів та захисту споживачів, Антимонопольного комітету України, правоохоронних органів, Державної податкової служби України), а також ініціювання широких кампаній, спрямованих на дискредитацію розробника (анти-PR).

Європейське законодавство

Коли надаються послуги споживачам, що є громадянами чи резидентами Європейського Союзу, то варто дотримуватися Закону про цифрові послуги (Digital Services Act — DSA).

DSA охоплює широкий спектр цифрових платформ та посередників, що надають послуги споживачам. Однак залежно від категорії платформи встановлюються специфічні обов'язки:

Дуже великі онлайн-платформи (Very large online platforms, VLOP):

Включають платформи з аудиторією понад 45 мільйонів користувачів у ЄС. Підлягають суворішому регулюванню, яке охоплює:

- Звітування про управління ризиками.
- Проведення незалежних аудитів.
- Забезпечення прозорості алгоритмів і модерації контенту.

Менші онлайн-платформи:

Включають платформи з меншою аудиторією, які мають зменшений обсяг зобов'язань. Їх вимоги можуть бути спрощеними, особливо щодо звітності.

Посередники, що забезпечують технічну інфраструктуру:

Це хостинг-провайдери, мережеві оператори та інші сервіси, які виконують посередницькі функції.

Їх обов'язки стосуються реагування на незаконний контент та надання інформації державним органам у межах закону.

Цей підхід гарантує пропорційне регулювання залежно від впливу та масштабів діяльності платформи.

DSA встановлює конкретні вимоги для розробників цифрових платформ, спрямовані на захист прав споживачів та підвищення прозорості у сфері цифрових послуг. Нижче наведено основні вимоги DSA, які мають виконувати розробники, та практичні рекомендації щодо їх реалізації.

1. Дотримання вимоги прозорості цифрових послуг, зокрема інформування користувачів про алгоритми та автоматизовані системи. Рекомендуємо розробникам забезпечити користувачам чітку інформацію про використання автоматизованих систем. Це передбачає пояснення, як працюють алгоритми персоналізації контенту, а також можливість вимкнення автоматизованих функцій, якщо це можливо. Наприклад ви можете додати автоматичне повідомлення, яке інформує користувачів про те, що вони взаємодіють із ШІ, а також пояснює, на базі чого формуються відповіді.
2. Механізми подання скарг та апеляції та запровадження ефективних механізмів скарг. На платформах повинні створюватись зручні та доступні механізми для подання скарг користувачами та оскарження автоматизованих рішень. Важливо також забезпечити оперативний розгляд скарг та можливість користувачам звернутися до незалежних органів.

Також важливо забезпечити прозорість рекламних оголошень, пояснюючи користувачам, що вони бачать рекламу, хто є її замовником і на яких умовах вона обирається для показу. Платформи повинні мати помітне маркування рекламного контенту та спеціальну вкладку або розділ, де користувачі можуть отримати інформацію про рекламу й надавати можливість для користувачів контролювати, чи вони хочуть бачити персоналізовану рекламу.

Закон про цифрові ринки (DMA) спрямований більше на захист конкуренції та бізнес-користувачів, ніж на прямий захист прав споживачів. Проте деякі положення впливають на права споживачів. зокрема, це стосується прозорості даних. Платформи повинні інформувати споживачів про використання їхніх даних і доступ до них третіх сторін, забезпечуючи їхню конфіденційність та безпеку. Також це стосується доступу до альтернатив. Споживачі отримують вигоду від зменшення монополістичних практик, що дає більше вибору та рівні умови доступу до цифрових послуг. Розробники повинні забезпечити користувачам можливість налаштовувати інтерфейс платформи під власні потреби, включно з контролем над видачею контенту та пріоритетами його відображення. Це передбачає можливість обирати, які категорії контенту чи товарів показувати, змінювати порядок їх пріоритезації, а також налаштовувати алгоритми рекомендацій відповідно до своїх уподобань. Такі функції повинні бути доступними на маркетплейсах, щоб забезпечити споживачам більше контролю над тим, як їм пропонується контент чи товари.

Окремо варто звернути увагу на заборону використання так званих **темних патернів**. Пункт 67 преамбули DSA визначає темні патерни як «практики, які суттєво спотворюють або погіршують, навмисно чи фактично, здатність одержувачів послуги приймати самостійні та усвідомлені вибори чи рішення», а стаття 25 забороняє їх використання. Аналогічні заборони містяться в The Unfair Commercial Practices Directive (стаття 8), DMA (стаття 13), The Data Act (пункт 38 преамбули), The Consumer Rights Directive (CRD) (забороняє темні патерни для фінансових послуг), European Parliament resolution of 12.12.2023 on addictive design of online services and consumer protection in the EU single market ([2023/2043\(INI\)](#)).

Водночас найповніше наразі правовий концепт захисту споживачів від впливу систем ШІ на підсвідомість розкрито в The Guidelines on prohibited AI practice, опублікованих 04.02.2025 року на виконання, зокрема, статті 5(1)(a), (b), (f) AI Act. Ця стаття визначає як високоризикові та забороняє використання систем ШІ, що впливають на підсвідомість, використовують маніпулятивні та оманливі техніки, експлуатують вразливості, а також розпізнають емоції. Ця заборона застосовується за умови завдання чи обґрунтованої ймовірності завдання істотної шкоди споживачам, залежить від ситуації використання системи ШІ, а для систем ШІ, які розпізнають емоції, — ще й від джерела даних: використання біометричних даних заборонено, а от семантичний аналіз наразі дозволений. При цьому підкреслюється необхідність індивідуальної оцінки кожного випадку.

4.6. Директива про відповідальність за продукцію (Product Liability Directive)

Директива про відповідальність за продукцію (Product Liability Directive) є важливим документом, який доповнює вимоги DSA та DMA. Вона передбачає:

- 1. відповідальність за дефекти продуктів:** Постачальники цифрових послуг та платформ можуть нести відповідальність за шкоду, спричинену дефектами їхніх продуктів чи послуг, включаючи випадки використання алгоритмів чи автоматизованих систем;
- 2. чіткі механізми компенсації:** Користувачі повинні мати можливість звертатися за компенсацією у разі шкоди, спричиненої дефектами продуктів або послуг;
- 3. захист прав споживачів:** Постачальники мають гарантувати прозорість щодо безпеки своїх продуктів та послуг.

Розділ 5. Авторське право при розробці

5.1. Охорона Моделі машинного навчання авторським правом / правом особливого роду

Права на використання Моделі можуть бути отримані компанією на різних правових підставах, зокрема:

- отримання Моделі на умовах ліцензії для розгортання у власному ШІ-проекті
- використання Моделі, як частини сервісу (послуг) третьої сторони
- розробка власної Моделі та отримання майнових авторських прав / прав особливого роду від працівників / підрядників / гіг-спеціалістів, які залучаються до розробки такої Моделі

Залежно від цього, компанія отримує різний обсяг прав щодо Моделі і, відповідно, може використовувати різні правові інструменти для її охорони.

Отримання Моделі на умовах ліцензії для розгортання у власному ШІ проекті.

Це можуть бути умовно «відкриті» безкоштовні ліцензії, адже незважаючи на можливість безоплатного використання Моделі у своїх проектах, можуть мати місце обмеження, зокрема, щодо можливості застосування моделі в певних типах проектів (наприклад медичні консультації). Зазвичай у таких ліцензіях визначається, що в ліцензію передається саме Модель у вигляді окремого об'єкта авторського права, що, зокрема, містить параметри моделі, але не зараховує передавання прав на навчальні матеріали, які використовували при створенні такої Моделі.

Використання Моделі, як частини сервісу (послуг) третьої сторони

У цьому разі компанія отримує можливість імплементувати певну модель у свої сервіси / продукти, додатково додавши свої «знання» у вигляді найкращих практик, текстів, формул тощо. У цьому разі сама Модель як така є частиною сервісу, що надається, і, відповідно, провайдер такого сервісу самостійно визначає умови її охорони.

Розробка власної Моделі

При розробці власної Моделі важливим є питання можливості охорони її об'єктивного вираження авторським правом / правом особливого роду. Модель, або ж нейронна мережа з визначеними параметрами (зсуви / ваги нейронів / умови їх активації) а також відповідним комп'ютерним кодом, що забезпечує надсилання запитів / отримання відповідей від такої нейронної мережі та її навчання, може бути об'єктивно виражена у формі комп'ютерної програми, а також систематизованого набору параметрів такої Моделі.

Питання можливості охорони систематизованого набору параметрів такої Моделі як бази даних авторським правом / правом особливого роду є дискусійним. Частина ж у вигляді комп'ютерної програми може охоронятись авторським правом на умовах, передбачених чинним законодавством.

5.2. Врегулювання відносин з командою в контексті авторських прав / прав особливого роду

Спеціалісти, що працюють над розробками систем штучного інтелекту, можуть бути залучені на умовах трудових договорів / гіг-контрактів, цивільно-правових договорів, або господарських угод з відповідними виконавцями, які перебувають у статусі фізичних-осіб підприємців.

Для компанії, яка здійснює таке залучення, важливо отримати всі необхідні майнові права інтелектуальної власності на результати послуг / робіт, що виконуються в межах розробки конкретних моделей машинного навчання / систем штучного інтелекту.

Ураховуючи, що в таких проєктах має місце використання технологій машинного навчання, ролі учасників та результати їх послуг мають свої особливості, порівняно зі звичайними ІТ-проєктами. Ризики неналежного передання майнових авторських прав від учасників команди до компанії можуть мати місце через некоректність / невідповідність реальним обставинам формулювань у відповідних договорах.

Можливо виокремити такі специфічні ролі в межах розробки проєктів у сфері машинного навчання та зіставлення відповідного вкладу з результатами послуг (ІР-активи), що надаються, поглянувши на них через призму авторського права:

1. Інженер з даних (англ. Data Engineer).

- Ключова функція: забезпечення якісного набору даних.
- Ключові ІР-активи: очищені, структуровані канали даних (data pipelines) та відповідна інфраструктура / описані процеси.

2. Спеціаліст з даних (англ. Data Scientist).

- Ключова функція: отримання інсайтів / аналітики з даних та їх використання.
- Ключові ІР-активи: Інсайти з даних, прогностичні моделі та алгоритми, отримані з даних, що впливають на бізнес-результати.

3. Інженер з розробки моделей машинного навчання (англ. ML Engineer).

- Ключова функція: створення, навчання та впровадження моделей машинного навчання
- Ключові ІР-активи: Моделі машинного навчання та оптимізовані алгоритми, а також код, що їх реалізує.

До проєкту, зазвичай, залучаються й інші спеціалісти, які так чи інакше взаємодіють із самою моделлю машинного навчання, наприклад, створюючи середовище для її масштабування (MLOps інженер), чи створюючи користувацькі інтерфейси, у межах яких відбуватиметься взаємодія моделі з кінцевими користувачами (front end розробник).

Залежно від структури команди та специфіки ролей різні функції можуть перетинатись, і в контексті захисту авторських прав / прав особливого роду та консолідації таких прав на відповідному суб'єкті правовідносин (юридичній особі) ключовим є відображення суті реальних правовідносин з командою у відповідних договорах, розуміння імперативного законодавства, що застосовується до правовідносин, та деталізація умов переходу конкретних прав на результати послуг.

5.3. Умови використання об'єктів права інтелектуальної власності та права особливого роду (sui generis) в наборі даних

До набору даних, на основі якого відбувається навчання та тестування алгоритмів і моделей системи ШІ, можуть входити об'єкти права інтелектуальної власності та права особливого роду (sui generis), права на які належать іншим особам. Об'єкти авторського права та суміжних прав можуть використовуватися шляхом відтворення. Таке використання не завжди підпадає під випадки вільного використання, визначені Законом України «Про авторське право і суміжні права».

Так, розробники систем ШІ зобов'язані забезпечити дотримання прав інтелектуальної власності та прав особливого роду (*sui generis*) інших осіб одним із таких способів:

- отримання дозволу на використання охоронюваного об'єкта, зокрема, на підставі договору з правоволодільцем у процесі навчання системи ШІ. Особливо ефективним такий підхід може бути для невеликих систем генеративного ШІ, під час навчання яких можлива однозначна ідентифікація охоронюваних об'єктів;
- запровадження механізмів, за допомогою яких правовласники об'єктів права інтелектуальної власності та права особливого роду (*sui generis*) можуть захистити свої права на такі об'єкти, зокрема шляхом їх вилучення із системи ШІ. Це може досягатися, наприклад, шляхом подання правовласником відповідного об'єкта заяви про вилучення такого об'єкта із системи ШІ, форма якої затверджується розробником або постачальником такої системи (зокрема, в електронній формі). Така форма повинна забезпечувати можливість указати всю інформацію, необхідну для однозначної ідентифікації об'єкта, права на який порушуються або можуть бути порушені. Зауважимо, що зазначений механізм захисту прав може запроваджуватися як з метою вилучення відповідних об'єктів із датасету для навчання системи ШІ, так і з генерованих нею AI-outputs. Проте варто наголосити, що вилучення об'єктів із генерованих системою ШІ AI-outputs (коли процес навчання системи ШІ вже завершено) є технологічно складним, про що докладніше йдеться в розділі **«Надання права на використання»**. Цей механізм захисту особливо ефективний для великих систем генеративного ШІ, під час навчання яких складно ідентифікувати, які саме охоронювані об'єкти були використані;
- заохочення розробників робити відкритими датасети, на яких тренуються системи ШІ, а також супроводжувати інформацією про використані для навчання об'єкти та джерела їх запозичення.

5.4. Використання суспільного надбання

Варто зазначити, що в окремих випадках дозвіл на використання охоронюваних авторським правом чи суміжними правами об'єктів, а також об'єктів прав особливого роду (*sui generis*) не потрібен, наприклад, у разі використання суспільного надбання, ліцензії та прямої вказівки закону.

Суспільним надбанням насамперед вважається об'єкт авторського права чи суміжних прав, щодо якого скінчився строк чинності майнових прав інтелектуальної власності. За загальним правилом строк чинності майнових прав інтелектуальної власності на твір спливає через 70 років, що обчислюються з 1 січня року, наступного за роком смерті автора. Строк чинності майнових прав інтелектуальної власності на відповідний об'єкт суміжних прав спливає через 50 років, що обчислюються з 1 січня року, наступного за роком, у якому відбувся певний факт, визначений законом, наприклад, здійснення виконання. Водночас зазначені строки можуть варіюватися залежно від юрисдикції.

Строк чинності майнових прав на відповідний об'єкт, згенерований з використанням системи ШІ, спливає за законодавством України через 25 років, що обчислюються з 1 січня року, наступного за роком, у якому він був згенерований. Водночас у більшості інших юрисдикцій наразі такі об'єкти не охороняються.

Також законодавство визначає об'єкти, які не охороняються авторським правом України, наприклад, народну творчість (фольклор), чи затверджені державні символи (ст. 8 Закону України «Про авторське право і суміжні права»).

Ба більше, правовласник може самостійно дозволити вільне використання своїх об'єктів у ліцензії, визначивши різні умови, які не суперечать законодавству. **Варто звернути увагу, що відмова від авторських чи суміжних прав не передбачена законодавством України, тому умови ліцензії ССО 1.0, згідно з якою правовласник відмовляється від своїх прав і передає їх у суспільне надбання, не відповідає законодавству України.**

5.5. Строк правої охорони майнових авторських прав не є синхронізованим на загальноосвітньому рівні

Окремі держави, зокрема Україна та держави-члени ЄС, передбачають загальний строк правової охорони майнових авторських прав протягом життя автора та 70 років, що обчислюються з 1 січня року, наступного за роком смерті автора.

Водночас є держави, які встановлюють строк правової охорони майнових авторських прав протягом життя автора та 50 років, що обчислюються з 1 січня року, наступного за роком смерті автора, а інші, навпаки, встановлюють триваліші строки охорони – наприклад, 80, 90 чи навіть 100 років, що обчислюються з 1 січня року, наступного за роком смерті автора. З переліком таких країн можна ознайомитися **за посиланням**.

Для визначення застосовного строку правової охорони майнових авторських прав використовується спосіб порівняння строків, який передбачений у Бернській конвенції про охорону літературних і художніх творів (Стаття 7(8)): «У будь-якому випадку термін визначається законом країни, в якій запитується охорона; проте, якщо законодавством цієї країни не передбачено інше, цей термін не може бути більш тривалим, ніж термін, установлений у країні походження твору». Це означає, що твір охороняється в країні, у якій запитується охорона, протягом строку, що не перевищує строку, встановленого в країні походження твору, якщо законодавство країни, де запитується охорона, не відмовилося від застосування порівняння строків.

Варто зазначити, що законодавство України не містить такого застереження про відмову. На практиці це означатиме таке: якщо автор із держави А, у якій передбачено строк охорони авторського права протягом життя та 50 років після смерті автора, то строк охорони такого твору в Україні становитиме строк життя автора та 50, а не 70 років після його смерті. У разі звернення за правовою охороною в країні А строк правової охорони буде визначатися національним законодавством країни А, тобто становитиме 50 років після смерті автора.

Щодо об'єктів інтелектуальної власності, для яких відсутні аналогічні міжнародно-правові норми, необхідно орієнтуватися на строки охорони, встановлені законодавством країни, де використовується відповідний об'єкт. Такі строки можуть відрізнятися від тих, що передбачені в законодавстві країни походження відповідного об'єкта інтелектуальної власності.

5.6. Copyright Shield й інші приватно-правові ініціативи захисту прав та охоронюваних законом інтересів користувачів систем генеративного ШІ

Системи генеративного ШІ надають безпрецедентні можливості для своїх користувачів. Водночас є обґрунтовані побоювання щодо потужності таких систем та можливості їх використання для завдання шкоди, а не користі. Забезпечення належних запобіжників для відповідального використання систем генеративного ШІ не обмежується лише законодавчим регулюванням; кожен розробник систем ШІ повинен розробити та впровадити свої політики управління ризиками, які виникають у зв'язку з використанням системи генеративного ШІ, і права інтелектуальної власності мають бути лише одним зі складників цих політик.

Згідно з положеннями ч. 7. ст. 33 Закону України «Про авторське право і суміжні права», передбачено, що в разі, якщо AI-output є результатом використання об'єктів авторського права та/або об'єктів суміжних прав, відповідний суб'єкт користується правом особливого роду (*sui generis*) на такий AI-output за умови дотримання ним прав суб'єкта (суб'єктів) авторського права або суміжних прав, твір (твори) або об'єкт (об'єкти) суміжних прав якого зазнав (зазнали) використання у процесі генерування такого AI-output. У разі, якщо AI-output є результатом використання іншого охоронюваного AI-output, відповідний суб'єкт користується правом особливого роду (*sui generis*) на цей AI-output, за умови дотримання ним прав суб'єкта (суб'єктів), AI-output якого зазнав використання у процесі генерування нового AI-output.

Проте питання захисту прав та охоронюваних законом інтересів користувачів систем генеративного ШІ від можливих претензій та позовів третіх осіб залишилося відкритим. Як наслідок, окремі розробники систем ШІ самостійно почали запроваджувати відповідні юридичні механізми.

- Під час використання платних версій комерційних служб Microsoft Copilot і Bing Chat Enterprise для отримання AI-output, щодо якого подається претензія, користувач не має вимикати, оминати, виводити з ладу фільтри вмісту, обмеження в Метапідказках (інструкції, закодовані в службі генеративного ШІ Microsoft, які надають вказівки службі для створення AI-output) чи інші системи безпеки в складі платних версій комерційних служб Microsoft Copilot і Bing Chat Enterprise або перешкоджати їхній роботі.
- Користувач не змінює, не використовує або не поширює AI-output способом, який, як йому відомо або має бути відомо, з високою ймовірністю призведе до порушення або присвоєння будь-якого майнового права третьої особи.
- Користувач має достатні права, щоб використовувати AI-input (всі дані користувача (дані, включно з текстовими, звуковими, відео- та графічними файлами, а також програмне забезпечення, які надаються корпорації Microsoft користувачем або від його імені під час використання системи ШІ, які користувач надає, позначає, вибирає або вводить для використання генеративною технологією ШІ з метою створення або налаштування AI-input) стосовно платних версій комерційних служб Microsoft Copilot і Bing Chat Enterprise, зокрема будь-які дані клієнта, за допомогою яких він здійснив налаштування моделі, що згенерувала AI-output, стосовно якого подано претензію.
- Претензія не передбачає, що AI-output у тій формі, у якій він використовується в торгівлі або професійній діяльності, порушує права на торговельну марку або пов'язані з нею права третьої особи.

Компанія Adobe також має власну політику та умови відшкодування своїм користувачам шкоди за порушення прав інтелектуальної власності в AI-output системи ШІ Firefly, що генерує зображення, а саме:

- Відшкодування Adobe поширюватиметься на позови, у яких стверджується, що AI-output Firefly безпосередньо порушують патенти, авторські права, торговельні марки, права на рекламу або права на недоторканність приватного життя будь-якої третьої особи.
- Відшкодування Adobe не поширюватиметься на претензії, якщо будь-яка претензія ґрунтується на діях або контенті користувача або впливає з них, що включає, без обмежень, такі випадки:
 - модифікацію користувачем AI-output системи Firefly за допомогою будь-якого продукту або послуги, включно з редагуваннями, зробленими за допомогою продуктів/послуг Creative Cloud;
 - будь-які інші матеріали, які використовує користувач в поєднанні з AI-output системи Firefly;
 - будь-яке використання AI-output системи Firefly з порушенням клієнтської ліцензійної угоди;
 - контекст, у якому користувач використовує AI-output системи Firefly;
 - будь-яке використання AI-output системи Firefly після того, як Adobe повідомить користувача про припинення використання відповідних AI-outputs.

Схожі підходи до компенсації запроваджені також компанією Shutterstock при використанні згенерованих результатів роботи системи Shutterstock's AI Image Generator.

6 листопада 2023 року компанія OpenAI (розробник та правоволоділець систем генеративного ШІ ChatGPT та DALL-E) задекларувала впровадження функції захисту прав користувачів ChatGPT Enterprise від претензій третіх осіб щодо порушення їхніх авторських прав.

Зазначені вище випадки щодо запровадження договірного регулювання сфери використання AI-outputs на рівні приватно-правових ініціатив свідчать про щонайшвидшу необхідність визначення уніфікованого балансу інтересів авторів творів, з одного боку, та розробників (користувачів) систем генеративного ШІ — з іншого.

⁵ [Умови використання продукту Microsoft](#)

⁶ [Поширені запитання про юридичні послуги Firefly](#)

⁷ [Генератор зображень зі штучним інтелектом](#)

⁸ [Нові моделі та продукти для розробників, анонсовані на DevDay](#)

Розділ 6. Приватність

6.1. Ключові українські та іноземні регуляторні інструменти, які охоплюють питання обробки ПД з використанням ШІ

Перед створенням нових продуктів розробникам потрібно проаналізувати застосовне регулювання у сфері захисту персональних даних (ПД) у державі, де зареєстрована компанія, а також державах, на чії ринки планується випускати продукти. Оскільки для України актуальними є національний ринок, правовий простір Європейського Союзу та часто американський ринок, у цих Рекомендаціях ми детальніше зупинимося саме на цих трьох регуляторних рамках.

Українське регулювання. Ключовим актом в українському контексті є Закон України **«Про захист персональних даних»**. Окрім загальних положень про юридичний статус особи (володілець та розпорядник) та норм щодо того, яка інформація належить до персональних даних, розробникам систем ШІ варто пам'ятати про:

- обов'язок чітко визначати мету обробки ПД (наприклад, тренування моделі або створення бази даних) і при зміні мети обов'язково повідомляти суб'єктів даних та отримувати згоду на обробку ПД з іншою метою (якщо дані обробляються на підставі згоди особи);
- коректно обирати **правову підставу** для обробки ПД;
- потребу перевірки точності, достовірності і релевантності ПД, їх ненадмірності відносно мети обробки (принцип мінімізації даних);
- необхідність забезпечення прав суб'єкта ПД (зокрема, на доступ до своїх ПД, інформацію про спосіб і порядок їх обробки, заперечення проти обробки даних, вимога видалення даних, відкликання згоди на обробку даних, захист від автоматизованих рішень, які мають правові наслідки для суб'єкта ПД тощо);
- вимогу повідомляти про передачу персональних даних третім особам із зазначенням підстави передачі даних та прав суб'єкта в цьому контексті;
- обов'язок забезпечувати захист ПД шляхом упровадження технічних та організаційних заходів безпеки даних.

Окрім цього, одним з важливих інструментів є Закон України **«Про доступ до публічної інформації»**, який дозволяє отримувати від розпорядника публічну інформацію за запитом. У випадку запитів, що стосуються персональних даних, знеособлення цих даних здійснює відповідний розпорядник (частина 3 статті 10) або дані надаються у випадках, якщо вони належать до категорії відкритих даних. Перелік даних, які можна отримати у форматі відкритих даних, міститься у відповідному **Положенні**. Для розробників це є актуальним у контексті розробки систем ШІ, що використовуються для аналізу публічної інформації, або таких, що використовуються в публічному секторі.

Уповноважений ВРУ з прав людини затвердив **Типовий порядок обробки персональних даних** та **Роз'яснення** до нього. Ці документи, зокрема, пояснюють, як організувати збір ПД у законний спосіб, а також наводять вимоги до згоди особи на обробку її ПД та процедури для здійснення ефективного внутрішнього контролю і нагляду за процесами, пов'язаними з обробкою ПД.

Важливо пам'ятати, що в межах євроінтеграційної реформи **планується гармонізація українського законодавства з вимогами ЄС**, зокрема із Загальним регламентом про захист даних (далі – GDPR). Наразі є декілька законопроектів, що пропонують упровадити європейські вимоги.

⁹ Зокрема, [законопроект 8153](#) (про захист ПД) і [законопроект 6177](#) (про орган контролю у сфері захисту ПД). Їх норми є предметом дискусій, тож радимо розробникам стежити за законодавчими змінами і вчасно реагувати на зміни в регуляторному полі.

Вимоги в межах Європейського Союзу та Європейської економічної зони (ЄЕЗ). Масив законодавства ЄС/ЄЕЗ складається з регламентів (актів, норми яких мають пряму дію) та директив (механізм імплементації яких розробляють держави-члени ЄС). Виходячи на ринок ЄС/ЄЕЗ або працюючи з компаніями з ЄС/ЄЕЗ, розробники мають зважати й на національні регулювання країн, які можуть встановлювати або додаткові обмеження, або додаткові можливості. У Рекомендаціях ми зосередимося на регулюванні рівня ЄС, а не індивідуальних держав. Основними актами, актуальними для розробників систем ШІ є:

- **Загальний регламент про захист даних (GDPR)**. GDPR встановлює загальні вимоги мінімізації даних та обмеження процесів обробки даних метою їх збору, анонімізації і псевдонімізації ПД, впровадження безпекових практик, а також закріплює права особи на доступ до ПД, пояснення процесів, що відбуваються з даними, і право на забуття. Хоч GDPR не містить прямих згадок систем ШІ, стаття 22 впроваджує вимоги до процесів автоматизованого ухвалення рішень, що, фактично, охоплює діяльність таких систем у більшості випадків їх взаємодії з ПД. У цьому контексті GDPR закріплює право особи не бути підданою автоматизованим рішенням, якщо вони мають для неї правові наслідки або суттєво впливають на положення особи, а отримувати послуги (доступ та користування продуктом) із залученням людини (за винятком випадків, коли така автоматизована обробка закріплена законодавчо). Важливо, що GDPR, на відміну від Акта про ШІ, застосовується екстериторіально – тобто дані суб'єктів на території ЄС, які обробляються в інших країнах, захищені за правилами GDPR.
- **Акт про штучний інтелект (Акт про ШІ)**. Документ встановлює обов'язки щодо відповідального управління даними в системах ШІ з високим ризиком (стаття 10), вимоги щодо прозорості (стаття 13), людського нагляду та моніторингу (статті 14 та 26, відповідно), а також вимога здійснювати оцінку впливу на права людини (стаття 27), яка також передбачає оцінку впливу процесів розробки і самих систем ШІ на ПД. Крім того, відповідно до Преамбули Акта про ШІ, розробникам і постачальникам варто належно захищати набори даних (датасети), де змішані персональні та неперсональні дані.
- **Акт про дані (ДА, набере чинності 12 вересня 2025 року)**. Акт регулює доступ та обмін як персональними, так і неперсональними даними, створеними за допомогою продуктів та цифрових послуг, а також сприяє розвитку конкурентного ринку хмарних послуг. Документ установлює вимоги до обміну даними здебільшого між такими суб'єктами в приватному секторі, як: бізнес-бізнес, бізнес-споживач, бізнес-уряд. До того ж якщо йдеться про персональні дані – Акт містить відсильну норму до GDPR і не знижує рівня захисту ПД;
- **Акт про управління даними (DGA)**. Документ регулює добровільний обмін даними з публічними інституціями, вимагаючи від держав ЄС ефективного законодавчого й технічного оснащення для належної передачі даних (як-от анонімізація ПД, наявність спеціальних контрактів для передання конфіденційних даних (проте лише за згодою суб'єкта даних)). Також Акт впроваджує поняття альтруїзму даних – можливості осіб у волонтерському порядку надавати дані для використання їх для досягнення публічного інтересу (досліджень у сфері медицини чи кліматичних змін, ведення статистики, інших некомерційних цілей);
- **Директива про відкриті дані і повторне використання даних публічного сектору**. За загальним правилом, Директива не застосовна до ПД, на які поширюється режим конфіденційності і захист відповідно до GDPR. І хоча сам по собі документ не унеможливорює повторного використання публічно доступних даних навіть у випадку, коли вони є ПД, це можливо лише в разі обмеження використання даних метою їх збору. Водночас суб'єкти можуть запитувати в державних органів можливість анонімізувати дані для подальшого їх використання. Така анонімізація може бути платною для запитувачів.

Вимоги законодавства США. Американська доктрина захисту ПД суттєво відрізняється від європейських стандартів.

На федеральному рівні США не мають єдиного закону про захист персональних даних, як, наприклад, GDPR у ЄС. Натомість захист даних регулюється низкою спеціалізованих законів й активною діяльністю Федеральної торговельної комісії США (Federal Trade Commission, FTC).

¹⁰ На момент написання відомо про 19 штатів, які ухвалили закони про захист персональних даних. Детальніше: <https://iapp.org/resources/article/us-state-privacy-legislation-tracker/>

Федеральна торгова комісія є ключовим регулятором у сфері захисту персональних даних на федеральному рівні. Основні напрями діяльності FTC:

- **Захист від недобросовісних та оманливих практик:** FTC має повноваження розслідувати й накладати штрафи на компанії, які вводять в оману споживачів або використовують недобросовісні методи обробки даних.
- **Моніторинг виконання законів: FTC забезпечує виконання ключових спеціалізованих законів, як-от:**
 - **Акт про конфіденційність дітей в Інтернеті (COPPA):** Захищає дітей віком до 13 років, зобов'язуючи компанії отримувати батьківську згоду на збір даних.
 - **Акт про портативність страхових медичних даних і підзвітність (HIPAA):** Регулює обробку та захист медичної інформації.
 - **Акт про захист фінансових даних (GLBA):** Зобов'язує фінансові установи інформувати клієнтів про обробку даних та забезпечувати їх захист.
 - **Акт про кредитну звітність (FCRA):** Встановлює стандарти для обробки кредитної інформації та надає права споживачам на доступ і виправлення даних.

На рівні штатів одним з найбільш відомих (і вимогливих з погляду стандартів) є **Каліфорнійський закон щодо приватності споживачів** (CCPA), утім, важливо пам'ятати, що на рівні штатів є великий масив інших актів, які можуть бути актуальними для розробників залежно від того, у якій юрисдикції вони планують працювати і для якого ринку створюють продукти. У цих Рекомендаціях ми детальніше зупинимося на вимогах CCPA як таких, що встановлюють один з найвищих стандартів в американській практиці:

- CCPA не розглядає загальнодоступної інформації з державних реєстрів на федеральному рівні, рівні штатів чи місцевому рівні як ПД (як-от професійні ліцензії та публічні записи про нерухоме і рухоме майно);
- CCPA не вважає ПД дані, які були псевдонімізовані та анонімізовані або агреговані та анонімізовані, оскільки їх неможливо прямо пов'язати з особою;
- CCPA зараховує до ПД дані не лише про особу, а й про домогосподарства;
- Акт передбачає права знати про обробку ПД, вимагати видалення, зміни чи обмеження використання ПД, право відмовитися від продажу чи передачі ПД третім особам та право не бути дискримінованим при реалізації інших прав, пов'язаних з ПД.

6.2. Інвентаризація даних та приватність за замовчуванням / проєктуванням

Оскільки використання персональних даних у наборах даних є їх обробкою і може бути надчутливою темою для суб'єктів ПД, розробникам, перш ніж працювати з конкретним набором даних, що можуть містити персональні відомості, необхідно провести його інвентаризацію, а також забезпечити належну імплементацію концептів приватності за замовчуванням та за проєктуванням в операції з обробки ПД з використанням ШІ / тренування Моделей машинного навчання (Моделей) з використанням ПД.

Інвентаризація персональних даних перед їх використанням в ШІ-проєктах.

Під інвентаризацією мається на увазі встановлення типу даних (конфіденційні / персональні / відкриті, тощо), типів запланованих операцій, цілей та способів використання.

Використання (обробка) ПД в Моделях може мати різні прояви, зокрема, але не виключно:

- використання ПД для навчання Моделі;
- використання ПД в межах взаємодії Моделі з користувачем (через відповідний інтерфейс у кінцевому продукті);
- використання ПД для збагачення (додавання нової інформації чи атрибутів до наявних даних на основі передбачень) даних про конкретного користувача шляхом використання Моделі.

Незалежно від способу та мети використання, перш ніж почати відповідні дії, корисно зробити таке:

1. Ідентифікація даних та їх класифікація:

- отримання розуміння щодо джерел даних — звідки саме взято конкретний набір даних та відповідні дані (додатки / бази даних / консолідація тощо);
- класифікація даних та визначення їх ризиковості — необхідне розуміння типів даних у наборах даних, а також наявності чутливих даних.

2. Оцінка правових підстав для обробки даних:

- розграфування правових підстав для обробки всіх відповідних ПД у відповідних наборах даних (згода / легітимний (законний) інтерес тощо);
- додаткова оцінка належності підстав перед використання ПД в запланованих активностях та перевірка можливості таких дій у контексті мети обробки ПД.

3. Перевірка точності / повноти даних:

- перевірка точності, релевантності даних та уточнення застарілої / неактуальної інформації.

4. Карти руху ПД та документація:

- розграфлення деталізованого маршруту / маршрутів руху та обробки ПД в межах запланованих операцій;
- визначення відповідальних осіб за використання ПД в конкретних операціях та умови їх доступу / роботи з такими даними на відповідному етапі.

5. Мінімізація даних:

- визначення й видалення інформації, що не є критично необхідною для запланованих операцій;
- анонімізація / псевдонімізація даних.

6. Оцінка безпеки в роботі з даними:

- оцінювання організаційних та технічних заходів захисту ПД, зокрема в контексті доступів і технічних заходів, як-от, зокрема, шифрування та імплементація відповідних заходів.

7. Ідентифікація та забезпечення відповідності регулюванню, що застосовується:

- скільки закони / акти про захист даних часто мають екстериторіальну дію, необхідним є аналіз не лише локального законодавства компанії, що планує здійснення відповідної операції а й відповідних нормативних актів країн, де планує оперувати / оперує компанія.

8. Побудова процесів роботи із запитам суб'єктів даних:

- визначення прав суб'єктів даних щодо своїх даних відповідно до регулювання, що застосовується, та пропрацювання алгоритму / процедур реалізації таких запитів.

9. Визначення умов та строків знищення даних / завершення операцій:

- документування умов припинення запланованих операцій з даними / припинення їх використання та запланованого видалення / вичерпання мети обробки.

10. Документування процесів:

- створення відповідного звіту щодо проведення інвентаризації даних та планування нових інвентаризацій / процедур, умов, обставин (тригерів) їх проведення (зміна умов обробки / наявність скарг / зміна мети тощо).

11. Визначення можливості / необхідності використання технологій, які посилюють приватність — PETs (privacy enhancing technologies):

- розгляд додаткових опцій забезпечення прав суб'єктів ПД шляхом використання шуму (технологій диверсифікованої приватності) та інших технічних рішень задля унеможливлення / зменшення можливості ре-ідентифікації суб'єктів ПД.

12. Проведення необхідних оцінювань щодо впливу обробки ПД на суб'єктів ПД:

- такі оцінювання, зокрема **DPIA** (оцінювання впливу на захист даних), як це передбачено в GDPR, та інші оцінювання мають на меті виявлення, документування та нівелювання / зменшення ризиків для суб'єктів даних.

До проведення інвентаризації ПД перед їх використанням в різного роду операціях із застосуванням технологій ШІ рекомендовано залучати спеціалістів із захисту даних / DPO (співробітник з питань захисту даних) на етапі початкового планування відповідних операцій.

Приватність за замовчуванням та за проектуванням у ШІ проєктах

Приватність за замовчуванням та за проектуванням (Privacy by Design & by Default / PbD) передбачені в статті 25 GDPR. Ключовою ідеєю PBD вважається імплементація дотримання приватності на всіх етапах проєкту, із самого його початку.

У контексті ШІ проєкту початком може бути планування створення конкретної моделі машинного навчання та пропрацювання умов її тренування й подальшого використання в контексті ризиків для суб'єктів даних.

Приватність має бути взята до уваги в процесі розробки проєкту, наприклад — навчання Моделі з урахуванням дотримання всіх прав суб'єктів даних (зокрема, але не виключно — наявності правової підстави для обробки ПД, впроваджених ефективних алгоритмів реалізації прав суб'єктів даних тощо).

З початком використання Моделі машинного навчання в конкретному проєкті, наприклад, у чатботі Системи генеративного ШІ, необхідний постійний аудит такого використання щодо ризиків для приватності як суб'єктів, чиї ПД були використані для навчання Моделі машинного навчання, так і суб'єктів даних, які взаємодіють з такою Моделлю в контексті автоматизованих рішень щодо них / дискримінації / неправомірного використання їх ПД тощо.

У контексті реалізації PbD для ШІ проєктів, серед іншого, доцільно звернути увагу на таке:

- належне інформування / наявність підстави на обробку даних від суб'єктів ПД на використання їх ПД в межах ШІ-проєктів;
- повідомлення суб'єктам ПД простою та доступною мовою, як саме їх ПД використовуватимуться в Моделі;
- дотримання принципів мінімізації ПД на всіх етапах ШІ проєкту;
- упровадження технічних та організаційних заходів захисту даних;
- моніторинг та постійні документування інвентаризацій / оцінювань / процесів і їх покращення.

6.3. Визначення правових підстав для обробки даних у межах розробки систем ШІ

Однією з основних вимог будь-якого законодавства про захист ПД є наявність правової підстави для обробки даних. Це означає, що перед початком обробки даних володілець (контролер) персональних даних повинен визначити підставу такої обробки та задокументувати її.

Правова підстава для обробки ПД визначається залежно від цілі обробки ПД, статусу суб'єкта, який обробляє дані (наприклад, держава або приватні особи) і категорії даних.

Визначаючи підстави, насамперед важливо зрозуміти, з якими категоріями даних розробник має справу. Відповідно до статті 7 ЗУ **«Про захист персональних даних»**, **особливі вимоги до обробки ПД встановлюються щодо інформації про:**

- расове або етнічне походження;
- політичні, релігійні або світоглядні переконання;
- членство в політичних партіях та професійних спілках;
- засудження до кримінального покарання;
- дані щодо здоров'я, статевого життя;
- біометричні та генетичні дані.

GDPR додає до цього переліку інформацію про сексуальну орієнтацію особи.

Нижче наведено **найбільш релевантні підстави**, на основі яких розробники можуть обробляти ПД на різних етапах життєвого циклу систем ШІ, а також ключові виклики, з якими можна стикнутися, обравши ту чи іншу підставу для обробки даних:

- **Згода особи.** Обробка ПД на підставі згоди є одним з найпоширеніших варіантів серед розробників систем ШІ. Вибираючи цю підставу, варто пам'ятати, що згода має бути добровільною, поінформованою і попередньою – це означає, що особі мають повідомити, які дії вчинятимуть з її ПД, якою є мета такої обробки та її строк. Обробка може здійснюватися лише з тією метою, на яку особа надала згоду. У разі зміни мети обробки варто також отримати згоду від особи. Загалом, за належного оформлення згоди особи, ця підстава для обробки ПД залишає розробника в «безпечній» позиції, оскільки суперечливі ситуації виникатимуть з досить низькою ймовірністю порівняно з іншими підставами. Водночас варто розуміти, що згода не завжди є оптимальною підставою. Наприклад, на етапі тренування систем ШІ можуть виникнути проблеми з огляду на можливість її відкликання. У цьому разі розробник має видалити дані з набору, що може спричинити значне погіршення якості моделі. Крім того, це може вплинути на відстежуваність і прозорість системи ШІ, оскільки цілісність набору даних, на якому тренувалася система, буде порушено.
- **Укладення і виконання правочину.** Обробка ПД може здійснюватися на підставі виконання положень договору, стороною якого є суб'єкт ПД. Водночас обробка ПД має бути необхідною саме для виконання договору або переддоговірних зобов'язань, а не просто згадуватися в договорі як одне з його положень. Ця підстава також не може застосовуватися для обробки ПД третіх осіб, які не є стороною договору. Утім, як і згода, укладення правочину не є універсальною підставою для обробки ПД. По-перше, у разі виконання правочину мета обробки ПД є дуже обмеженою і не може бути змінена (для обробки з іншою метою потрібно буде отримувати згоду особи або шукати іншу правову підставу). По-друге, коли йдеться про обробку особливих категорій даних, може знадобитися згода особи на додачу до укладення договору. Такі ситуації повсякчас виникають у сфері охорони здоров'я, іноді – в освітній сфері.
- **Необхідність захисту законних інтересів володільця ПД.** Ця правова підстава має дві обов'язкові умови: 1) наявність надзвичайної ситуації, у якій інтереси володільця потребують захисту; 2) переважання інтересів володільця над інтересами суб'єкта ПД. Із цього можна зробити висновок, що така підстава, наприклад, незастосовна на етапах розробки системи, адже ситуація не сягає рівня, коли потреба обробки переважає інтереси суб'єкта даних. Водночас ця підстава може бути застосовною в надзвичайних випадках, коли йдеться про інциденти під час застосування системи ШІ. Такі ситуації виникатимуть нечасто і в кожній з них розробнику потрібно буде наводити обґрунтовані докази того, що його інтерес переважає над правами та інтересами суб'єкта ПД. Також важливо, що ця підстава не стосується обробки особливих категорій даних.

Вибираючи правову підставу для обробки ПД, розробники систем ШІ мають надати обґрунтування щодо її відповідності виду діяльності, який планується здійснювати (тренування або тестування моделі, створення бази даних, використання системи ШІ тощо). Також **необхідно зберігати всі супровідні документи, що підтверджують наявність правових підстав на обробку даних**: наприклад, згода особи в письмовій формі або договори з державними органами, які надають доступ до ПД для розробки систем ШІ для державних потреб в межах державно-приватного партнерства.

В американській системі за загальним правилом відсутня вимога щодо наявності правових підстав для обробки ПД, але вводиться обов'язок надати **можливість відмовитися від продажу чи передачі ПД третім особам**. У виняткових випадках є вимога згоди на обробку ПД дітей чи ПД, що обробляються в комерційних цілях. Працюючи з американським законодавством, варто зважати, що регулювання різняться залежно від штату.

6.4. Оцінка впливу на приватність при розробці систем ШІ

Однією з вимог GDPR є проведення компаніями оцінки впливу їх продуктів (процесів, проєктів, систем) на захист даних. У контексті систем ШІ доцільно наголосити, що йдеться про проведення такої оцінки на всіх етапах життєвого циклу систем ШІ. Стаття 35 GDPR указує, що процес оцінки має здійснюватися **до початку будь-якої обробки даних** і є обов'язковим у випадках:

- систематичної та широкої оцінки особистісних характеристик людини, яка ґрунтується на автоматизованій обробці, включно з профілюванням, і на основі якої ухвалюють рішення, що мають юридичні наслідки або суттєво впливають на особу;

- масова обробка чутливих даних або ПД, що стосуються кримінальних засуджень та правопорушень;
- систематичний масовий моніторинг у публічних місцях.

Водночас інші випадки, коли обробка ПД може суттєво впливати на права людини (незалежно від характеру цього впливу) також потребують дотримання вимог щодо проведення оцінки впливу та розробки механізмів для запобігання інцидентам та реагування на них інцидентів (наприклад, кризові протоколи). Стаття 35 GDPR, зокрема, виокремлює такі елементи оцінки впливу:

- систематичний опис операцій обробки ПД та цілей обробки, включно, де актуально, з описом законного інтересу, який переслідує контролер ПД;
- оцінка необхідності та пропорційності операцій з обробки ПД відносно цілей;
- оцінка ризиків для прав і свобод суб'єктів даних;
- розробка і впровадження заходів з усунення ризиків, включно з гарантіями, заходами безпеки та механізмами для забезпечення захисту ПД, беручи до уваги права та законні інтереси суб'єктів ПД й інших зацікавлених осіб.

GDPR при проведенні оцінки ризиків не вимагає залучення третіх сторін, передбачаючи її як внутрішній процес. Водночас Європейська комісія у своєму **дослідженні** щодо ШІ в контексті регулювання із захисту даних наголосила, що важливою є співпраця між різними стейкхолдерами в процесі проведення оцінки ризиків. Зокрема, йдеться про залучення академічної спільноти, представників громадянського суспільства, а також вразливих і маргіналізованих груп, чії права можуть зазнати значного впливу, до проведення таких заходів.

GDPR передбачає, що у випадках, коли результати оцінок свідчать про високі ризики для прав людини, якщо контролер даних не вживатиме певних заходів, контролер має проконсультуватися з національними регуляторами для отримання рекомендацій.

Які питання мають враховуватися, коли проводиться оцінка впливу на рівень захисту персональних даних? Ми пропонуємо невичерпний перелік ключових аспектів, що потребують особливої уваги з боку розробників:

- ризики для суб'єктів ПД, пов'язані з неправомірним використанням даних, що містяться в наборі навчальних даних;
- ризик алгоритмічної дискримінації, спричинений системою ШІ на стадії розробки, наприклад, пов'язаний із нижчою продуктивністю системи для певних категорій;
- ризик створення неправдивого контенту про реальних осіб, який може мати наслідки для їх репутації;
- ризик автоматизованого ухвалення рішень, що виникає, якщо на етапі розробки не вжито заходів для пояснення, як саме функціонує система ШІ, що обмежує здатність користувача перевірити її роботу;
- ризики, пов'язані з кібератаками, характерними для систем ШІ, як-от атаки шляхом отруєння даних або моделі, створення бекдору (прихованого доступу) для зловживань або інверсія моделі;
- ризики, пов'язані з конфіденційністю ПД, які можуть бути отримані із системи ШІ;
- системні та серйозні етичні та правові ризики, пов'язані з розгортанням систем, як-от вплив на демократичні процеси або дотримання основних прав (наприклад, дискримінація);
- ризик втрати користувачами контролю над своїми опублікованими в інтернеті та вільно доступними ПД, оскільки широкомасштабний збір часто необхідний для навчання системи ШІ, зокрема, коли він збирається за допомогою вебскрепінгу.

6.5. Управління процесами, пов'язаними з реалізацією прав суб'єктами персональних даних

Усі розглянуті регулювання¹¹ щодо захисту персональних даних передбачають можливість збереження контролю суб'єкта персональних даних за інформацією про себе. Це можливо, зокрема, шляхом реалізації суб'єктом своїх прав. Акти передбачають такі основні групи прав:

- право бути поінформованим (суб'єкт персональних даних має право знати, які та як дані збираються, з якою метою вони обробляються, хто є їхнім володільцем (контролером), кому дані можуть передаватися та як довго ці дані будуть зберігатися);

¹¹ Розділ зосереджується на Законі України «Про захист персональних даних» («Закон України»), GDPR та CCPA. Закони щодо регулювання захисту персональних даних інших країн або штатів США можуть містити інші права суб'єктів персональних даних або особливості їхньої реалізації. Рекомендується ознайомитись з вимогами локальних регулювань перед початком обробки відповідних персональних даних;

- право на доступ до персональних даних (суб'єкт може запитати інформацію про те, які саме персональні дані про нього обробляються, отримати їхню копію та дізнатися, хто має до них доступ);
- право на виправлення даних (якщо дані є неточними або неповними, суб'єкт має право вимагати їх виправлення);
- право на видалення даних (суб'єкт персональних даних має право вимагати видалення даних про себе, якщо вони більше непотрібні для обробки, обробляються незаконно або суб'єкт відкликав свою згоду на їх обробку (якщо підставою для обробки була згода)); та
- право на заперечення проти обробки (суб'єкт має право заперечувати проти обробки його персональних даних, наприклад, для цілей прямого маркетингу або автоматизованого ухвалення рішень, які мають правові наслідки для нього).

Розподіл обов'язків, пов'язаних з реалізацією прав суб'єктів

Обов'язок сприяння реалізації прав суб'єктами персональних даних повністю покладено на володільця (контролера). У разі спільного контролю над даними¹² або залучення розпорядників (процесорів), рекомендується передбачити у відповідному договорі механізм розподілу функцій та відповідальності за управління процесами, пов'язаними з реалізацією таких прав.

Особливості сприяння реалізації деяких прав суб'єктів розробниками Систем ШІ

При дизайні системи управління процесами, пов'язаними з реалізацією прав суб'єктів, розробникам Систем ШІ варто зважати на специфіку ведення бізнесу у ШІ-сфері. Серед факторів, які впливають на ефективність системи та можливість якісно модерувати здійснення прав суб'єктами, можуть бути такі особливості життєвого циклу Систем ШІ та ШІ-технологій загалом:

- набори даних для потреб розробки, тренування, тестування тощо Системи ШІ можуть отримуватись з джерел, не пов'язаних із суб'єктом. Такий збір інформації може підпадати під спеціальні вимоги щодо інформування суб'єктів персональних даних і вимагати від володільців (контролерів) даних проактивних дій щодо оповіщення суб'єктів. Щоб уникнути порушення застосовного законодавства, рекомендується відстежувати застосовні правила та джерела збору інформації. Це дасть змогу діяти відповідно до вимог застосовного порядку інформування суб'єктів¹³;
- системи ШІ можуть використовувати великі та неструктуровані набори даних, що може ускладнити виокремлення персональних даних, ідентифікацію їхньої належності до конкретного суб'єкта та видалення даних за його запитом. Рішенням може слугувати впровадження механізму виявлення персональних даних у наборах даних, який дасть змогу відстежувати походження та місцезнаходження персональних даних у таких наборах;
- якщо персональні дані були використані для тренування Систем ШІ, їх видалення може вимагати перетренування Системи ШІ, що є або ресурсно затратним або технічно складним. Рекомендується використовувати методи тренування, які передбачають навчання закономірностей, а не поглинання інформації та дають змогу в подальшому видаляти дані без шкоди для Системи ШІ. Такими методами можуть бути покрокове навчання¹⁴ або використання «шлюзів забуття»¹⁵, які вирішують, яку інформацію з довгострокової пам'яті системи потрібно зберігати чи видаляти тощо;
- персональні дані можуть оброблятися на різних етапах життєвого циклу Системи ШІ. Наприклад, персональні дані можуть міститися в наборах даних для тренування, використовуватися під час тестування та перевірки Системи ШІ, використовуватися для прогнозування під час розгортання¹⁶ і навіть міститися в самій Системі ШІ. У разі отримання запиту суб'єкта щодо доступу до даних, зміни, видалення тощо, необхідно розуміти всі місця збереження даних та могли вільно взаємодіяти з ними відповідно до запиту суб'єкта. Рекомендується вести та зберігати журнали даних, використаних для тренування, тестування тощо, щоб полегшити відстеження та видалення інформації в майбутньому;
- псевдонімізовані персональні дані досі залишаються персональними. У разі використання таких даних протягом життєвого циклу Систем ШІ та при отриманні запитів від суб'єктів щодо доступу / видалення даних, можливі труднощі з ідентифікацією такої інформації та асоціацією даних із суб'єктом. Рішенням також може стати ведення журналів використання даних та операцій з ними;

¹² Англ., Joint Controllship відповідно до GDPR;

¹³ Наприклад, стаття 14(3) GDPR зобов'язує повідомляти суб'єкта про збір та умови обробки даних не пізніше 30 днів з моменту збору даних, а у випадку, якщо планується передача даних іншим – до такої передачі;

¹⁴ Англ., incremental learning

¹⁵ Англ., Forget Gate

¹⁶ Англ., deployment

- системи ШІ можуть робити висновки про суб'єкта, спираючись на його персональні дані, або генерувати синтетичні персональні дані про нього (робити припущення високого ступеня ймовірності). Іноді розробники не маркують такої інформації як персональні дані про суб'єкта і, відповідно, не беруть їх до уваги, працюючи із запитом суб'єктів. Рекомендується розглядати таку інформацію, як персональні дані, застосовувати до них такі ж правила захисту, відстежувати їх рух так само, як рух первинних даних, та забезпечити прозорість їх обробки;
- іноді запит суб'єкта на мобільність та перенесення даних до інших провайдерів Систем ШІ може бути складно виконати через унікальні формати даних, інтеграцію даних у складні структури та розграфлення або їх невіддільність від Системи ШІ. Рекомендується використовувати стандартизовані та загальнодоступні в професійній сфері формати даних (наприклад, CSV для табличних даних, JSON для структурованих даних тощо). При дизайні Систем ШІ варто впровадити процедури, які дадуть змогу швидко виокремити та передати дані за запитом суб'єкта;
- дедалі більша обізнаність суб'єктів про можливість неавторизованої обробки персональних даних Системами ШІ та відсутність технологічно ефективних способів перевірки цього факту суб'єктами може призводити до збільшення кількості запитів від них та, як наслідок, навантаження на розробників і затримки з наданням відповідей. Рішенням може бути автоматизація процесів обробки запитів за допомогою спеціалізованих інструментів та технічних рішень, наприклад, збір запитів чи надання доступу до даних за допомогою вебінтерфейсу управління даними на вебсайті розробника Системи ШІ. Хоча за своєю суттю цей процес не може бути автоматизовано повністю і він вимагає залучення професійного людського контролю, автоматизація може значно спростити процес отримання та категоризації запитів тощо.

6.6. Комунікація щодо обробки ПД шляхом використання ШІ

Обробка ПД суб'єктів даних із застосуванням ШІ може мати різні формати. Доведення інформації до суб'єктів даних щодо умов такої обробки може бути частиною маркетингової стратегії організації.

Багато компаній ідуть за межі простого дотримання регуляторних вимог і впроваджують найбільш ефективні рішення для захисту ПД суб'єктів даних та забезпечення реалізації їх прав, одночасно, активно комунікуючи такий підхід.

Це в багатьох випадках може збільшити рівень довіри споживачів до бренду та продукту загалом, наприклад, через формування впевненості, що система ШІ використовуватиме їх дані лише з указаною метою і вони мають ефективний контроль над своїми ПД.

Ефективна комунікація складних речей (наприклад, логіки використання ПД для тренування моделей) простими словами може одночасно демонструвати користувачам високий технічний рівень відповідної системи ШІ та бути частиною виконання вимог статті 5 GDPR щодо забезпечення прозорості в комунікації.

Комунікацію щодо дотримання вимог із захисту персональних даних у ШІ проєктах можна розділити на внутрішню і зовнішню. Внутрішня комунікація буде направлена на команду проєкту (ML-інженерів / датаінженерів тощо) з метою забезпечення високого рівня знань про забезпечення приватності в організації та підвищення культури приватності загалом. До інструментів такої комунікації можуть належати: навчання з тестуванням, роз'яснювальні матеріали, комунікація DPO з учасниками проєкту.

Дієвим лейтмотивом внутрішньої комунікації може бути донесення до команди проєкту важливості побудови довіри з кінцевими споживачами ШІ-рішення завдяки турботі про їхню приватність.

Зовнішня комунікація дотримання вимог із захисту персональних даних у ШІ проєктах є частиною виконання компанією вимог GDPR щодо принципу відповідності.

Засобами зовнішньої комунікації privacy compliance в ШІ проєктах можуть бути такі:

1. Належне виконання запевнень щодо забезпечення приватності в ШІ проєктах.

Наприклад, компанія може заявляти про можливість реалізовувати свої права суб'єктами даних та, зокрема, можливість відкликати згоду на обробку ПД для навчання ШІ в будь-який момент. Якщо це не буде реалізовано в простому та зручному функціоналі, довіра користувачів і їх користувацький досвід можуть постраждати, що вплине на бренд компанії і на її можливість використовувати ПД суб'єктів даних у своїх ШІ проєктах.

2. Залученість та проактивність спеціалістів з приватності / DPO компанії.

Наявність проактивної особи, що відповідає за дотримання вимог приватності ШІ проєктів компанії та її/його активна комунікація в соціальних мережах / онлайн- та офлайн-заходах щодо дотримання приватності в компанії може демонструвати відповідальну позицію організації щодо піклування про приватність користувачів

3. Своєчасні та ефективні повідомлення про наслідки дій користувача щодо приватності.

Оскільки користувач ШІ проєктів може надавати / змінювати умови використання власних ПД на різних умовах реалізації ШІ проєктів, у момент здійснення конкретних дій користувачем, наприклад – надання згоди / зміни умов використання його ПД, такому користувачу має бути зрозумілим, які наслідки це буде мати. Це може бути реалізовано шляхом коротких та вичерпних повідомлень перед відповідною операцією та наявністю посилань на деталізовані документи / політики тощо.

4. Деталізовані матеріали / відеогайди про практики privacy compliance.

Бізнес-клієнти можуть мати сумніви щодо можливості довірити ПД своїх кінцевих користувачів постачальникам ШІ-рішень. Деталізовані розділи на сайті (AI privacy FAQ) / спеціальні файли з детальними поясненнями щодо умов обробки таких ПД / відеоматеріали, які демонструють найкращі privacy-практики, що реально втілені в організації (шифрування / межі використання даних, окремі технічні заходи тощо) можуть допомогти залучити таких клієнтів.

6.7. Технічні та організаційні заходи захисту даних

Системи ШІ потребують забезпечення високого рівня захисту даних на всіх етапах їх обробки, зокрема за допомогою організаційних та технічних заходів. Законодавство України та GDPR вимагають наявності технічних й організаційних заходів, які мають враховувати права користувачів, ризики і вимоги до безпеки даних. Нижче наведено основні приклади заходів щодо захисту даних.

Закон України «Про захист персональних даних» та Типовий порядок обробки персональних даних:

- заходи, спрямовані на запобігання їх випадкових втрати або знищення, незаконної обробки, зокрема незаконного знищення чи доступу до персональних даних;
- наявність порядку доступу до персональних даних працівників;
- наявність порядку ведення обліку операцій, пов'язаних з обробкою персональних даних суб'єкта та доступом до них;
- розробка плану дій на випадок несанкціонованого доступу до персональних даних, пошкодження технічного обладнання, виникнення надзвичайних ситуацій;
- Регулярне навчання співробітників, які працюють з персональними даними;
- наявність письмових зобов'язань працівників про нерозголошення персональних даних, які їм було довірено або які стали їм відомі у зв'язку з виконанням професійних чи службових або трудових обов'язків.

GDPR:

- псевдонімізація та шифрування персональних даних;
- здатність забезпечити постійну конфіденційність, цілісність, доступність і стійкість систем обробки даних і послуг;
- можливість своєчасного відновлення доступності та доступу до персональних даних у разі фізичного чи технічного інциденту;
- процес регулярного тестування, оцінки та оцінки ефективності технічних та організаційних заходів для забезпечення безпеки обробки.

6.8. Технології, які посилюють приватність¹⁷ (PETs), як засіб захисту ПД при обробці ШІ

У контексті захисту інформації, технології, що посилюють приватність (PETs), є підгрупою технічних заходів захисту персональних даних¹⁸, спрямованих на додаткове підвищення рівня безпеки їхньої обробки. PETs втілюють фундаментальні принципи захисту даних шляхом:

- мінімізації обробки персональних даних;
- посилення інформаційної безпеки;
- розширення можливостей суб'єктів у контексті захисту персональних даних.

PETs можна поділити на такі категорії¹⁹:

- 1. Інструменти для приховування даних²⁰** покладаються на криптографічний захист інформації та дозволяють безпечну обробку даних завдяки приховуванню ідентифікувальних деталей, додавання «інформаційного шуму», заміну персональних даних синтетичною інформацією тощо. Інструменти мають такий ефект:
 - **анонімізація персональних даних** дає змогу повністю видаляти ідентифікувальні елементи з персональних даних, що теоретично робить повторну ідентифікацію суб'єкта даних неможливою;
 - **псевдонімізація персональних даних**, подібно до анонімізації, дає змогу видалити чи замінити ідентифікувальні елементи персональних даних, щоби зменшити ризик повторної ідентифікації суб'єкта даних. На відміну від анонімізації, цей інструмент дозволяє зворотне відновлення інформації;
 - **генерування наборів синтетичних даних** на основі таких самих або схожих статистичних властивостей, як і в оригінальних персональних даних, дає змогу частково чи повністю уникнути використання реальних персональних даних. Залежно від цілей використання такі набори можуть містити повністю синтетичні (набір не містить оригінальних персональних даних), частково синтетичні (у наборі даних відбувається заміна синтетичними найбільш чутливих ідентифікувальних елементів) чи гібридні дані (створюється «двійник» набору даних на основі синтетичної інформації, який потім змішується з оригінальним набором);
 - **технологія диференційної приватності** дає змогу знизити або ліквідувати ідентифікаційну властивість числового масиву інформації за допомогою введення контрольованої кількості випадкового інформаційного шуму в дані за умови збереження статистичних характеристик набору. Під шумом зазвичай мають на увазі випадкові числові значення або результати обчислень за конкретними формулами;
 - **технологія доведення з нульовим розголошенням** дозволяє опосередковано отримати підтвердження певної властивості набору даних без розкриття таких даних.
- 2. Інструменти для обробки зашифрованих даних²¹** дають змогу обробляти персональні дані, які не будуть видимими або розкритими завдяки захищеності криптографічними методами. На відміну від інструментів приховування даних, шифрування не передбачає їх зміну. Виокремлюють такі інструменти:
 - **гомоморфне шифрування**. За загальним правилом, за відсутності відповідного ключа зашифровані дані є непридатними для аналізу чи обробки за призначенням. За допомогою технології гомоморфного шифрування є можливість залишати дані зашифрованими під час їхнього аналізу та використання для обчислень, отримуючи результат, який буде зашифровано тим самим способом, як і оригінальні дані;
 - **багатостороннє обчислення**. Технологія дає змогу кільком учасникам обробки даних (або кільком різним наборам даних у межах одного учасника) обмінюватися інформацією для обчислювальних завдань без розкриття даних одне одному, але з доступом до кінцевого результату аналізу чи обчислень;
 - **довірене середовище виконання**. За допомогою цього інструменту можливо обробляти, зокрема, аналізувати дані в спеціальному виокремленому середовищі обчислювального процесора обчислювальної системи, недоступному для решти системи. Це гарантує, що дані будуть оброблені лише для певної операції та не використовуватимуться надмірно й не зберігатимуться в просторі, до якого можуть отримати доступ сторонні особи, комп'ютерні програми чи обладнання.

¹⁷ Англ., privacy enhancing technologies;

¹⁸ До таких заходів захисту належать, наприклад, шифрування інформації при зберіганні та передачі, мінімізація, логування та контроль доступу до даних, відокремлене зберігання різних типів інформації та створення резервних копій тощо.

¹⁹ На дату оприлюднення цих рекомендацій універсальної класифікації не існує. У цьому розділі за основу взято поділ, запропонований OECD: https://www.oecd-ilibrary.org/science-and-technology/emerging-privacy-enhancing-technologies_bf121be4-en

²⁰ Англ., Data obfuscation tools: Anonymisation/Pseudonymisation, Synthetic data, Differential privacy, and Zero-knowledge proofs;

²¹ Англ., Encrypted data processing tools: Homomorphic encryption, Multi-party computation, and Trusted execution environments

3. Інструменти об'єднаної та розподіленої аналітики²² дають змогу отримувати аналітичний результат з даних, які не передаються та не розкриваються тому, хто отримує результат їх обробки. Описують такі технології:

- **об'єднаному навчанню** характерний підхід, коли необроблені дані попередньо обробляються на рівні джерела даних (наприклад, на рівні кожного суб'єкта даних або кожного окремого набору даних на рівні володільця (контролера) персональних даних). Лише зведені статистичні дані від усіх суб'єктів / наборів даних та результати їх аналізу передаються в централізовану базу даних для подальшого об'єднання з аналогічними даними з інших джерел;
- **розподіленій аналітиці**, на відміну від попереднього інструменту, притаманне збереження оригінальної інформації в центральній захищеній базі даних та надання доступу до обмеженого обсягу зашифрованої інформації тим, кому він необхідний для обробки, зокрема, аналізу. Особливістю цього інструменту є те, що аналіз різних фрагментів інформації зі зниженою ідентифікаційною властивістю відбувається багатьма гравцями паралельно, результати аналізу акумулюються централізовано і потім повідомляються кожному, хто здійснював аналіз, без необхідності розкриття повного набору даних.

4. Інструменти підзвітності даних²³ пропонують засоби контролю над тим, як персональні дані можуть збиратися та оброблятися. Серед них виокремлюють:

- **підзвітні системи.** Такі системи дають змогу контролювати та відстежувати, як і коли дані можуть збиратися, оброблятися й використовуватися, шляхом прив'язки метаданих, які обмежують сценарії взаємодії з такими даними та дають змогу відстежувати історію дозволеної взаємодії;
- **порогове розподілення таємної інформації.** Цей інструмент дає змогу встановити особливе правило доступу та можливості подальшої обробки зашифрованих даних. Таким правилом є умовно одночасне використання ключів шифрування всіма тримачами та ухвалення спільного рішення про доступ до даних;
- **системи зберігання та управління персональними даними.** Ця технологія дає змогу суб'єктам персональних даних вільно обирати, де і як їхні дані зберігатимуться, у кого та у якому обсязі до них буде доступ та як вони оброблятимуться. Такі системи є місцем збереження персональних даних про конкретного суб'єкта, який повністю контролює (як логічно, так і фізично) обробку його даних.

Задля реалізації принципу мінімізації обробки персональних даних, забезпечення високого рівня безпечності обробки та зниження ризиків витоку персональних даних, розробникам Систем ШІ рекомендується впровадити інструменти PETs при проектуванні, тренуванні та тестуванні Систем ШІ, якщо це передбачає обробку персональних даних. Набір інструментів PETs, які можуть використовувати розробники, відрізнятиметься залежно від специфіки системи ШІ та сфери її подальшого застосування. Пропонуються такі приклади реалізації описаних інструментів:

- за допомогою генерування та використання синтетичних наборів даних чи використання інструментів диференційної приватності розробники можуть тренувати системи ШІ без використання наборів реальних персональних даних, але зберігаючи їхні статистичні та кореляційні властивості. Наприклад, банки (або розробники, залучені банками чи інші фінансові установи) можуть використовувати штучні дані, створені на базі історичної інформації їхніх клієнтів, для тренування систем ШІ, спрямованих на оцінку ризикованості операцій та виявлення ознак шахрайства на ранніх стадіях;
- з використанням інструментів об'єднаного навчання при тренуванні та використанні систем ШІ розробники й користувачі зможуть поєднувати потужності різних систем ШІ чи їхніх «джерел знань» без фактичної передачі оригінальних даних з них. Наприклад, дослідницькі центри при лікарнях чи спеціалізованих інститутах можуть співпрацювати для створення моделі прогнозування та виявлення захворювань, зберігаючи дані пацієнтів локально. У централізовану систему передаватимуться лише агреговані статистичні результати, інформація про тенденції та результати обчислень;
- за допомогою інструментів підзвітності розробники можуть відстежувати, як персональні дані використовують для тренування систем ШІ, і контролювати доступ до них. Наприклад, при розробці та тренуванні систем розпізнавання мовлення й віртуальних асистентів розробники можуть розграфити набори даних та додати відповідні метадані до аудіозаписів, які лімітуватимуть аналіз таких даних лише за конкретними параметрами задля уникнення створення упереджень та хибних тенденцій, які потенційно можуть виникнути при визначенні статі, емоційного стану чи расової належності спікера на них.

²² Англ., Federated and distributed analytics: Federated learning and Distributed analytics

²³ Англ., Data accountability tools: Accountable systems, Threshold secret sharing, and Personal data stores/Personal Information Management Systems

6.9. Хороші Vs погані практики при обробці ПД з використанням ШІ

Нижче наведено приклади різних практик приватності при обробці персональних даних з використанням ШІ. Ці приклади вигадані, а коментарі – ілюстративні та не є юридичною порадою. Збіги з реальними бізнес-процесами є випадковими.

Медична приватність та ШІ-помічник

Приклад 1

Фабула:

інноваційна медтех-компанія «МедТехКомпанія1» використала частково анонімізовані набори даних з персональними даними клієнтів (без перевірки віку) партнерських медичних закладів, а саме було вказано стать, вік, історії хвороб та їх лікувань з метою тренування власної Моделі, яка планується для використання вказаними партнерськими закладами як ШІ-помічник для лікарів при діагностуванні / лікуванні пацієнтів. Пацієнти не давали своєї згоди на відповідну обробку / передачу даних і датасет не був достатньо диверсифікованим і в результаті Модель генерувала явно шкідливі / сумнівні поради.

Коментар:

Дії компанії «МедТехКомпанія1» є неправомірною практикою використання ПД для навчання Моделі. Такі дії мають явні ознаки порушення законодавства про захист персональних даних. Відсутність правових підстав для обробки, ризик деанонімізації, ризик незаконної обробки ПД дітей, ризик дискримінації та інші питання мають місце. Украв імовірно, що найкращим рішенням буде негайне знищення відповідної Моделі та відповідного набору даних, що був спеціально скомпільований (упорядкований).

Кастомізована реклама в соціальній мережі

Приклад 2

Фабула:

Соціальна мережа для публікації коротких відео та фото «СоцМережа1» використовує Модель1 з метою аналізу контенту, що завантажений користувачами на предмет локації та часу зйомки, зіставляє ці дані з профілем користувача (включно з його активними діями на платформі) та враховує цю інформацію про демонстрації таргетингової реклами третіх осіб таким користувачам, використовуючи при цьому Модель2. Модель1 постійно тренується на нових фото та відео, що завантажують користувачі, а Модель2 опрацьовує не зашифровані та неанонімізовані персональні дані, здійснюючи кореляцію та «покращуючи» результати, з огляду на інформацію про активні кліки користувачів на рекламні оголошення. Умова про використання Моделей була додана в Угоду користувача та Політику приватності компанії маленькими абзацами на 8 та 10 сторінках відповідних документів на 30-й день від початку використання вказаних Моделей.

Коментар:

Указана фабула ілюструє цілий комплекс можливих порушень та поганих практик. Комунікація щодо використання Моделей є непрозорою, невчасною, правова підстава для обробки даних є вкрай сумнівною і в користувача немає реального впливу та контролю як обробкою його / її персональних даних Моделлю1, так і Моделлю2. Також мають місце підвищені ризики розкриття персональних даних у результаті недостатніх технічних заходів захисту, дискримінації та інших порушень.

**Модель, що
відстежує
шахрайські
транзакції**

Приклад 3

Фабула:

Банк «ДужеІнноваційнийБанк1» впровадив у свої бізнес-процеси Модель для виявлення можливих шахрайських операцій у межах транзакцій своїх клієнтів. Зокрема, Модель була натренована на анонімізованих, правомірно отриманих наборах даних виявляти ризикові транзакції на основі зіставлення даних про конкретну операцію, водночас отримуючи лише мінімально необхідний доступ до даних користувача в момент аналізу. Банк провів Data protection impact assessment (оцінювання про захист даних) перед інтеграцією Моделі в систему, а також проводить системний аналіз її роботи. Можливість використання ШІ-систем для запобігання шахрайству передбачається договором та повідомленням про захист персональних даних, що укладається з клієнтами банку.

Коментар:

Описані privacy-практики відображають відповідальний підхід до використання / обробки персональних даних системами ШІ в реальних бізнес-процесах. До того ж необхідно враховувати, що вказані практики мають бути частиною комплексної програми приватності, у межах якої забезпечуються належні організаційні та технічні заходи для дотримання відповідного законодавства й захисту прав суб'єктів персональних даних.

Розділ 7. Антимонопольні ризики

Оскільки цифрові послуги легко масштабуються, при розробці системи ШІ виникає низка ризиків порушити антимонопольне законодавство та законодавство про захист добросовісної конкуренції. Вказані питання в Україні регулюються насамперед Законами **«Про захист від недобросовісної конкуренції»** та **«Про захист економічної конкуренції»**. Вказані закони, як і **звіт** Антимонопольного комітету України (далі — АМКУ) за 2023 рік, наразі не містять згадок про виявлені порушення саме у сфері розробки систем ШІ. Проте з огляду на норми вказаних законів та напрями розвитку правозастосування у світі, зонами ризику є такі:

1. Зловживання домінантним становищем на ринку

Кожна кафедра або програма аналізує специфіку своїх курсів і визначає відповідну позицію щодо використання генеративного ШІ. Позиція може бути обраною на власний розсуд.

Приблизний перелік позицій передбачає таке:

Цей ризик виникає в розробників систем ШІ, якщо вони:

- контролюють великий сегмент ринку (наприклад, ринки генеративного ШІ, хмарних обчислень чи аналітики даних);
- упроваджують практики, які ускладнюють доступ конкурентів до інфраструктури чи даних, необхідних для розробки ШІ (наприклад, монополія на навчальні набори даних або обчислювальні ресурси);
- вимагають від користувачів ексклюзивного використання їхніх платформ, обмежуючи можливість співпраці з іншими постачальниками.

Варто підкреслити, що станом на сьогодні відсутні методики визначення окремих ринків у сфері систем ШІ, тому застосовується **Методика визначення монопольного (домінуючого) становища суб'єктів господарювання на ринку**, затверджена розпорядженням Антимонопольного комітету України від 05 березня 2002 року № 49-р.

2. Антиконтурентні узгоджені дії

Угоди між компаніями, що працюють у сфері ШІ, можуть порушувати антимонопольне законодавство, якщо вони:

- містять положення про фіксацію цін, розподіл ринків чи обмеження обсягів виробництва;
- передбачають обмін комерційно чутливою інформацією між конкурентами.

Детальний аналіз можливих антимонопольних ризиків наведено в контексті зловживання домінантним становищем на ринку та антиконкурентних узгоджених дій щодо законодавства ЄС і США у звіті **ОЕСД**.

Варто звернути увагу, що в Україні діє **Положення про порядок подання заяв до органів Антимонопольного комітету України про надання дозволу на узгоджені дії суб'єктів господарювання**, яке застосовується як до українських, так і до іноземних компаній та транснаціональних корпорацій.

3. Концентрації

Угоди про придбання, злиття, набуття контролю над іншими компаніями у сфері ШІ може потребувати дозволу або попереднього висновку АМКУ відповідно до **Положення про порядок розгляду заяв та справ про концентрацію суб'єктів господарювання**, яке також застосовується як до українських, так і до іноземних компаній та транснаціональних корпорацій.

Вказані угоди можуть бути заблоковані, зокрема, якщо:

- угода значно зменшує конкуренцію на ринку;
- є ризики «поглинання конкурента» (killer acquisitions), коли більші компанії купують новаторів, щоб обмежити їх вплив.

4. Дискримінація, дискредитація та привілейоване ставлення

Розробники можуть порушити антимонопольне законодавство, якщо вони:

- віддають перевагу окремим постачальникам або користувачам (наприклад, через ціноутворення чи доступ до функцій);
- дискредитують певні групи постачальників чи користувачів;
- дискримінують певні групи постачальників чи користувачів;
- використовують власні платформи для непропорційного просування своїх продуктів, порушуючи принципи рівного доступу;
- створюють інші обмеження для конкуренції (наприклад, завищення цін через автоматизацію картелів).

З метою мінімізації вказаних ризиків рекомендується вживати таких заходів, як:

1. Розробка прозорих алгоритмів:

- розробляти системи ШІ з урахуванням принципів справедливої конкуренції;
- забезпечити відкритість та прозорість алгоритмів, особливо щодо рішень, які впливають на ціноутворення або доступ до ринку;
- уникати створення алгоритмів, що сприяють антиконкурентній координації між учасниками ринку;
- використовувати анонімізацію та агрегування даних, щоб уникнути звинувачень у маніпуляції ринком через асиметрію інформації.

2. Дотримання принципів недискримінації:

- гарантувати рівні умови доступу всім користувачам та партнерам до даних і функціональності системи ШІ;
- забезпечувати, щоб дані, які використовують для навчання моделей ШІ, відповідали принципам добросовісності та не створювали бар'єрів для конкуренції;
- забезпечувати нейтральність платформ у випадку, якщо розробнику належить інфраструктура, яка є критичною для його конкурентів;
- залучати незалежних експертів для аналізу розроблених систем ШІ щодо упередженості та дискримінації.

3. Антиконтурентний комплаєнс:

- упровадити програми з дотримання антимонопольного законодавства, включно з навчанням персоналу;
- перевіряти умови угод з постачальниками та споживачами щодо їх відповідності антимонопольному законодавству.

4. Взаємодія з регуляторами:

- перед укладанням угод про узгоджені дії або концентрацію консультиватися з профільними юристами щодо можливих ризиків та необхідності отримання дозволів або попередніх висновків від АМКУ, антимонопольних органів ЄС та інших країн (у випадку, якщо розробник представлений більше, ніж в одній країні, або планує укласти вказані угоди з резидентом іншої країни);

- застосовувати екосистемний підхід та створити механізми для оцінки впливу розроблюваних систем ШІ на конкуренцію, а також чи вони сприяють або обмежують інновації на ринку за всім ринковим ланцюжком, включно з постачальниками, партнерами і кінцевими споживачами;
- бути готовими до перевірок з боку антимонопольних органів, зберігаючи документацію щодо розробки, впровадження та використання ШІ-систем;
- у разі надходження запитів від АМКУ або антимонопольних органів ЄС та інших країн, надавати їм необхідну інформацію для перевірки, оскільки штрафи за ненадання інформації є суттєвими.

Варто підкреслити, що штрафи за порушення антимонопольного законодавства як в Україні, так і в іноземних країнах є надзвичайно високими й часто вимірюються у відсотках від доходу компанії за рік, а відповідні процеси розгляду є тривалими й такими, що потребують значних ресурсів. Тому вкрай важливо приділяти увагу постійному моніторингу антимонопольних ризиків та вживати заходів щодо їх мінімізації. Наприклад, у ЄС почав діяти **Digital Market Act** (DMA), який запроваджує поняття та критерії gatekeepers («брамники»), установлює їх обов'язки, а також суттєві санкції за порушення.

Тому дотримання наведених рекомендацій дасть змогу розробникам систем ШІ не лише мінімізувати ризики у сфері антимонопольного права, а й підвищити довіру споживачів та регуляторів до їхніх продуктів, забезпечуючи сталий розвиток ринку.

Додаток Інструменти

Як обирати інструменти

Використання інструментів для розробки штучного інтелекту, які гарантують дотримання прав людини, є важливим кроком у розробці етичних та відповідальних систем ШІ. Вибір інструментів передбачає баланс між технічними можливостями й етичними міркуваннями. Ухвалюючи рішення, зверніть увагу на такі рекомендації:

- обирайте інструменти, які зосереджені на розробці ШІ, орієнтованого на людину, тобто вони спроектовані з урахуванням прав і добробуту користувачів. Деякі платформи пропонують функції, що підтримують етичний дизайн ШІ, включно з інструментами для моніторингу та оцінки впливу систем ШІ на людину;
- вибирайте платформи розробки AI, які забезпечують прозорість своїх алгоритмів, даючи змогу розробникам і користувачам зрозуміти, як ухвалюють рішення;
- використовуйте інструменти, які мають функції для виявлення та зменшення упередженості в наборах даних і моделях;
- переконайтеся, що інструменти, які ви використовуєте, відповідають вимогам щодо захисту даних і мають потужні функції конфіденційності. Інструменти повинні дозволяти анонімізацію, мінімізацію даних та безпечне зберігання даних для захисту прав на приватність;
- використовуйте платформи, які сприяють залученню зацікавлених сторін та збору зворотного зв'язку, забезпечуючи залучення в процес розробки ШІ інтересів та потреб тих, на кого може вплинути ця система;
- обирайте інструменти від постачальників, які мають чітку відданість правам людини та етичним практикам у сфері ШІ. Перевірте їхні політики, досвід та участь в ініціативах щодо етичного ШІ.

Список популярних інструментів



Aequitas Bias and Fairness Audit Toolkit

набір інструментів з відкритим кодом для аудиту упередженості та справедливості моделей машинного навчання, розроблений Університетом Чикаго.



AI Fairness 360

AI Fairness 360 від IBM

набір інструментів з відкритим кодом для забезпечення справедливості штучного інтелекту, призначений для виявлення та усунення упередженостей у моделях.



Anthropic Claude

генеративна модель штучного інтелекту, використовується для генерації тексту та обробки природної мови.



Apache Mahout

масштабована бібліотека машинного навчання з відкритим кодом, призначена для створення алгоритмів рекомендацій, кластеризації та класифікації.



AWS SageMaker Clarify

інструмент для виявлення упередженостей і забезпечення пояснюваності штучного інтелекту, використовується для аналізу моделей на наявність упередженостей.



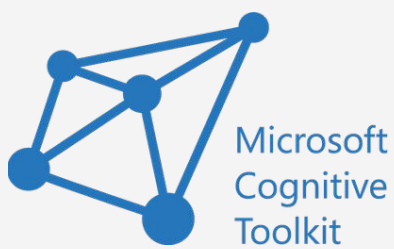
Azure AI

хмарна платформа для штучного інтелекту від Microsoft, що надає сервіси для розробки, тренування та розгортання моделей машинного навчання.



Caffe

фреймворк глибокого навчання з відкритим кодом, розроблений в Університеті Каліфорнії (Берклі), оптимізований для швидкості та модульності.



CNTK (Microsoft Cognitive Toolkit)

фреймворк глибокого навчання з відкритим кодом, використовується для створення нейронних мереж.



Cursor

редактор програмного коду, який використовує технології машинного навчання для надання розумних підказок, автодоповнення та виявлення помилок у коді.



Deeplearning4j

фреймворк глибокого навчання для Java та Scala з підтримкою розподілених обчислень.



Fairlearn від Microsoft

набір інструментів з відкритим кодом для забезпечення справедливості штучного інтелекту, використовується для оцінки та зменшення упередженостей у моделях.



Fast.ai

високорівнева бібліотека для глибокого навчання на основі PyTorch, спрямована на спрощення розробки моделей машинного навчання.



GitHub Copilot

інструмент на основі штучного інтелекту, розроблений GitHub у співпраці з OpenAI. Інтегрується в середовища розробки коду, як-от Visual Studio Code, і надає рекомендації та автодоповнення під час написання програмного коду.



Google Gemini

генеративна модель штучного інтелекту, використовується для автоматизації завдань та створення контенту.



Vertex AI

Google Vertex AI

хмарна платформа для машинного навчання та генеративного штучного інтелекту, надає інструменти для розробки й розгортання моделей.



AI Explainability 360

Google's AI Explainability 360

набір інструментів для забезпечення пояснюваності штучного інтелекту, використовується для аналізу та пояснення рішень моделей.

H2O.ai

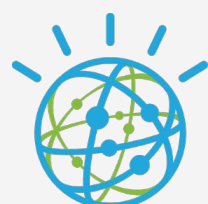
H2O.ai

платформа з відкритим кодом для машинного навчання, використовується для створення моделей аналізу даних і прогнозування.



Hugging Face Transformers

бібліотека з відкритим кодом для обробки природної мови (NLP), надає попередньо навчені моделі та інструменти для роботи з текстом.



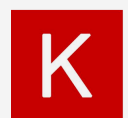
IBM Watson

платформа штучного інтелекту, що надає сервіси для обробки природної мови, машинного навчання та аналізу даних.



JAX

бібліотека з відкритим кодом для чисельних обчислень від Google Research, використовується для розробки алгоритмів глибокого навчання та оптимізації моделей.



Keras

Keras

високорівнева бібліотека для нейронних мереж, що працює з TensorFlow, використовується для створення моделей машинного навчання.



LightGBM

LightGBM

високопродуктивний градієнтний бустинг фреймворк від Microsoft, оптимізований для швидкості та продуктивності.



Microsoft Copilot

інструмент на основі штучного інтелекту, який працює на базі пошукової системи Bing та великих мовних моделей від OpenAI, що використовується для генерування відповідей, пошуку інформації та аналізу даних.



Midjourney

генеративна модель штучного інтелекту, яка перетворює текстові описи на зображення. Використовується для створення художніх та креативних візуалізацій, надаючи можливість генерувати зображення на основі заданих текстових запитів.



MXNet

платформа з відкритим кодом для глибинного навчання від Apache Software Foundation, підтримує різні мови програмування.



ONNX Runtime

інструмент з відкритим кодом для виконання моделей машинного навчання у форматі ONNX, підтримує різні апаратні платформи.



OpenAI ChatGPT

генеративна модель для обробки природної мови від OpenAI, застосовується для генерації тексту та відповіді на запитання.



Perplexity

вебсервіс від Perplexity AI, використовується для пошуку та обробки інформації.



PyTorch

фреймворк з відкритим кодом для машинного навчання від Meta, використовується для розробки та тренування моделей.



RapidMiner

платформа для аналізу даних та машинного навчання, надає інструменти для побудови моделей без необхідності програмування.



Scikit-learn

бібліотека Python для машинного навчання, використовується для класифікації, регресії та кластеризації.



TensorFlow

платформа з відкритим кодом для машинного навчання від Google, застосовується для створення та тренування моделей.



Theano

бібліотека Python з відкритим кодом для чисельних обчислень, дає змогу ефективно визначати та оптимізувати математичні вирази.