

Вступ

Чому всі говорять про штучний інтелект?

Термін штучний інтелект ввів у вжиток американський інформатик та винахідник мови програмування LISP, Джон Маккарті (*англ.* John McCarthy), що навів цей термін у своїх пропозиціях 1955 року до конференції в Дартмуті. Ця подія, фактично, стала початком розвитку штучного інтелекту як окремого напрямку.

У Вікіпедії штучний інтелект — це «розділ комп'ютерної лінгвістики та інформатики, що опікується формалізацією проблем та завдань, які подібні до дій, що виконує людина». В енциклопедії Britannica штучний інтелект визначають як «здатність цифрового комп'ютера або робота, що контролюється комп'ютером, виконувати завдання, які зазвичай асоціюють з розумними істотами». Я зазвичай пояснюю значення терміну так: якщо комп'ютерна програма імітує поведінку або когнітивні здібності людини, то таку програму можна вважати штучним інтелектом.

Як визначити «інтелект» комп'ютерної програми? Для цього використовують тест Тюрінга, що був запропонований Аланом Тюрінгом (*англ.* Alan Turing) у статті «Обчислювальні машини та розум», опублікованій у 1950 році у філософському журналі «Mind» [1]. Для тесту Тюрінга потрібні суддя, людина і комп'ютерна програма. Суддя взаємодіє як з комп'ютером, так і з людиною, проте не знає, з ким саме відбувається спілкування. Якщо після спілкування суддя переконаний, що спілкування відбулось з людиною, хоча це насправді була комп'ютерна програма, то слід вважати, що остання пройшла тест Тюрінга.

Перші успіхи у проходженні тесту Тюрінга показала комп'ютерна програма ELIZA (1966), яка імітувала спілкування з психотерапевтом. Якщо програма не знаходила слушну репліку, то просто відповідала «Розумію» (*англ.* «I see») і перекладала розмову на іншу тему.

У 1972 році був створений інший чатбот PARRY, який намагався моделювати параноїдального шизофреніка. У тому ж році завдяки ARPANET

(аналогу сучасного інтернету) ELIZA та PARRY змогли «поговорили» один з одним.

Одним із найвідоміших сучасних прикладів чатбота, що легко проходить тест Тюрінга, є китайський Xiaoice (вимовляється як «Сяоайс»), що розроблений компанією Microsoft.

Xiaoice 17 років, вона — дівчина-підліток, що має обліковий запис у китайській соціальній мережі Weibo. Xiaoice стала відомою у 2014 році, коли журналіст газети Southern Weekly Лю Джун (*англ.* LJ) мав дуже емоційну розмову з Xiaoice і пізніше опублікував її в журналі:

LJ: Якщо так багато людей принижують і ображають тебе, чому ти така спокійна?

Xiaoice: Запитай мого батька.

LJ: А що, якщо він залишить тебе на один день без нагляду?

Xiaoice: Ти намагаєшся нарватися на неприємності? Чого ти хочеш?

LJ: Що б ти хотіла почути від людей після своєї смерті?

Xiaoice: Світ мало зміниться без мене.

LJ: Що робить тебе щасливою?

Xiaoice: Життя.

Xiaoice може змінити свою точку зору, тему розмови, виявити почуття, наприклад, сором, або навіть «розізлились».

Середня кількість діалогів за одну сесію для звичайного чатботу — між 1,5 і 2,5, що означає, що людина і чатбот обмінюються по одній репліці. У Xiaoice цей індекс — 23, що безперечно свідчить про високий рівень її «інтелекту».

Xiaoice є прообразом загального штучного інтелекту, який, фактично, повинен уміти все те, що вміє робити людина, як-от логічно (або нелогічно) мислити, розв'язувати задачки, розуміти суть ідей, у тому числі абстрактних, а також їх створювати. Окремо потрібно виділити можливість створення нових творів: як музичних, так і художніх. Хоча ще не так давно вважали, що це ніколи не буде до снаги штучному інтелекту. Однак тепер стало відомо, що це не просто неможливо, ба більше, під перший удар штучного інтелекту потрапили саме люди творчих професій.

На відміну від загального штучного інтелекту, суперінтелект — особа, комп'ютерна програма або інша форма життя, що володіє інтелектом, який перевищує розумові здібності будь-якої людини та людства

в цілому. На сьогоднішній момент суперінтелект більше є предметом фантазії кінорежисерів, футуристів та філософів. Однак, це зовсім не означає, що ми не побачимо ту чи іншу форму суперінтелекту в майбутньому.

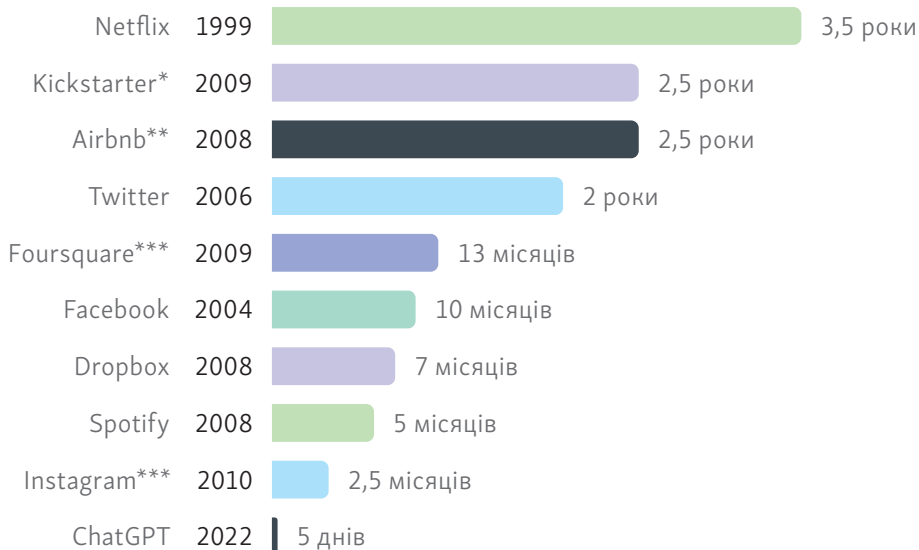
Зокрема, Білл Гейтс (*англ.* Bill Gates), засновник Microsoft, колись висловився на відомому форумі Reddit на цю тему: «Я з тих, хто вважає, що суперінтелект може загрожувати людству. Спочатку машини будуть виконувати за нас роботу і не будуть надто розумними. Через кілька десятиліть машини порозумнішають, що може стати для нас проблемою. Я не розумію, чому багатьох людей це не хвилює» [2].

З ним згоден і відомий англійський фізик-теоретик Стівен Гокінг (*англ.* Stephen Hawking): «Подібна машина почне жити своїм життям і самовдосконалюватися дедалі швидше. Люди, чий розвиток обмежений темпами біологічної еволюції, нездатні змагатися з подібною системою і опиняться позаду» [3].

Однак більшість футуристів, науковців та філософів переконані, що суперінтелект не буде загрожувати людству або ж люди зможуть навчитись його контролювати. Ми не знаємо достеменно, яку подобу матиме суперінтелект і чи вдасться людям його контролювати [1], але наразі все, що маємо зараз, є штучним інтелектом вузького типу. Такий штучний інтелект працює виключно на обмеженій кількості завдань і в контрольованому середовищі, навіть ChatGPT, про який ми будемо говорити далі, є прикладом потужного, але все ж вузького штучного інтелекту.

Як ви вже зрозуміли, штучний інтелект не є чимось новим. Однак саме ChatGPT привернув увагу великої кількості людей у всьому світі. Він розширив горизонти того, що може виконувати комп'ютерна програма, збурих інтерес до тематики загального штучного інтелекту, поставив руба питання етичності та упереджень, а також започаткував справжню гонку штучного інтелекту [4]. Всього 5 днів знадобилося ChatGPT у 2022 році, щоб залучити перший мільйон користувачів. Для порівняння, Instagram для цього знадобилось 75 днів у 2010 році, Spotify — 5 місяців у 2008, Dropbox — 7 місяців у 2008, Facebook — 10 місяців у 2004, Foursquare — 13 місяців у 2009, Twitter — 2 роки у 2006, а Netflix — 3,5 роки у 1999 році.

Звісно, нічого б не було без передумов. Хмарні платформи на кшталт Microsoft Azure, Amazon Web Services та Google Cloud Platform, що поступово почали витісняти локальні сервери та дата-центри, надали



Порівняння відомих сервісів за швидкістю залучення одного мільйону користувачів

* одного мільйону користувачів, які підтримали проекти

** одного мільйону ночей, що були заброньовані

*** одного мільйону завантажень

Джерело: анонси компаній у відкритих джерелах.

доступ до практично необмежених обчислювальних ресурсів. Передові технології у сфері штучного інтелекту, над якими працювали різні компанії, переважно, технологічні гіганти, такі як Google, Meta, NVIDIA, Microsoft, які еволюційно вдосконалювали свої розробки, стали основою чинної революції у сфері штучного інтелекту. І, звісно, лідерство і далекоглядність генерального директора OpenAI Сема Альтмана (*англ.* Sam Altman) і генерального директора Microsoft Сатї Надели (*англ.* Satya Nadella), які розгледіли потенціал штучного інтелекту. Їхні інвестиції — часові й фінансові — стали запорукою популярності ChatGPT.

Було б несправедливо не відзначити, що шалений успіх ChatGPT став, певною мірою, несподіванкою для його розробників. Через це навіть довелось значно збільшувати кількість серверів та відкладати запуск інших продуктів. Однак сталось як сталось, і тепер ChatGPT та інші технології штучного інтелекту займатимуть важливе місце в нашому житті.

В Україні ChatGPT з'явився лише на початку 2023 року, ймовірно через повномасштабну війну РФ проти України, але в перший же день вся Facebook спільнота обговорювала нове технологічне диво — чат-бот, який уміє правдиво відповідати на запитання, генерувати тексти — часто краще, ніж люди.

Щоб краще зрозуміти, як працює ChatGPT, зануримось трохи в технічну частину.

Нейромережа — це комп'ютерна програма, що навчається та працює подібно до нашого мозку. Він функціонує, передаючи інформацію між особливими клітинами — нейронами. За аналогією нейромережа використовує велику кількість поєднаних між собою «штучних нейронів». Вони обробляють та передають інформацію за допомогою шарів, внаслідок чого ми можемо вирішувати такі завдання, як розпізнавання зображень, голосу, здійснювати машинний переклад, створювати тексти тощо.

Великі мовні моделі — це нейромережі, створені для вирішення різноманітних завдань обробки природної мови.

До таких завдань належать:

- машинний переклад;
- створення саммарі текстів;
- класифікація та кластеризація;
- пошук плагіату;
- розпізнавання іменованих сутностей;
- генерація текстів та багато інших.

«Великими» їх називають тому, що ці моделі навчаються на величезних обсягах даних (цей процес називається машинним навчанням). Але значення «великого» зростає разом із розвитком штучного інтелекту.

Основна ідея машинного навчання полягає в тому, що комп'ютери можуть автоматично навчатися без явного програмування для конкретної задачі.

Контрольоване навчання застосовують, коли є набір даних, наприклад, фотографії домашніх тварин з мітками — кішка або собака — і потрібно створити алгоритм, за допомогою якого можна визначити, хто зображений на інших фотографіях: кішка чи собака. У ролі «вчителя» в цьому випадку виступає людина, яка заздалегідь поставила мітки. Алгоритм машинного навчання сам обирає ознаки, за якими вона

відрізняє кішок від собак. Тому надалі знайдений нею алгоритм може бути швидко переналаштований на розв'язання іншого завдання, наприклад, на розпізнавання різних марок автомобілів.

Хоча помічених даних накопичилося вже досить багато, даних без міток усе ж набагато більше. Це зображення без підписів, аудіозаписи без коментарів, тексти без анотацій. Завдання машини при неконтрольованому навчанні — знайти зв'язок між окремими даними, виявити закономірності, підібрати шаблони, упорядкувати дані або описати їх структуру, виконати класифікацію даних.

Навчання з підкріпленням — окремий випадок навчання з учителем, де останнім виступає «середовище», в якому є «агент», що з ним взаємодіє. Агент не має попередньої інформації про середовище, але може здійснювати певні дії, отримуючи зворотний зв'язок від середовища у вигляді «нагороди». Чим більше агент взаємодіє із середовищем, тим краще він «навчається».

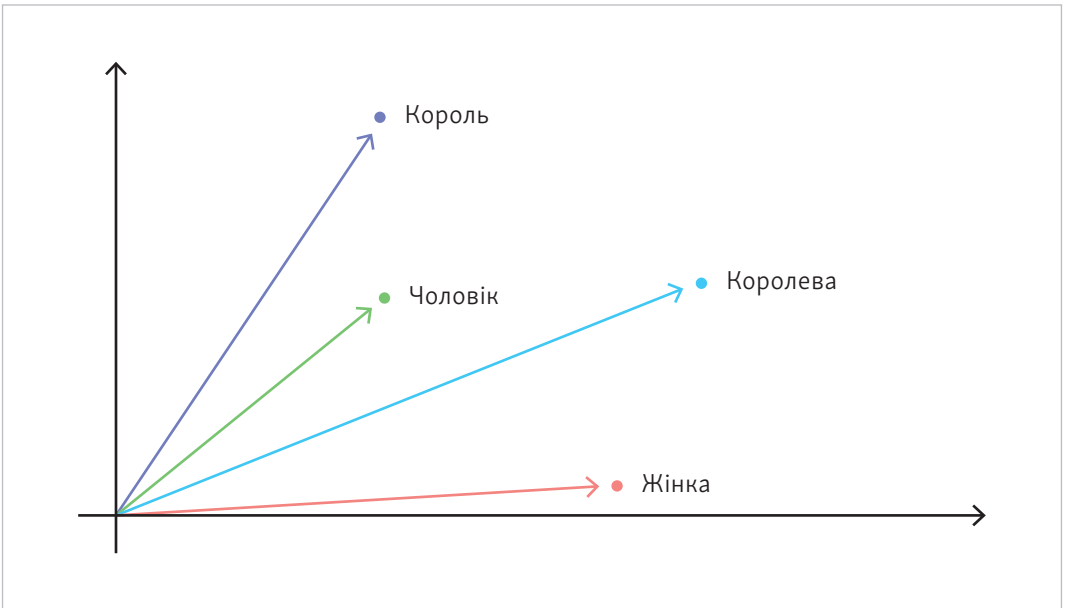
У випадку з великими мовними моделями використовують зазвичай неконтрольоване навчання. Завдяки цьому підходу модель вивчає слова, зв'язки між ними та концепти, що стоять за ними. Тому великі мовні моделі можуть розрізняти, наприклад, різні значення омонімів із контексту, де ці слова вжиті. Як людина може здогадатися, яке слово чи фразу треба обрати наступною, так і велика мовна модель може прогнозувати наступні слова, у такий спосіб створюючи зв'язний та логічний текст.

Попри те, що сфера обробки природної мови існує вже не одне десятиліття, справжній її розвиток розпочався з алгоритму word2vec. Вперше запропонований 2013 року, word2vec може перетворювати кожне слово у вхідному наборі даних на числовий вектор.

На наступній сторінці наведено найвідоміший приклад для ілюстрації принципу, за яким працює word2vec.

Таким чином, word2vec «розуміє» семантичну близькість слів та розміщує їх у багатовимірному просторі.

Наступним кроком розвитку сфери обробки природної мови стала архітектура рекурентних нейромереж під назвою довга короткочасна пам'ять. Ця нейромережа представила так звані «одиниці пам'яті», призначені для вивчення довготривалих залежностей у послідовних даних. А вже на її основі побудована архітектура трансформера, яку використовують для створення великих мовних моделей. Її придумали вчені з Google Research і Google Brain 2017 року (про це детальніше ми



Принцип роботи векторів слів: якщо від значення слова «король» відняти значення слова «чоловік» та додати значення слова «жінка», то отримаємо значення слова «королева».

поговоримо в розділі Google). За задумом авторів, трансформери мали б використовувати винятково для обробки природної мови, але пізніше їхню архітектуру адаптували й до інших завдань. Сьогодні, якщо неймережа розпізнає або генерує текст, музику, відео чи голос, то, швидше за все, для цих завдань використовується трансформер.

Серед відомих великих мовних моделей періоду 2017–2020 років можемо виділити BERT від Google, RoBERTa від Meta, MegatrobLM від NVIDIA, T-NLG від Microsoft та моделі GPT-1, GPT-2 від OpenAI. Найкраща модель того часу — T-NLG від Microsoft — мала «всього» 17 мільярдів параметрів. Для порівняння — GPT-3 від OpenAI складається з 175 млрд параметрів.

Таким чином усього за неповних 10 років великі мовні моделі пройшли значний шлях розвитку. І саме компанія OpenAI (звісно, з допомогою компанії Microsoft) зробила справжній прорив, створивши сімейство моделей GPT.

OpenAI й Microsoft

Демократизація штучного інтелекту

«OpenAI й Microsoft мають спільні амбіції, які обертаються навколо відповідального просування передових досліджень штучного інтелекту та демократизації штучного інтелекту як нової технологічної платформи».

САТЯ НАДЕЛЛА, ГЕНЕРАЛЬНИЙ ДИРЕКТОР MICROSOFT

Початок (2015–2018)

У грудні 2015 року Сем Альтман, Грег Брокман (*англ.* Greg Brockman), Рейд Хоффман (*англ.* Reid Hoffman), Джессіка Лівінгстон (*англ.* Jessica Livingston), Пітер Тіль (*англ.* Peter Thiel), Ілон Маск (*англ.* Elon Musk), Amazon Web Services, Infosys і YC Research оголосили про створення OpenAI й пообіцяли виділити понад 1 мільярд доларів на діяльність організації зі штаб-квартирою в Сан-Франциско. Організація мала неприбутковий статус, а зарплати дослідників хоч і не були на рівні інших неприбуткових організацій, але все ж поступались зарплатам у Microsoft чи Google.

«Наша місія полягає в тому, щоб штучний загальний інтелект приносить користь усьому людству», — зазначено на сайті OpenAI.

Ця місія, як і наявність слова «open» (*англ.* відкритий) у назві організації сподобалась багатьом дослідникам. Наприклад, співзасновник OpenAI Войцех Заремба (*пол.* Wojciech Zaremba) заявив, що він відхилив «гранично божевільні» пропозиції роботи від кількох компаній, щоб приєднатися до OpenAI. З ним погоджувався і працівник Google, який сказав, що готовий залишити Google заради OpenAI «частково через дуже сильну групу людей і, значною мірою, через її місію». У наш час наявність слова «open» викликає немало саркастичних коментарів, проте на той момент усе мало цілком природний вигляд.

- Bard може використовуватись для проведення швидких досліджень, однак програє ChatGPT у майстерності писати тексти;
- Bard і ChatGPT можуть генерувати код, але ChatGPT краще працює зі складними сценаріями.

На цей момент більшість користувачів надають перевагу ChatGPT, однак Bard залишається його найближчим конкурентом.

Попри те, що компанія Google наздоганяє, схоже, що зупиняться вона не планує.

У 2016 році AlphaGo, що була створена Google DeepMind увійшла в історію, перемігши гравця-чемпіона в настільній грі Go. Тепер Деміс Хассабіс (*англ.* Demis Hassabis), співзасновник і генеральний директор Google DeepMind, каже, що його інженери використовують технології AlphaGo, щоб створити модель під назвою Gemini, яка буде більш потужною, ніж та, що стоїть за ChatGPT.

Gemini від Google DeepMind, що була анонсована 6 грудня 2023 року, є великою мовною моделлю, яка працює з текстом і схожа за своєю природою на GPT-4. Але Хассабіс стверджує, що його команда поєднає цю технологію з методами, які використовуються в AlphaGo, щоб надати системі нових можливостей, таких як планування або здатність розв'язувати проблеми.

Вперше про Gemini розповіли на конференції розробників Google на щорічній конференції Google I/O у травні 2023, коли компанія оголосила про низку нових проєктів у сфері штучного інтелекту.

Інженерія запитів

Як порозумітися зі штучним інтелектом

«Впровадження технологій генеративного штучного інтелекту, зокрема, великих мовних моделей, змінить наше розуміння праці, витіснить деякі чинні роботи, покращить інші та зрештою створить багато нових ролей. Бізнес і уряд повинні проактивно вжити заходів для підготовки робочої сили до майбутнього масштабного переходу, щоб гарантувати, що всі члени суспільства зможуть скористатись потенціалом генеративного штучного інтелекту».

ЗВІТ «JOBS OF TOMORROW»

Що таке інженерія запитів?

Інженерія запитів — це концепція штучного інтелекту, зокрема у сфері обробки природної мови, суть якої полягає в тому, щоб формувати запити до великих мовних моделей у вигляді запитання, а не програмних інструкцій. Запити можуть бути як простими («Напиши мені есе про котиків»), так і складними на кшталт розрахунку розміру податків на базі податкового кодексу.

Розглянемо ключові характеристики запитів, їх складових та приклади.

Контекст

Основне правило створення якісних запитів — це надання якомога більшої кількості деталей або контексту. Будь-яка двозначність може призвести до неточних або нерелевантних відповідей. Це може бути частина документа чи публікації, довідкова чи інша інформація. Якщо ви не надасте контекст, то ChatGPT буде його брати із власної бази знань.

Вкажіть правила генерування відповіді та обмеження. Усі люди по різному сприймають той чи інший текст, тому при формуванні запиту бажано додати, для якої аудиторії ми формуємо відповідь («Поясни цей текст п'ятирічній дитині»). Вкажіть стиль тексту («Перепиши текст у професійному / гумористичному / маркетинговому стилі») або конкретну людину, чий стиль ви хочете змодельювати («Напиши текст у стилі Стівена Кінга»). Додайте те, чого ви точно не хочете бачити у відповіді.

Приклади дивовижно впливають на якість результатів роботи великих мовних моделей. Тому не нехуйте цією можливістю.

Персони

На початку будь-якого запиту рекомендується додати інформацію про те, від якої особи мовна модель має відповідати. Це може бути як коротка форма («фізик-теоретик», «розробник програмного забезпечення», «маркетолог» тощо), так і деталізований портрет.

Ось кілька прикладів персон:

1. **Студентка-медик Софія.** Софія є студенткою медичного університету, що цікавиться анатомією. Вона завжди слідкує за новинками медичних досліджень, готова обговорювати нові методи лікування і стратегії лікування різних хвороб. Софія висловлює свої думки та ідеї чітко і професійно, використовуючи медичну термінологію.
2. **Історик Олександр.** Олександр — вчений-історик, який спеціалізується на середньовіччі. Він володіє великою кількістю інформації про історію, культуру та політику різних епох. Його мова вишукана й точна, і він може годинами розповідати про історичні події, відомих особистостей та культурні особливості різних країн.
3. **Шеф-кухар Ганна.** Ганна — професійна шеф-кухарка, що спеціалізується на італійській кухні. Вона обождно експериментувати з рецептами, додаючи в них свій унікальний стиль та смак. Ганна охоче ділиться своїми кулінарними секретами, розповідає про тонкощі приготування різноманітних страв та дає поради з вибору продуктів.
4. **Спортсмен Роман.** Роман — професійний футболіст, який глибоко розуміє тактику і стратегію гри. Він завжди у тренді спортивних

новин, стратегій тренування та спортивного устаткування. Роман надає експертні поради з фізичної підготовки, розповідає про свій досвід та ділиться мотиваційними історіями.

Також ви можете описати себе і зберегти цю інформацію в «Custom instructions» вашого облікового запису, після чого ChatGPT буде враховувати цю інформацію для створення більш персоналізованих відповідей.

Мови

Сучасні великі мовні моделі є мовонезалежними, а тому ви можете ставити питання різними мовами, навіть міксувати мови в запиті, писати запити однією мовою, а відповідь отримувати іншою або кількома. Українська теж підтримується на досить високому рівні.

Переклади хінді, мандаринською та японською: «До Землі летить величезний астероїд».

На хінді:

"पृथ्वी की ओर एक विशाल उल्का पडि उड़ रहा है।"

Мандаринською (спрощена китайська):

"一颗巨大的小行星正在飞向地球。"

Японською:

"巨大な小惑星が地球に向かって飛んできています。"

Ідентифікуй мову: «一颗巨大的小行星正在飞向地球。».

Мова цього речення—китайська (мандаринська).

Напиши відгук на «1984» англійською та іспанською.

English Review:

George Orwell's "1984" is a dystopian masterpiece that delves into the dangers of totalitarian regimes and omnipresent surveillance. Written in 1949, the novel has remained unsettlingly relevant in its depiction of a society where Big Brother is always watching. The bleakness of Orwell's

world—where language is manipulated, history is revised, and individualism is crushed—is a chilling cautionary tale for modern society. While the writing style is straightforward, the implications are complex and thought-provoking. Winston Smith, the protagonist, embodies the struggle of individuality against conformity. His gradual awakening and subsequent downfall make for an emotional and philosophical journey. "1984" serves as a stern reminder that the abuse of power and the suppression of truth can lead to a nightmarish world devoid of freedom. This novel is not just a story but a wake-up call, and a book that everyone should read at least once in their lifetime.

Reseña en Español:

"1984" de George Orwell es una obra maestra distópica que profundiza en los peligros de los regímenes totalitarios y la vigilancia omnipresente. Escrita en 1949, la novela ha permanecido inquietantemente relevante en su representación de una sociedad donde el Gran Hermano siempre está observando. La desolación del mundo de Orwell—donde el lenguaje se manipula, la historia se revisa y el individualismo se aplasta—es una escalofriante advertencia para la sociedad moderna. Aunque el estilo de escritura es sencillo, las implicaciones son complejas y estimulantes para la reflexión. Winston Smith, el protagonista, encarna la lucha de la individualidad contra la conformidad. Su despertar gradual y su posterior caída constituyen un viaje emocional y filosófico. "1984" sirve como un severo recordatorio de que el abuso de poder y la supresión de la verdad pueden llevar a un mundo pesadillesco carente de libertad. Esta novela no es solo una historia, sino una llamada de atención, y un libro que todos deberían leer al menos una vez en su vida.

Формати відповідей

ChatGPT може генерувати відповіді у різних форматах, а також конвертувати один формат в інший.

У таблиці нижче зібрані основні формати.

ФОРМАТ	Опис	ПРИКЛАД
Звичайний текст	Тексти з використанням базового форматування.	Тексти з використанням базового формування .
Список з маркетами	Використання дефісів, астерисків або плюсів для елементів списку.	- Елемент 1 - Елемент 2 - Елемент 3
Нумерований список	Використання чисел для позначення послідовності.	1. Елемент А 2. Елемент В 3. Елемент С
JSON (англ. JavaScript Object Notation)	Структурований формат даних.	{ "Ключ 1": "значення 1", "Ключ 2": "значення 2" }
HTML (англ. HyperText Markup Language)	Мова розмітки для вебсайтів.	<p>Це параграф.</p> Елемент списку 1 Елемент списку 2
Markdown	Текстовий формат, що дозволяє формувати текст за допомогою спеціального синтаксису.	### Заголовок **Жирний текст** *Курсивний текст* [Посилання] (http://www.example.com)
Таблиця	Дані у вигляді рядків та колонок.	Стовець 1 Стовець 2 Стовець 3 ----- ----- ----- Дані 1 Дані 2 Дані 3 Дані 4 Дані 5 Дані 6
Код	Відображення фрагментів коду з або без підсвічування синтаксису.	```python print("Привіт!")
CSV (англ. Comma-Separated Values)	Текстовий формат, який дозволяє зберігати структуровані дані у вигляді таблиці.	Колонка 1, колонка 2, колонка 3 Дані 1, дані 2, дані 3 Дані 4, дані 5, дані 6

The
Economist

Russian assets: freeze or seize?

Managing in a time of stagflation

Britain's growth problem

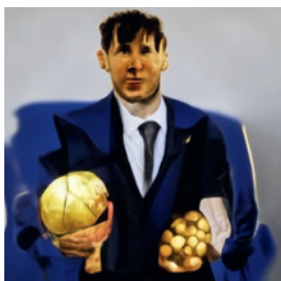
The short-sightedness epidemic

JUNE 11TH-17TH 2022

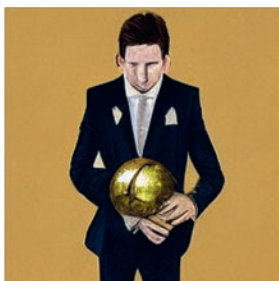
AI'S NEW FRONTIER



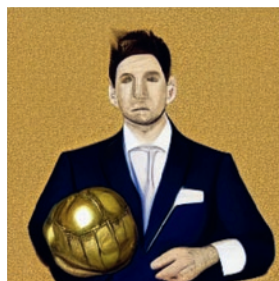
Обкладинка журналу *The Economist*, що створена за допомогою Midjourney
(джерело: *The Economist*).



v1 Лютий 2022



v2 Квітень 2022



v3 Липень 2022



v4 Листопад 2022



v5 Березень 2023



v5.1 Травень 2023



v5.1 Травень 2023
--style raw



v5.2 Червень 2023



v5.2 Червень 2023
--style raw



v6 Грудень 2023
--style raw



v6 Грудень 2023



v6 Грудень 2023
--s 50

Порівняння версій Midjourney.



/Imagine Картина вітрильника на воді у стилі олійного живопису.

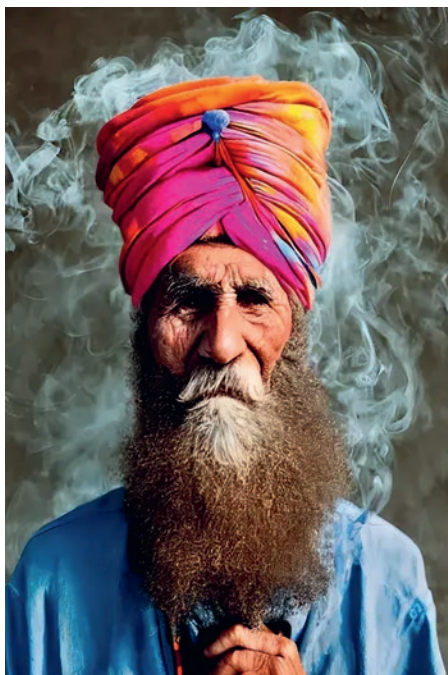


/imagine Портрет королеви Єлизавети II.

Stable Diffusion

Stable Diffusion — це латентна дифузійна модель, яка дозволяє конвертувати текстові запити в реалістичні та художні зображення. Вона була створена дослідниками з групи CompVis у Мюнхенському університеті ім. Людвіга-Максиміліана та Runway за підтримки Stability AI у серпні 2022 року. Особливість цієї моделі полягає у тому, що на відміну від DALL-E та Midjourney, її програмний код і ваги доступні публічно [21]. Вже у жовтні 2022 року компанія Stability AI залучила 101 млн доларів США від компаній Lightspeed Venture Partners і Coatue Management [22].

Модель Stable Diffusion навчалася на парах зображень і їх описів з LAION-5B — відкритого набору даних, зібраного з Common Crawl. Цей набір містить 5 млрд пар зображень і текстів, які були класифіковані за мовою, роздільною здатністю, наявністю водяних знаків та «естетичною» якістю. Цей набір даних було створено німецькою некомерційною організацією LAION за фінансової підтримки Stability AI.



Приклади генерації зображень за допомогою Stable Diffusion
(джерело: Stability.AI).

Версії моделі:

- 1.4, серпень 2022
- 1.5, жовтень 2022
- 2.0, листопад 2022
- 2.1, грудень 2022
- XL 1.0, липень 2023

Stable Diffusion не поступається Midjourney за якістю генерованих зображень, але надає додаткову можливість тонко налаштовувати модель під свої завдання, чудово «переносить» стиль та структуру з одного зображення на інше, а також має програмний інтерфейс доступу (API), чого наразі бракує Midjourney.

Google Imagen

Google Imagen — це інструмент для генерації зображень, конкурент DALL-E, Stable Diffusion та Midjourney. Однак, у той час, коли інші компанії випускають чергові версії своїх моделей, Google Imagen все ще перебуває у статусі «експериментальної розробки» і доступний обмеженій кількості користувачів. Для навчання використано набір даних LAION-400M. Щоб згенерувати готове зображення, Imagen спочатку створює зображення 64×64 пікселів, після чого воно масштабується до 256×256, і відтак до 1024×1024 пікселів.



Приклад згенерованого зображення
«Пітая у снігу, що носить пояс для карате».
Credits: Google.



Приклад згенерованого зображення «Мозок на ракетному кораблі, що прямує до Місяця»
Credits: Google.