

ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ ТА БЕЗПЕКА ДАНИХ

Лекція 7

Дані як актив і ризик: типи даних, персональні дані за законом



Освітня програма «Економіка і аналітика» · 1 курс · 4 кредити ECTS

Місток від технологій до безпеки

- Лекції 1–6 показали, що таке ШІ, як він працює, де помиляється і як його застосовують в економіці
- Кожне застосування — від чат-бота підтримки до фінансового скорингу (лекція 6) — працює з даними: клієнтів, співробітників, комерційною інформацією
- **Друга половина курсу:**
 - як безпечно й відповідально працювати з цими даними — організаційно та юридично

Ця лекція закладає фундамент: класифікація даних, триада безпеки, персональні дані, основи законодавства.

Дані як економічний актив

У сучасній економіці дані дедалі частіше називають «ною нафтою» — ресурсом, що набуває цінності через переробку.

- На відміну від фізичних активів, дані можна копіювати необмежену кількість разів практично без витрат
- Цінність часто зростає з обсягом — мережевий ефект: більше даних дають точніші прогнози й персоналізацію
- Дата-брокери купують і продають споживчі профілі; технологічні компанії оцінюються значною мірою через якість зібраних даних
- Для тренування сучасних ШІ-моделей (пор. лекцію 2) якісні дані стали дефіцитним і дорогим ресурсом

Дані варто розглядати як актив на балансі уваги компанії, навіть якщо бухгалтерськи вони рідко відображаються як такі.

Тріада інформаційної безпеки: CIA

Конфіденційність

Confidentiality

Дані доступні лише тим, хто має на це право

Протилежність: витік, несанкціонований доступ

Цілісність

Integrity

Дані точні, повні й не змінені без дозволу

Протилежність: підробка, зловмисне редагування

Доступність

Availability

Дані й системи доступні тоді, коли потрібно

Протилежність: збої, DDoS-атаки, вимагачі

Ці три властивості часто конфліктують: надмірна конфіденційність (складні паролі, 2FA) може знижувати доступність. Баланс — управлінське рішення.

Чотири рівні чутливості даних

Публічні

Прес-релізи, публічна звітність, маркетингові матеріали

Наслідки витоку: відсутні

Внутрішні

Внутрішні процедури, робоче листування

Наслідки витоку: помірні

Конфіденційні / комерційна таємниця

Фінансові показники, стратегічні плани, клієнтські бази, вихідний код

Наслідки витоку: серйозні

Персональні дані

Будь-яка інформація про фізичну особу

Наслідки витоку: юридичні (найсуворіші)

Перш ніж вводити інформацію в загальнодоступний ШІ-інструмент (лекція 8) — подумки класифікуйте її за цією шкалою.

Що таке персональні дані?

Будь-яка інформація, що відноситься до фізичної особи, яку прямо чи опосередковано ідентифіковано або можна ідентифікувати.

Прямі ідентифікатори

Ім'я, номер паспорта чи ІПН, номер телефону, домашня адреса

Непрямі ідентифікатори

Дата народження + індекс, IP-адреса, фото обличчя, геолокація, ID пристрою (у комбінації з іншими даними)

Псевдонімізація (заміна імені кодом) — це НЕ анонімізація: дані лишаються персональними, доки існує ключ депсевдонімізації.

Особливі категорії персональних даних

Здоров'я

Біометрія

Генетичні дані

Расове/етнічне
походження

Політичні погляди

Релігійні переконання

Членство у профспілках

Сексуальне життя чи
орієнтація

- Обробка таких даних вимагає додаткових підстав (явна згода, законодавчі виключення) і посилених заходів захисту

Приклад ризику для теми курсу: якщо студент вводить медичну довідку в загальнодоступний чат-бот «для перевірки формулювань» — це вже може бути передачею особливої категорії даних третій стороні (лекція 8).

Правове регулювання в Україні

Закон України «Про захист персональних даних» №2297-VI (2010, з подальшими змінами)

- Визначає персональні дані, встановлює права суб'єктів даних
- Встановлює обов'язки тих, хто обробляє дані (володільців і розпорядників)
- Визначає повноваження органу контролю
- Україна рухається в напрямку гармонізації з європейським GDPR у межах євроінтеграційного курсу

Тривають законодавчі процеси з оновлення українського законодавства для більшого узгодження з GDPR — перевіряйте актуальний стан окремо для конкретної практичної ситуації.

Основні принципи обробки персональних даних

- 1 Законність і прозорість
- 2 Обмеження мети
- 3 Мінімізація даних
- 4 Точність
- 5 Обмеження зберігання
- 6 Цілісність і конфіденційність
- 7 Підзвітність

Ролі: володільць, розпорядник, суб'єкт даних



Коли компанія вводить дані клієнтів у сторонній ШІ-інструмент, вона не перестає бути володільцем і залишається відповідальною за їх захист.

Права суб'єкта персональних даних

Доступ

Знати, чи обробляються дані, отримати копію

Виправлення

Вимагати виправлення неточних даних

Видалення

«Право бути забутим» за певних умов

Заперечення

Проти певних видів обробки (напр. маркетинг)

Обмеження обробки

Тимчасово «заморозити» обробку у спірних ситуаціях

Портативність

Отримати дані у зручному форматі для передачі

Ці права стосуються й вас як споживачів цифрових сервісів, включно з ШІ-інструментами.

Дані як ризик: чому витік дорого коштує

Юридичний

Штрафи регуляторів, судові позови, обов'язок публічно повідомити про витік

Фінансовий

Розслідування інциденту, усунення вразливості, компенсації, страхові премії

Репутаційний

Втрата довіри клієнтів і партнерів — довші й важче вимірювані наслідки

Ці ризики стосуються не лише великих корпорацій — будь-яка організація, що обробляє персональні дані, потенційно під ризиком.

Чому тема даних стала критичною саме зараз

- **Навчальні дані (лекція 2, RLHF):**
 - моделі тренуються на величезних масивах тексту з інтернету й діалогів — чи потрапляють туди персональні дані, залишається предметом судових суперечок
- **Введені дані (тема лекції 8):**
 - кожен запит до чат-бота потенційно містить чутливу інформацію, яка залишає межі організації
- **Упередженість (лекція 4):**
 - дефіцит якісних, етично зібраних даних підвищує спокусу використовувати дані сумнівного походження

Захист даних — не окрема від теми ШІ дисципліна, а необхідна умова відповідального використання ШІ.

Чек-лист класифікації даних

- 1 Чи стосується інформація конкретної фізичної особи, яку можна ідентифікувати?
- 2 Якщо так — чи належить вона до особливої категорії (здоров'я, біометрія тощо)?
- 3 Чи є законна підстава й дозвіл використовувати ці дані саме так?
- 4 Чи є це комерційною таємницею, навіть якщо не персональні дані?
- 5 Якби це стало публічним завтра — які були б наслідки?

Сумнів чи занепокоєння з будь-якого пункту — сигнал зупинитися і проконсультуватися, перш ніж ділитися інформацією з третіми сторонами.

ПІДСУМОВУЄМО

Що варто запам'ятати з цієї лекції

- Дані — водночас цінний економічний актив і джерело серйозних юридичних, фінансових і репутаційних ризиків
- Тріада CIA (конфіденційність, цілісність, доступність) — базова мова опису захисту даних
- Дані класифікуються за чутливістю: публічні → внутрішні → конфіденційні → персональні
- Персональні дані — будь-яка інформація про ідентифіковану чи ідентифіковану особу; особливі категорії вимагають посиленого захисту
- В Україні діє Закон «Про захист персональних даних», з рухом у бік гармонізації з GDPR
- Кожен, хто працює з ШІ, водночас лишається володільцем даних, які вводить у систему

Наступна лекція: приватність у діалозі з ШІ — куди насправді йдуть введені дані.