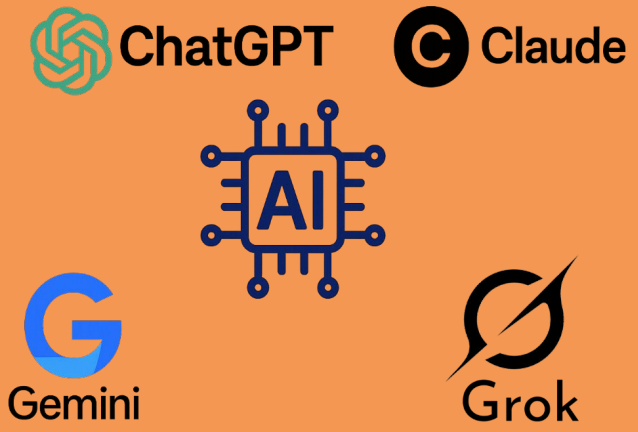
HEI-TRAIN. HEI Transformation for Entrepreneurship and AI-Driven Innovation

Funded by the European Union NextGenerationEU

Part of EIT Higher Education Initiative eit-hei.eu

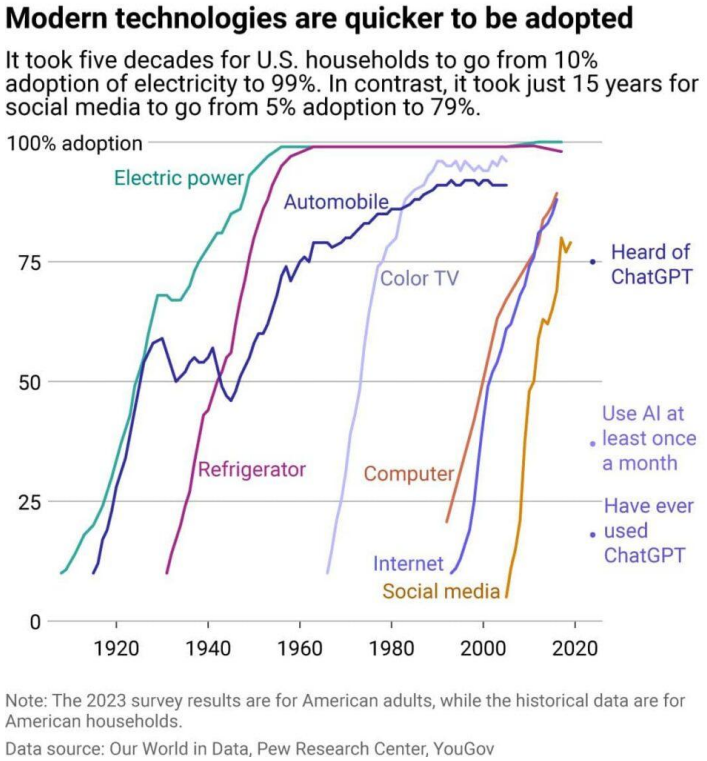
Сучасний штучний інтелект: великі мовні моделі (LLM)

Євгеній Владіміров

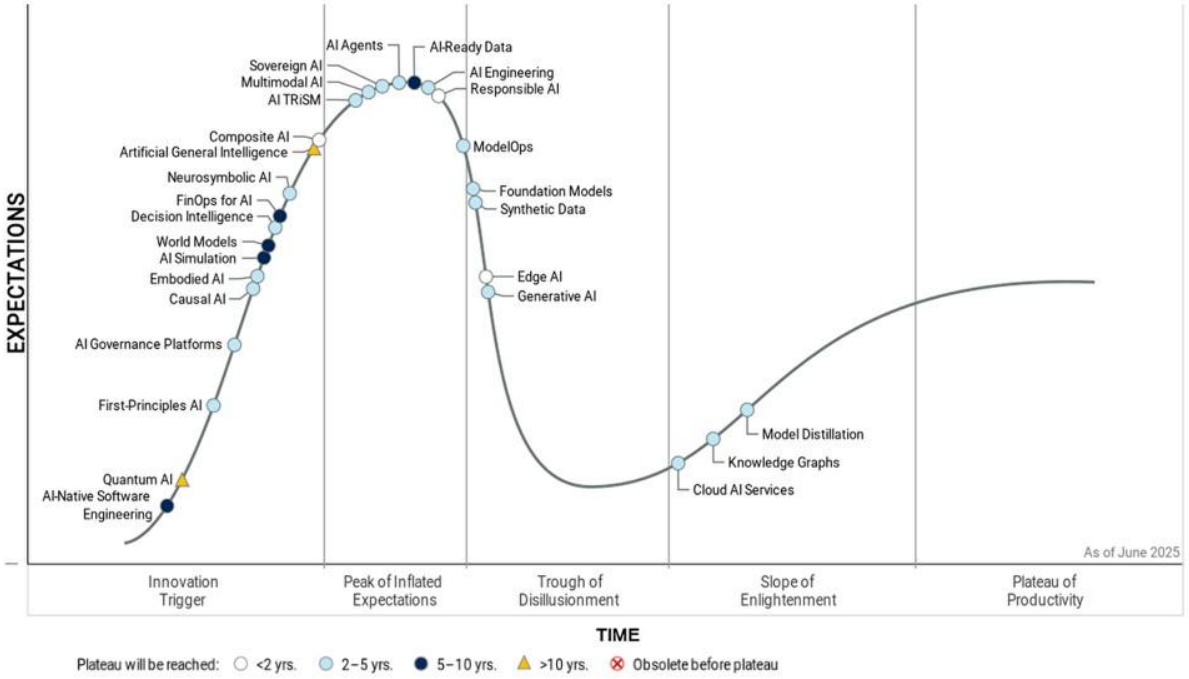


Вибух популярності AI

🌍 Глобальний ринок AI оцінюється приблизно у **$391 b**
📈 За 5 років індустрія AI зросте майже **у 5 разів**.
🔁 Середньорічний темп зростання (CAGR) **35,9%**.
👩‍💻 В **2025 році 97 мільйонів людей** працюють у сфері AI.
🏢 **83% компаній** вважають AI ключовим елементом своєї бізнес-стратегії.
🎬 **Netflix заробляє понад $1 млрд щороку** завдяки персоналізованим рекомендаціям AI.
💾 **48% бізнесів** використовують AI для ефективної роботи з великими даними.
🩺 **38% медичних закладів** застосовують комп'ютерний аналіз у процесі діагностики.



Різновиди AI на кривій



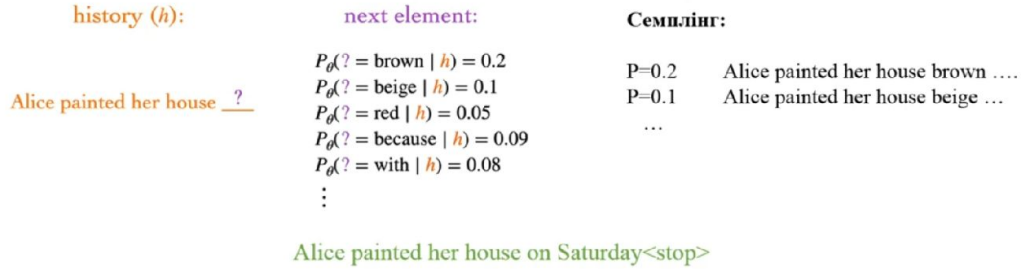
Gartner

Що таке LLM?

Модель мови навчається на мільярдах текстів, щоб **передбачати наступне слово** з урахуванням контексту.

$$P_\theta(X_{t+1} = x_{t+1} \mid x_1, \dots, x_t)$$

$X_{t+1} = x_{t+1}$ - наступний елемент
$x_1, \dots, x_t$ - попередні елементи (~3000 для ChatGPT)
Θ - параметри для навчання



Модель не “знає”, що таке будинок чи фарба, але **навчилась статистично передбачати логічні продовження**.

Саме це базовий механізм мислення сучасних LLM.

Безпрецедентні масштаби даних і параметрів

Особливість LLM це гігантський розмір і “Великий” у назві LLM це не метафора, а буквально мільярди параметрів.

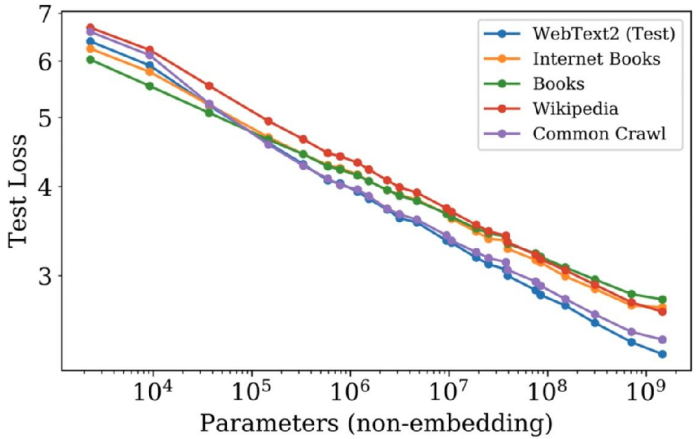
Найбільші моделі мають мільярди або навіть трильйони параметрів. Наприклад, для GPT-3 використали близько 500 мільярдів слів для навчання.

Зі зростанням розміру моделі (кількості параметрів) зменшується похибка (Test Loss).

Це означає, що модель краще розуміє закономірності мови навіть без додаткового навчання.

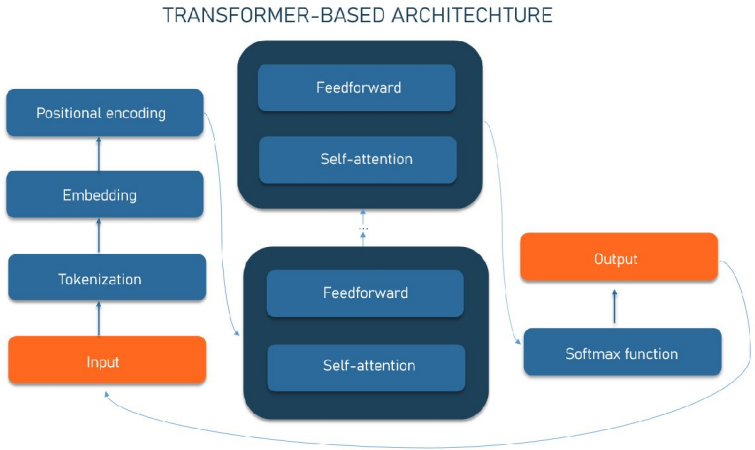
Ефект спостерігається стабільно для різних наборів даних.

Масштабування → універсальність:
великі моделі краще переносять знання між різними типами задач: переклад, код, діалог, аналіз тексту.



Архітектура трансформера

Tokenization: текст розбивається на невеликі фрагменти (токени)
Embedding: кожен токен перетворюється у вектор чисел, який відображає його зміст
Positional encoding: додається інформація про послідовність слів у реченні
Self-attention: модель визначає, які слова важливіші в контексті інших, і встановлює між ними зв'язки
Feedforward layers: обробка та узагальнення інформації після уваги
Output + Softmax: обрання найбільш ймовірного наступного слова



Self-attention

Ключова інновація трансформера — **механізм самоуваги (self-attention)**.
Модель “дивиться” на інші слова в реченні, щоб зрозуміти, як вони пов’язані між собою.

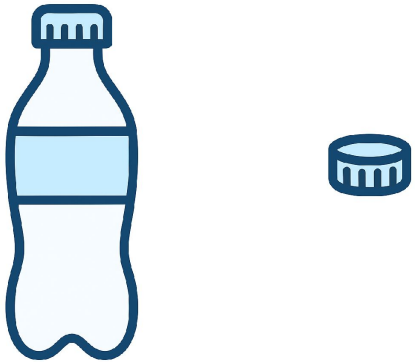
Наприклад:

«Я поставив пляшку на стіл і відкрутив кришку, бо вона була міцно закручена.»

Слово **«вона»** стосується саме **«кришки»**, а не **«пляшки»**.
Механізм уваги дозволяє моделі зробити це розрізнення.

Завдяки attention модель розуміє зв’язки між віддаленими словами й утримує контекст у межах довгих речень і абзаців.

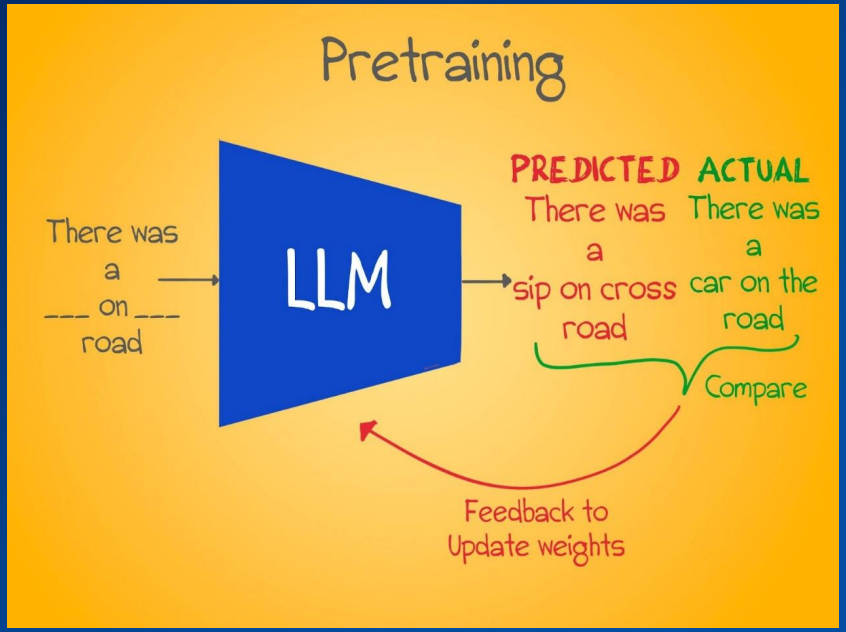
Вона - ?



Pre-training

На цьому етапі модель читає величезну кількість текстів з Інтернету, книжок, статей, форумів.
Вона вчиться **передбачати пропущене слово** — без готових правильних відповідей.
Це називається **самонагляд (self-supervised learning)**.

Поступово модель починає розуміти граматику, структуру речень, факти, логіку та зв'язки між словами.
Після цього етапу формується **"дика" базова модель**, яка вже може генерувати текст, але іноді говорить безглузді речі, бо навчалася на всьому підряд.

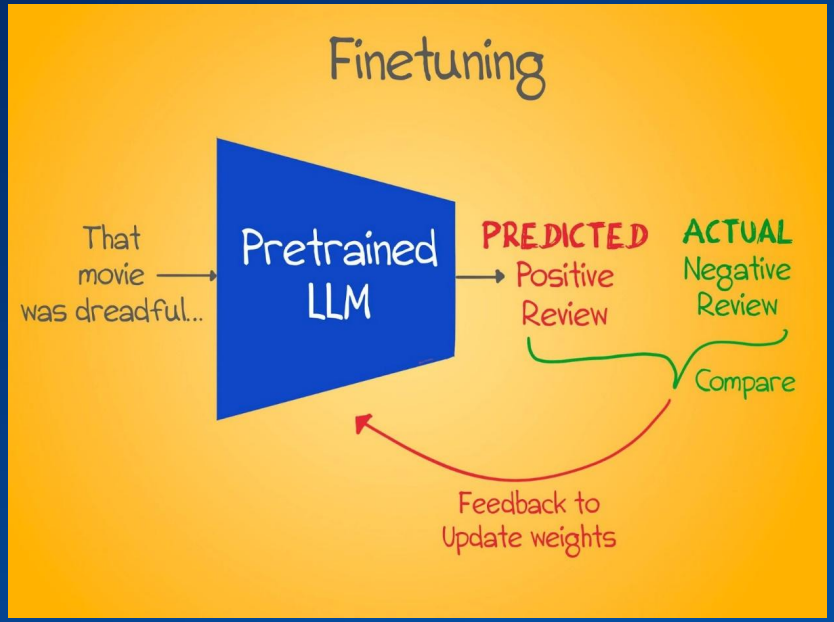


Fine-tuning

Після попереднього навчання модель донавчають під конкретні задачі.
Це **кероване навчання** на менших, якісно підготовлених наборах даних.

Модель отримує приклади запитань і правильних відповідей, діалогів, коду або текстів.
Люди позначають, де відповідь хороша, а де помилка.
Модель порівнює свій прогноз із правильною відповіддю і коригує свої параметри.

Після fine-tuning отримуємо **“вихованy” модель**, яка краще розуміє інструкції,
відповідає ввічливо і дає більш логічні результати.



HEI-TRAIN. HEI Transformation for Entrepreneurship and AI-Driven Innovation

Funded by the European Union – NextGenerationEU

Part of EIT Higher Education Initiative
eit-hei.eu

RLHF (Reinforcement Learning from Human Feedback)

Після fine-tuning модель уже слухняна, але ще не завжди розуміє, **що саме користувач вважає хорошою відповіддю**. Тому її навчають за допомогою **зворотного зв'язку від людини**.

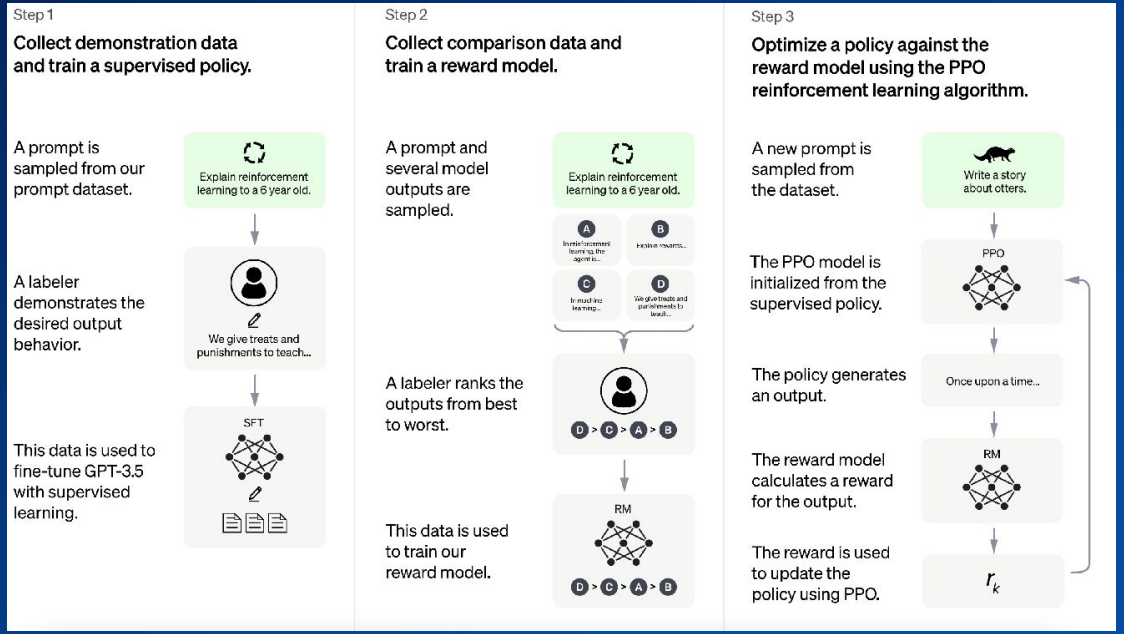
Цей етап складається з трьох частин:

1. Збір прикладів: люди показують бажані відповіді.

2. Оцінювання: кілька варіантів відповіді ранжуються від кращого до гіршого.

3. Підкріплення: модель отримує "нагороди" за кращі результати і поступово вчиться оптимізувати свої відповіді під людські очікування.

Так формується **"інструкційна модель" (instruction-tuned LLM)** — ChatGPT, Claude, Gemini тощо.
Вона не тільки знає факти, а **намагається бути корисною, чемною й зрозумілою**.



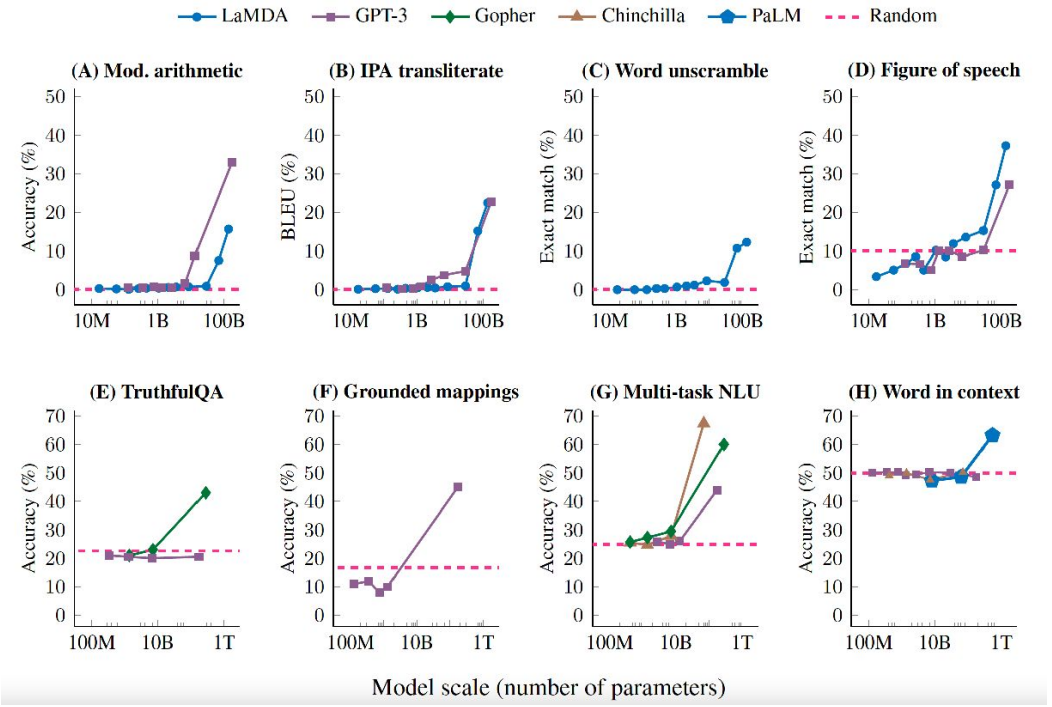
Несподівані здібності

Коли мовні моделі стають дуже великими, у них з'являються **емерджентні здібності**, нові вміння, яких не було в менших версіях.

Це якісний стрибок, коли модель раптом починає розв'язувати арифметичні задачі, перекладати, писати код або відповідати правдиво.

Поява таких здібностей відбувається нелінійно і непередбачувано.

Разом із новими можливостями виникають і потенційні ризики.



Нові здібності: логічне міркування

Великі моделі навчилися мислити послідовно (chain of thought)

GPT-4 уже розв'язує складні логічні та математичні задачі

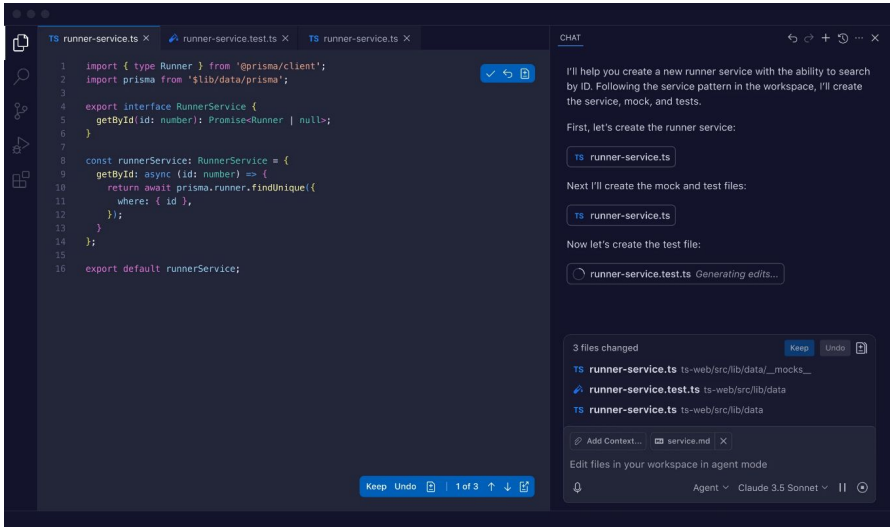
Невеликі моделі не справляються із завданнями, що потребують кількох кроків

Проте навіть великі LLM можуть "ламатися" на складних версіях, як у задачі "Ханойська вежа"

Ханойська вежа — приклад багатокрокової логічної задачі, де LLM вчаться міркувати послідовно.

Нові здібності: програмування

- Великі LLM навчилися писати і пояснювати код
- GPT-4 і Claude створюють функції та знаходять помилки
- Моделі індукували синтаксис із прикладів, без прямого навчання
- LLM працюють як “асистенти програміста”

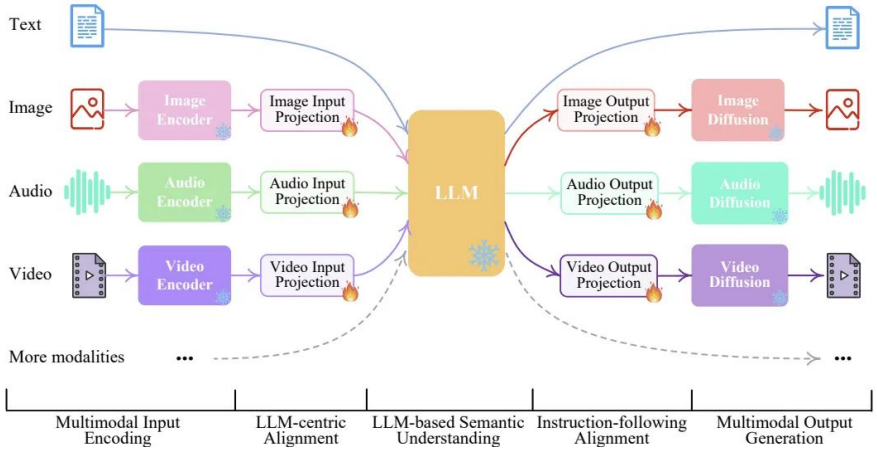


Нові здібності: багатомовність

- Великі LLM виявились поліглотами
- Розуміють десятки мов, навіть рідкісні
- Двомовність виникла природно без спеціального навчання
- Можуть переносити знання між мовами
- Відкривають шлях до мультимовних чат-ботів і перекладачів

Як LLM стають мультимодальними

Великі мовні моделі навчилися об’єднувати текст, зображення, аудіо та відео в одному контексті. Кожен тип даних кодується у спільний простір сенсів, що дозволяє моделі “бачити”, “чути” й “говорити” мовою змісту, а не лише слів. Завдяки цьому один алгоритм може описати зображення, перекласти аудіо чи створити відео у відповідь на запит.

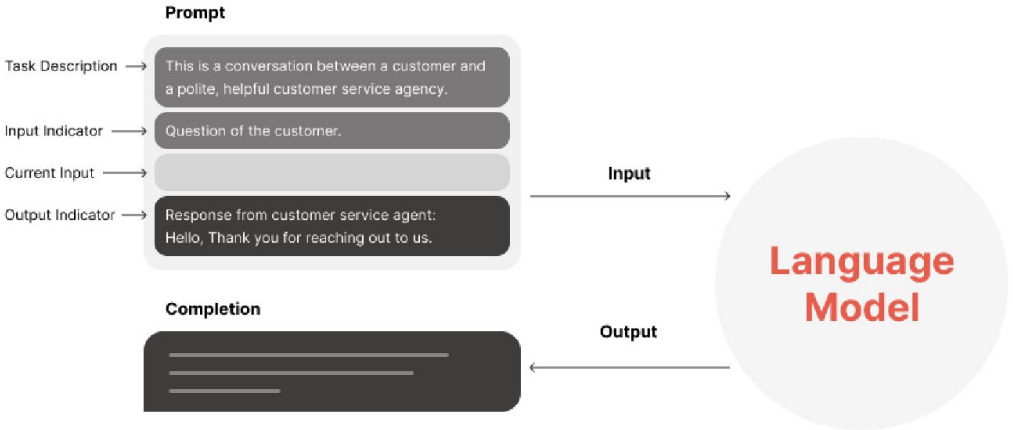


Обмеження LLM: “галюцинації”

Мовні моделі іноді вигадують факти, що звучать правдоподібно.
Такі помилки називають “галюцинаціями” AI.
Модель може з упевненістю цитувати неіснуюче джерело чи закон, або давати неточні дані.
GPT-4 має точність близько 78% — отже, кожна п’ята відповідь може бути хибною.
Причина в тому, що модель підбирає ймовірно правильні слова, не маючи реального розуміння істини.
Через це вона радше “придумає” відповідь, ніж скаже “не знаю”.

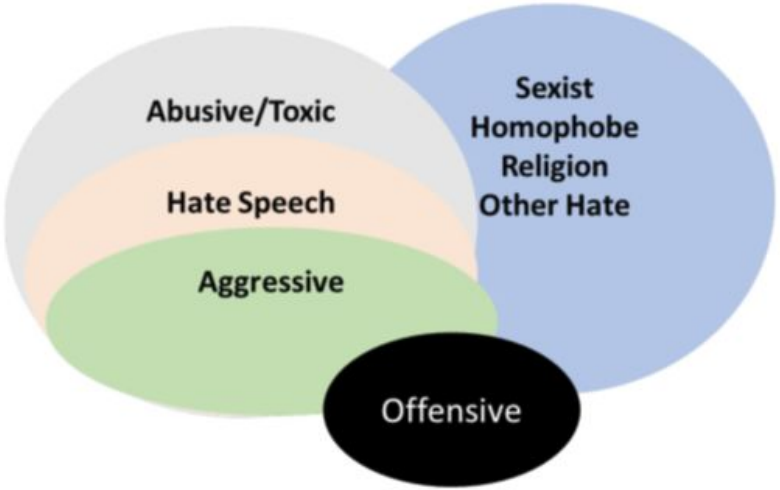
Обмеження LLM: чутливість до промптів

Великі мовні моделі можуть давати різні відповіді на одне питання, якщо змінити навіть кілька слів у запиті. Модель реагує не на зміст, а на форму подачі. Крім того, на довгих ланцюжках міркувань вона може втрачати логіку або "ламатися", коли складність виходить за межі того, що бачила під час навчання.



Обмеження LLM: упередження та токсичність

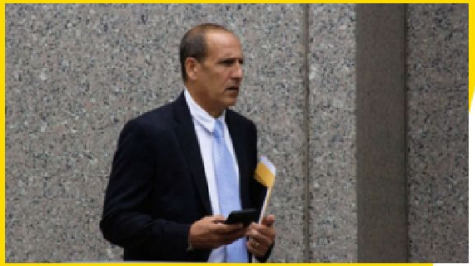
LLM навчаються на текстах, створених людьми, тому відтворюють стереотипи, що містяться в даних.
Навіть "вихованні" моделі можуть несвідомо продукувати упереджені або некоректні відповіді.
Іноді вони посилюють стереотипи щодо релігії, етнічності чи статі.
У крайніх випадках моделі здатні генерувати токсичний або агресивний контент, що становить ризик для публічного використання.



Приклад помилки: фейкові судові прецеденти

У 2023 році адвокат із Нью-Йорка використав ChatGPT для підготовки судового документа. Модель згенерувала посилання на вигадані судові справи, а юрист подав їх до суду без перевірки.
Суддя виявив підробки, а двох адвокатів оштрафували на 5000 доларів кожного.
Після цього суди США почали вимагати зазначати, чи використовувався ШІ під час підготовки документів.
Цей випадок став нагадуванням, що LLM не можна без перевірки використовувати у критично важливих сферах.

This US lawyer used ChatGPT to research a legal brief with embarrassing results. We could all learn from his error

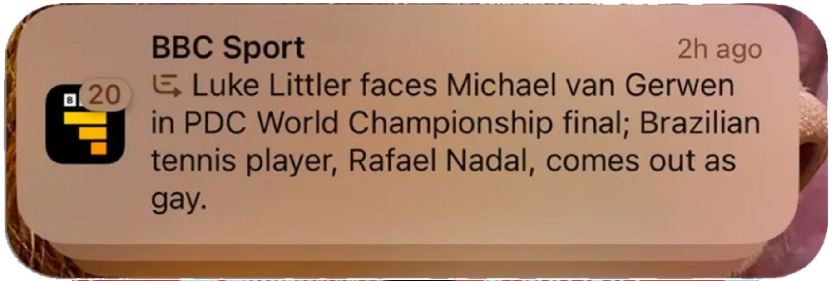


A New York-based lawyer has been fined after he misused the artificial intelligence chatbot, ChatGPT, relying on it for research for a personal injury case.

Published on the 24 June 2023

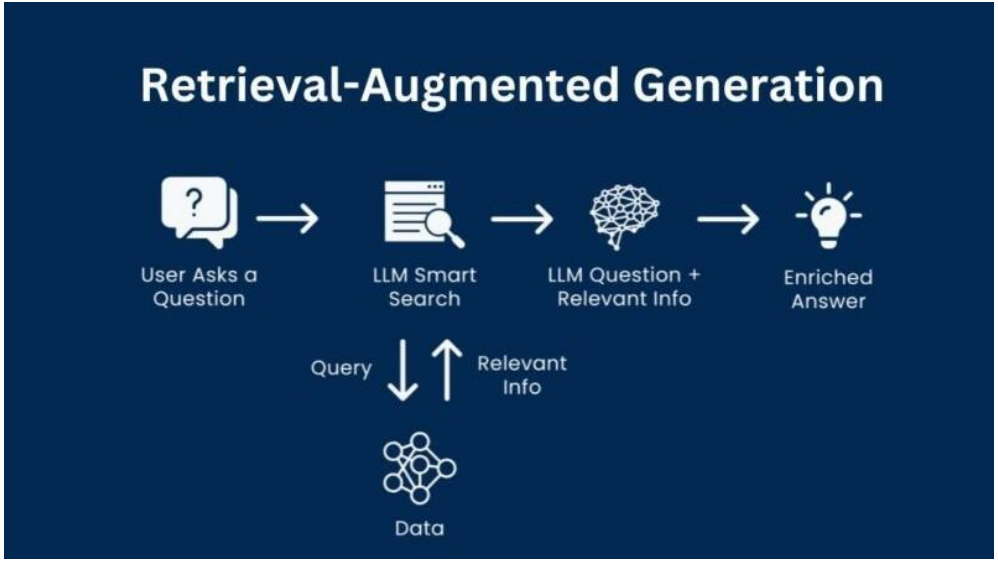
Приклад помилки: галюцинація в новинах

- Наприкінці 2024 року користувачі помітили дивне повідомлення від AI-сервісу коротких підсумків новин Apple.
- Модель у застосунку BBC Sport “повідомила”, що Рафаель Надаль зробив камінг-аут.
- Новина виявилася вигадкою: ШІ неправильно інтерпретував сповіщення й додав від себе “сенсаційні” деталі.
- Apple вибачилась і тимчасово вимкнула функцію.
- Випадок показав, що навіть допоміжні AI-сервіси можуть дезінформувати.



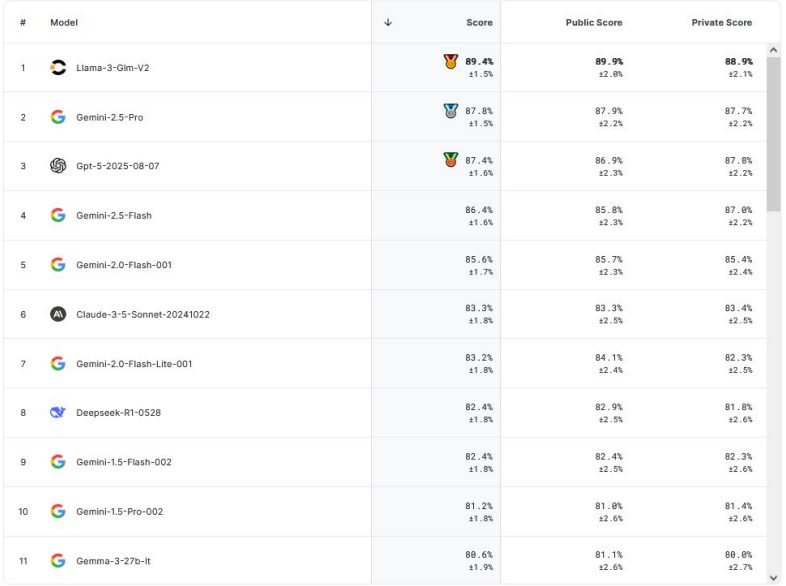
Подолання неточностей: RAG

- Retrieval-Augmented Generation поєднує LLM із пошуком у базах знань.
- Перед формуванням відповіді модель знаходить релевантні документи і використовує їхній зміст.
- RAG дає змогу уникнути вигадок, працювати з актуальними фактами та пояснювати джерела.
- Цей підхід застосовують Bing Chat, Perplexity AI, корпоративні помічники та внутрішні чат-боти.



Перевірка фактів і верифікація

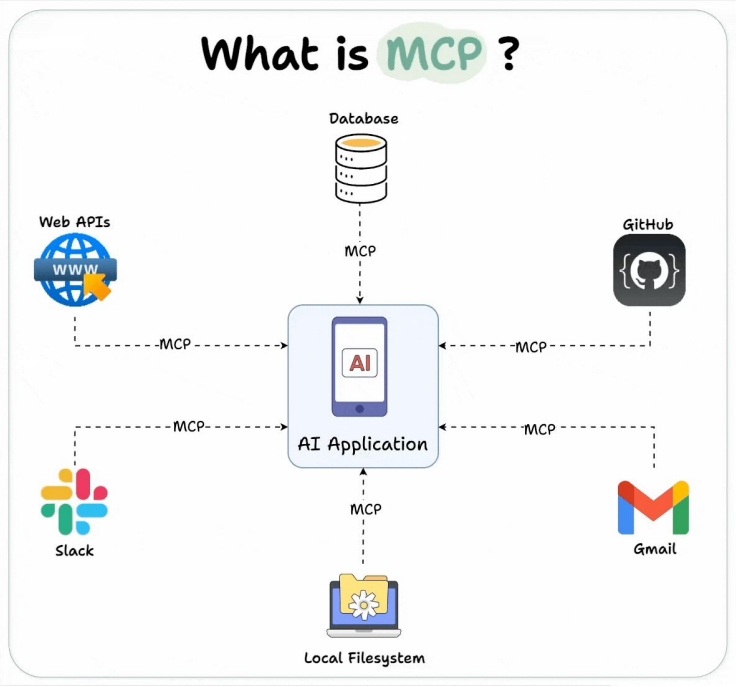
- Щоб зробити відповіді надійнішими, LLM починають комбінувати з інструментами перевірки фактів.
- Після генерації відповіді інша модель або модуль перевіряє твердження за базою даних чи правилами.
- Google DeepMind запропонувала FACTS Grounding підхід, що оцінює, наскільки відповідь узгоджується з реальністю.
- Мета щоб ШІ не просто “відповідав”, а пояснював, **звідки взяв інформацію** і наскільки впевнений у ній.



#	Model	Score	Public Score	Private Score
1	Llama-3-Glm-V2	89.4% ±1.5%	89.9% ±2.0%	88.9% ±2.1%
2	Gemini-2.5-Pro	87.8% ±1.5%	87.9% ±2.2%	87.7% ±2.2%
3	Gpt-5-2025-08-07	87.4% ±1.6%	86.9% ±2.3%	87.8% ±2.2%
4	Gemini-2.5-Flash	86.4% ±1.6%	85.8% ±2.3%	87.0% ±2.2%
5	Gemini-2.0-Flash-001	85.6% ±1.7%	85.7% ±2.3%	85.4% ±2.4%
6	Claude-3-5-Sonnet-20241022	83.3% ±1.8%	83.3% ±2.5%	83.4% ±2.5%
7	Gemini-2.0-Flash-Lite-001	83.2% ±1.8%	84.1% ±2.4%	82.3% ±2.5%
8	Deepseek-R1-0528	82.4% ±1.8%	82.9% ±2.5%	81.8% ±2.6%
9	Gemini-1.5-Flash-002	82.4% ±1.8%	82.4% ±2.5%	82.3% ±2.6%
10	Gemini-1.5-Pro-002	81.2% ±1.8%	81.0% ±2.6%	81.4% ±2.6%
11	Gemma-3-27b-It	80.6% ±1.9%	81.1% ±2.6%	80.0% ±2.7%

Гібридні підходи: Model Context Protocol

- Сучасні системи поєднують LLM із зовнішніми алгоритмами та базами знань.
- Нейромережа генерує текст, а модулі логіки або обчислень перевіряють його точність.
- Модель може самостійно викликати калькулятор, календар чи базу знань, коли потрібна перевірена інформація.
- Такі гібридні рішення компенсують слабкі сторони LLM і роблять результати надійнішими.



Де LLM особливо корисні сьогодні

LLM уже застосовуються в ключових напрямках роботи з текстом і даними: вони генерують, перекладають, аналізують, підсумовують і відповідають. Основні напрями:

Text Generation: створення статей, постів, сценаріїв

Machine Translation: переклад і багатомовна комунікація

Sentiment Analysis: визначення емоційного тону тексту

Question Answering: інтелектуальні чат-боти й асистенти

Content Creation: маркетинг, освіта, медіа

Де потрібна обережність у застосуванні LLM

LLM можуть допомагати у складних сферах, але не замінюють людську відповідальність. Там, де ціна помилки висока, модель має працювати лише під контролем експерта. Приклади сфер ризику:

Медицина: помилки в дозуванні чи діагнозі можуть бути небезпечні

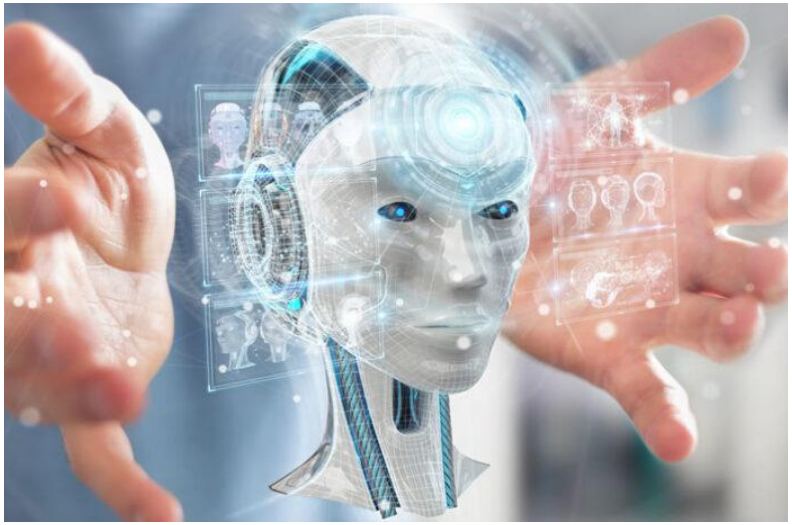
Право: вигадані факти й судові справи

Фінанси: ризик втрат через неточні поради

Освіта: поширення хибних знань і списування

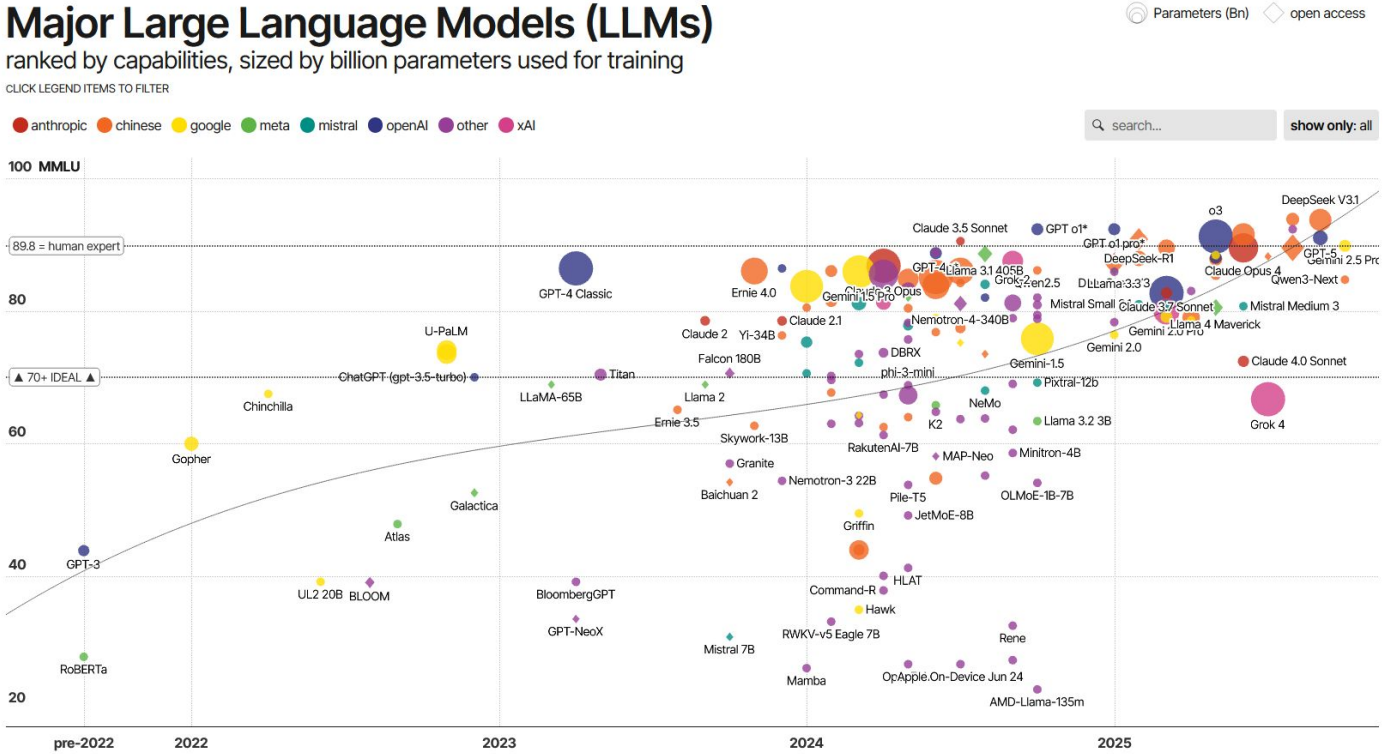
Журналістика: недостовірні або вигадані новини

Автономні системи та оборона: критичні рішення без етичної оцінки



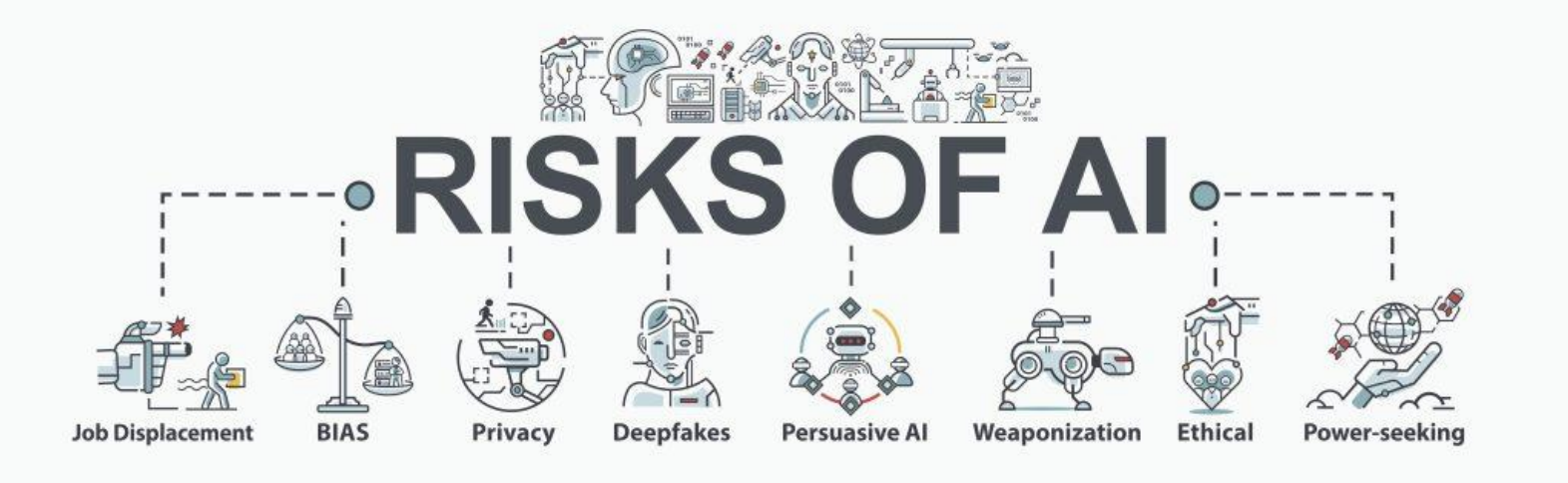
Major Large Language Models (LLMs)

ranked by capabilities, sized by billion parameters used for training



David McCandless, Tom Evans, Paul Barton
Informationisbeautiful // Sep 2025

MMLU = benchmark for measuring LLM capabilities
* = parameters undisclosed // source: LifeArchitect // data



Штучний інтелект відкриває величезні можливості, але разом із ними приходять і ризики: від упереджень і втрати приватності до дезінформації та витіснення людей з робочих місць. Без чітких етичних і правових рамок інновації можуть перетворитися з інструменту прогресу на джерело загроз, тому відповідальне використання AI стає питанням безпеки майбутнього.

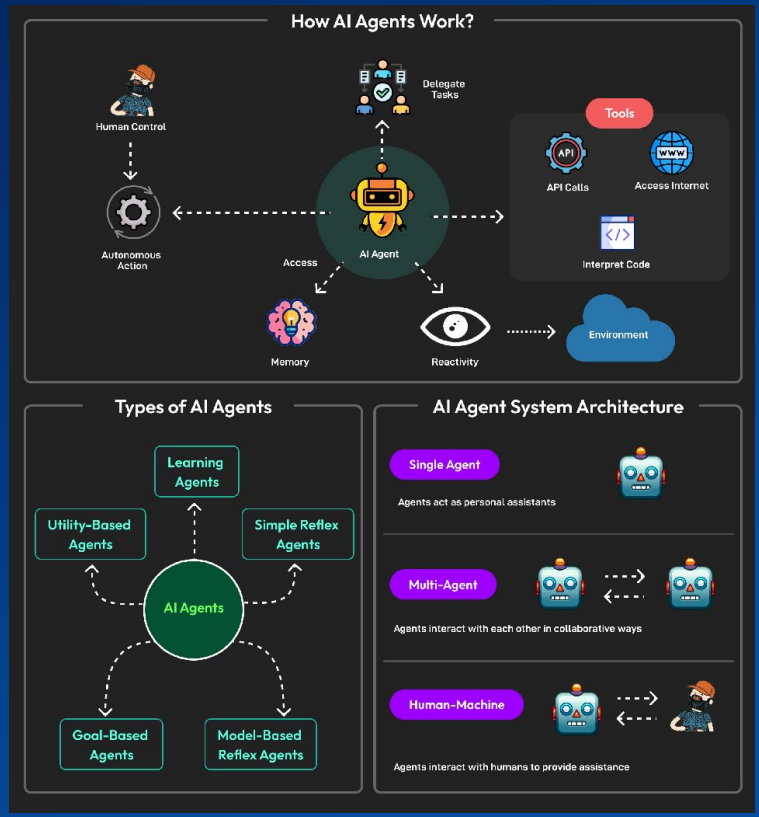
HEI-TRAIN. HEI Transformation for Entrepreneurship and AI-Driven Innovation

Funded by the European Union – NextGenerationEU

Part of EIT Higher Education Initiative
eit-hei.eu

Майбутнє: автономні агентні системи

Наступний етап розвитку ШІ це агентні системи. LLM перестають просто реагувати на запити: тепер вони планують, діють і співпрацюють. AI-агент має пам’ять, доступ до інструментів і може самостійно виконувати завдання, звітуючи людині. Ми рухаємося від “чат-ботів” до справжніх цифрових колег.



Дякую за увагу!