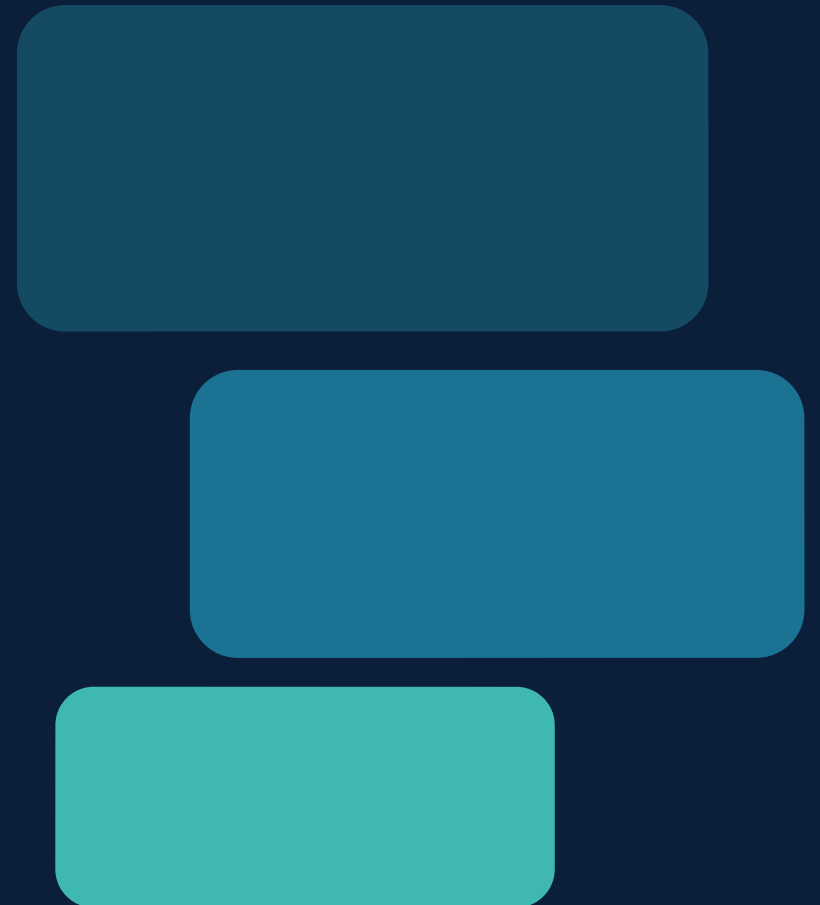


ТЕХНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ ТА БЕЗПЕКА ДАНИХ

Лекція 2

# Як ми навчили машину говорити: можливості генеративного ШІ

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



# Класичний підхід до обробки мови — і його провал

## 1950–1980-ті: підхід Хомскі

Мову намагалися кодувати строгими математичними правилами та лінгвістичними деревами.

Ідея спиралась на ранню філософію мови (Вітгенштайн): мова — «дзеркало» реальності з чіткою логічною структурою.

## Реальність: жива мова

Хаотична, повна винятків, метафор і контекстних значень.

Жорсткі правила не витримували зіткнення зі справжнім людським мовленням — комп'ютери «ламались» об виключення.

*Перші ~30 років спроб «зрозуміти» мову зазнали поразки саме через намагання формалізувати те, що погано піддається формалізації.*

# Нечіткість логіки змушує ШІ імітувати людину

## Каса самообслуговування

Скануй штрихкод → додай у чек.

Жорсткий алгоритм справджується, бо задача чітко формалізована: один товар — один код — одна дія.

## Базар / ринок

Торг, невербальні сигнали, контекст стосунків продавця й покупця.

Це неможливо звести до жорсткого алгоритму — забагато змінних і винятків.

*Людське спілкування набагато ближче до базару, ніж до каси самообслуговування — тому сучасний ШІ навчився імітувати реакцію людини, а не «рахувати» мову.*

# Чому бізнес нарешті «загорівся» генеративним ШІ

## Було: інструмент

Генератори тексту існували й раніше, але залишались нішевою цікавинкою — «приблудою для копірайтерів».

Обмежений інтерес бізнесу, обмежені інвестиції.

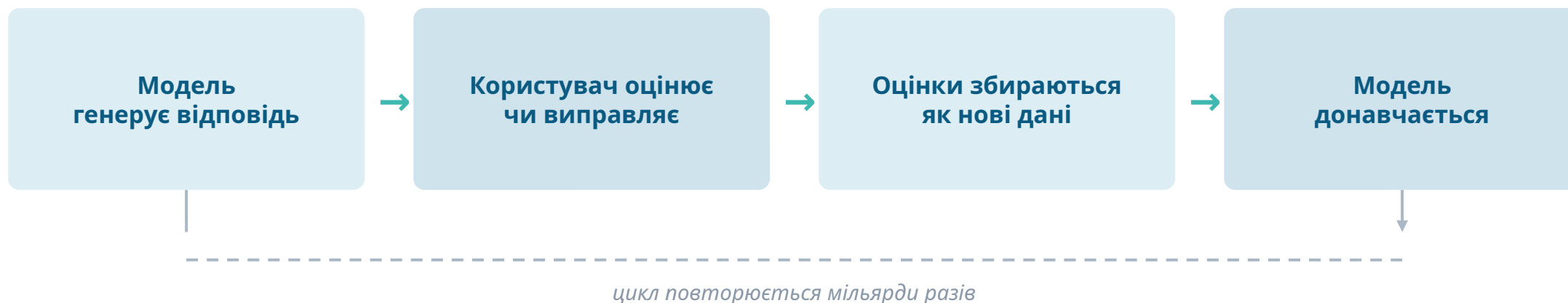
## Стало: співрозмовник

Користувачі почали сприймати LLM не як «папу з феноменальною пам'яттю», а як живого співрозмовника.

Саме ця зміна сприйняття відкрила двері для інвестицій у величезних масштабах.

*Урок для економістів: ринковий успіх технології іноді визначається не лише її можливостями, а тим, як її сприймає користувач.*

# RLHF: «ми навчили ШІ думати, хоча нас про це не просили»



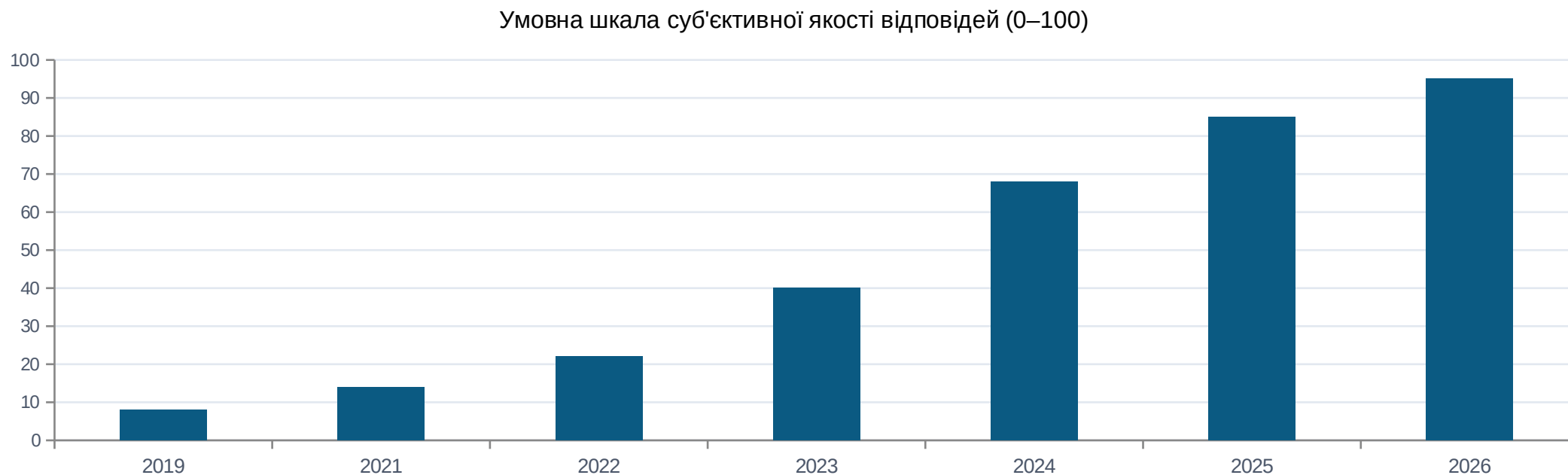
- До статичних текстів з інтернету додався новий тип даних — мільярди діалогів з користувачами
- Reinforcement Learning from Human Feedback (RLHF) — навчання з підкріпленням на основі людських оцінок
- Кожен користувач ChatGPT / Claude / Gemini, сам того не усвідомлюючи, ставав «вчителем» для наступної версії моделі

# Ми й раніше ненавмисно тренували ШІ

| Механізм        | Що робили люди                               | Що це тренувало                             |
|-----------------|----------------------------------------------|---------------------------------------------|
| CAPTCHA         | Розпізнавали розмитий текст на картинці      | Оцифрування книг і карт                     |
| reCAPTCHA       | Обирали «всі квадрати зі світлофорами»       | Розпізнавання об'єктів для безпілотних авто |
| Alt-текст       | Описували зображення для людей з вадами зору | Дані для генераторів зображень              |
| RLHF (сьогодні) | Оцінюють / виправляють відповіді LLM         | Якість генеративних мовних моделей          |

*Різниця лише одна: раніше нас не питали прямо, а тепер (RLHF) — прямо запитують «яка з двох відповідей краща?».*

## Вибухова еволюція якості моделей за лічені роки



*Це нетипова для технологій крива: зазвичай прогрес уповільнюється з часом, а тут — прискорюється завдяки поєднанню обчислень + даних всього інтернету + мільярдів людських оцінок (RLHF).*

# Тест Тьюрінга більше не актуальний



*Суддя листується текстом з A і B та намагається визначити, хто з них людина*

- Класичний тест Тьюрінга (1950) перевіряв, чи машина, будуючи мову «з нуля» за правилами, зможе ошукати суддю
- Тест був «заточений» саме під класичний (провалений) підхід до NLP з п. 1
- Сучасні LLM не будують мову «з нуля» — вони імітують людське мовлення, навчившись на мільярдах реальних текстів
- **Тому «пройти» тест Тьюрінга для LLM — очікувана поведінка за дизайном, а не доказ «інтелекту»**

# Що зараз вміє добре робити генеративний ШІ

1

## Обробка обсягів

Перебирати величезні обсяги даних під наглядом користувача

2

## Рутинні задачі

Виконувати шаблонну роботу швидко й послідовно

3

## Персональний асистент

Звільняти людину від рутини заради часу на мислення

*Спільна риса всіх трьох: людина лишається «в контурі» і перевіряє результат — ШІ пришвидшує, але не замінює нагляд.*

# Що зараз НЕ вміє добре робити генеративний ШІ

1

## Не відчуває ціну помилки

Не розуміє обмеження задачі без явно вказаного контексту

2

## «Додумує» факти

Не вміє видавати виключно перевірене — завжди генерує якусь відповідь

3

## Оминає етику

Не заглиблюється в етичні нюанси, якщо про це прямо не попросити

***Причина всіх трьох обмежень одна: первинна задача генеративної моделі — не дати правильну відповідь, а згенерувати текст, максимально схожий на людський.***

## ПРАКТИЧНИЙ ВИСНОВОК

**«ШІ — ваш асистент у вирішенні рутини,  
а не ви — асистент ШІ, що копіює  
результати його «мислення» без перевірки»**

*Відповідальність за результат генерації завжди лежить на тому, хто дав завдання.*

## ПІДСУМОВУЄМО

# Що варто запам'ятати з цієї лекції

- Класична спроба «порахувати» мову жорсткими правилами провалилась — мова хаотична, як базар, а не каса самообслуговування
- Бізнес «загорівся» ШІ тоді, коли люди почали сприймати його як співрозмовника, а не інструмент
- RLHF — мільярди людських оцінок, що донавчають модель; ми й раніше ненавмисно тренували ШІ (CAPTCHA, alt-текст)
- Тест Тьюрінга втратив сенс: LLM не будують мову з нуля, а свідомо імітують людське мовлення
- ГШІ добре опрацьовує рутину, але не відчуває ціну помилки й схильний «додумувати» факти — бо його задача генерувати гарний текст, а не істину

*Наступна лекція: побічні ефекти масового ШІ — ринок праці, когнітивний борг, довкілля.*