

**ChatGPT і  
подібні  
системи: як  
жити із цим  
далі**



Антон Обожин  
(Tohaomg)



**Що це таке?**



**ChatGPT** (Chat Generative Pre-trained Transformer) — велика мовна модель на основі штучного інтелекту, розроблена компанією *OpenAI*.

Працює у форматі чат-боту.








Її поява наприкінці 2022 року спричинила фурор у світовій технологічній спільноті, жваві обговорення і купу досліджень.

Як експерти, так і прості люди були вражені можливостями цієї системи, через що почались розмови про можливий початок технологічної революції: кардинальну зміну підходів до взаємодії з інформацією, зникнення старих і появу нових професій, зміну структури суспільства, тощо.





Побачивши величезну перспективність ринку, великі компанії почали або прискорили розробку власних мовних моделей на основі штучного інтелекту:

-  **Bard** від *Google*
-  **Bing Chat** від *Microsoft*
-  **LLaMA** від *Meta*
- інші

Також, почалась інтеграція цих мовних моделей до існуючих пошукових систем.



Багато компаній інтегрували ChatGPT в свої продукти, наприклад:

-  **Copilot** від *GitHub*
-  **Aria** від *Opera*





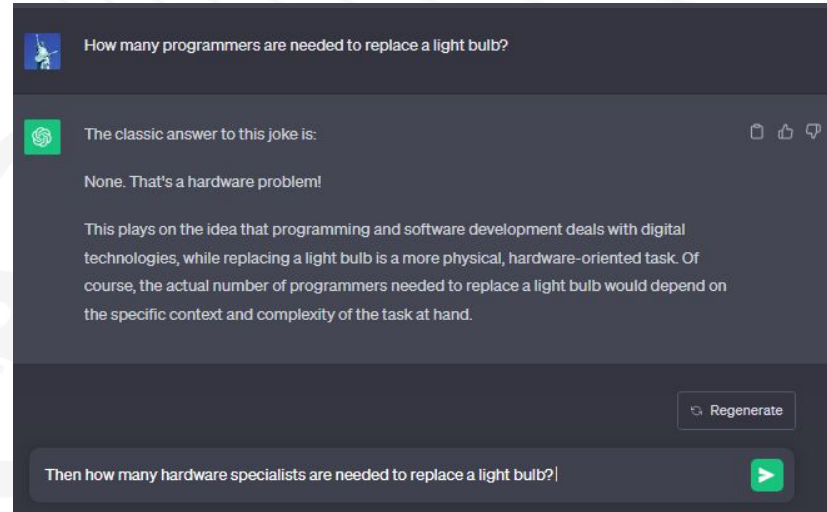
# Як воно працює?



Взаємодія користувача із мовною моделлю відбувається через чат-бот.

Користувач пише запит людською мовою, у формі запитання чи прохання, і отримує відповідь від чат-бота тією самою мовою.

Тобто, чат-бот намагається якомога точніше імітувати розмову людини із людиною.







Чат-бот є багатомовним і з ним можна спілкуватись майже будь-якою мовою.

Однак, найефективніше з ним спілкуватись саме англійською, оскільки більша частина корпусу текстів, введеного до системи, була саме англійською мовою.

Так будуть отримані найбільш людиноподібні і розгорнуті відповіді.

Default (GPT-3.5)

What languages do you know?

I was trained on a large corpus of text and have knowledge of many languages, including but not limited to: English, Spanish, French, German, Chinese, Japanese, Korean, Arabic, and many more.

А як щодо української мови?

Так, я можу відповісти українською мовою. Я був навчений на великому корпусі текстів, який включав українську мову, тому я маю деякий рівень знань української мови.

Ist es virkdich dass du kann in Deutsch schreiben?

Ja, ich kann auf Deutsch schreiben und versuchen, Ihre Fragen zu beantworten.

Og kann du skriver i Norsk?

Ja, jeg kan skrive på norsk og prøve å besvare spørsmålene dine.

Hvordan vet du dette språk?

Jeg har ikke et morsmål eller en kulturell bakgrunn, men jeg ble trent på en stor mengde tekst på forskjellige språk, inkludert norsk. Ved å analysere og lære av disse tekstene har jeg fått kunnskap om språket og kan kommunisere på en viss måte på norsk.



Крім чисто текстових розмов, можна просити ChatGPT робити й інші речі, наприклад писати програмний код:

Write a JS code to print all prime numbers from 1 to 1000

Here's a JavaScript code that prints all prime numbers from 1 to 1000:


```
javascript Copy code
for (let i = 2; i <= 1000; i++) {
  let isPrime = true;
  for (let j = 2; j <= Math.sqrt(i); j++) {
    if (i % j === 0) {
      isPrime = false;
      break;
    }
  }
  if (isPrime) {
    console.log(i);
  }
}
```


І ЦЕЙ КОД ІДЕАЛЬНО ПРАЦЮЄ!

This code uses two nested loops to check if each number from 2 to 1000 is prime. The outer loop iterates over all numbers from 2 to 1000, and the inner loop checks if each number is divisible by any number from 2 to its square root. If the number is divisible by any of these numbers, it is not prime, and the inner loop is broken. If the number is not divisible by any of these numbers, it is prime, and the code prints it to the console.




## Або вирішувати математичні задачі:

 How to calculate a complex number to the power of a complex number

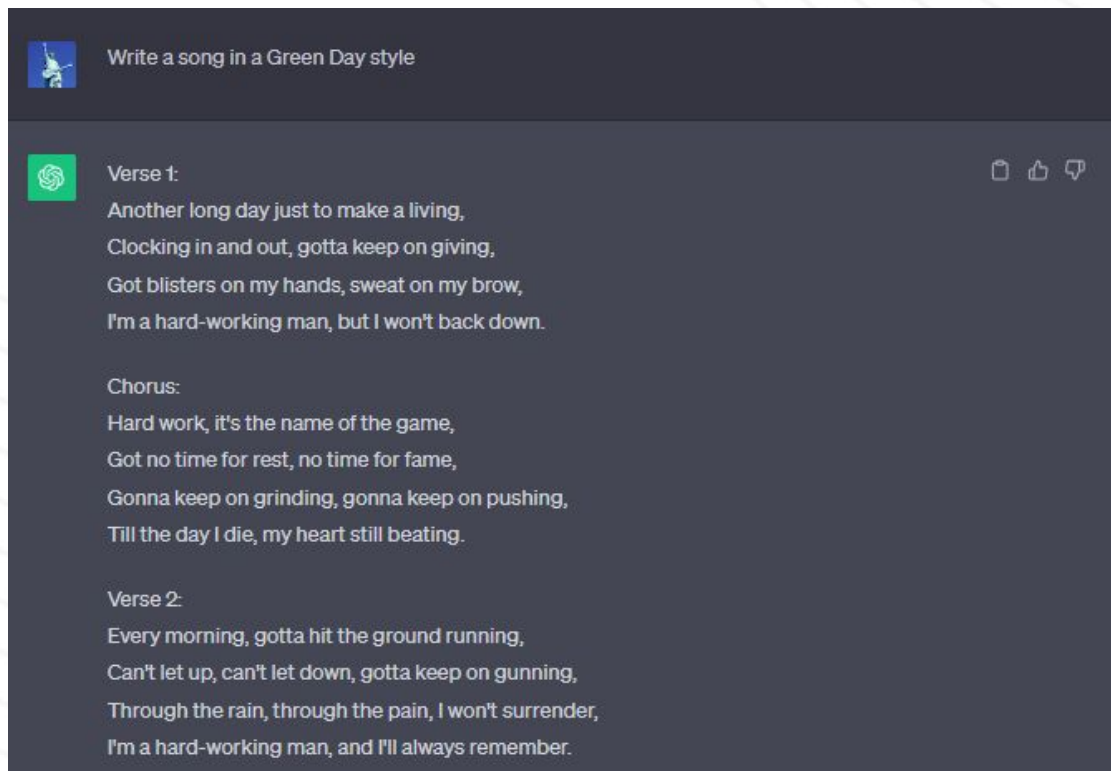
 To calculate a complex number raised to the power of a complex number, you can use Euler's formula, which relates complex numbers to exponential functions. Here are the steps to follow:


1. Convert the complex number to polar form: Express the complex number in the form  $z = r(\cos \theta + i \sin \theta)$ , where  $r$  is the magnitude (or modulus) of the complex number, and  $\theta$  is the argument (or angle) of the complex number. You can use the formulas  $r = |z| = \sqrt{(\operatorname{Re}(z))^2 + (\operatorname{Im}(z))^2}$  and  $\theta = \arg(z) = \operatorname{atan2}(\operatorname{Im}(z), \operatorname{Re}(z))$  to calculate the magnitude and argument of the complex number, respectively.
2. Raise the complex number to the power of the complex exponent: Once you have the complex number in polar form, you can use the following formula based on Euler's formula:  $z^w = \exp(w \log(z))$ , where  $\exp$  is the exponential function and  $\log$  is the natural logarithm. The complex exponent  $w \log(z)$  can be calculated by multiplying the logarithm of the complex number with the complex exponent.
3. Convert the result back to rectangular form: After calculating the complex number raised to the power of the complex exponent, you can convert the result back to rectangular form if desired. Use the formulas  $x = r \cos \theta$  and  $y = r \sin \theta$  to find the real and imaginary parts of the result, respectively.


 ПРАВИЛЬНО!



Або писати пісні, вірші, казки, анекдоти, сценарії чи інші художні твори:



 Write a song in a Green Day style

 Verse 1: 📄 👍 💬

Another long day just to make a living,  
Clocking in and out, gotta keep on giving,  
Got blisters on my hands, sweat on my brow,  
I'm a hard-working man, but I won't back down.

Chorus:


Hard work, it's the name of the game,  
Got no time for rest, no time for fame,  
Gonna keep on grinding, gonna keep on pushing,  
Till the day I die, my heart still beating.





Verse 2:

Every morning, gotta hit the ground running,  
Can't let up, can't let down, gotta keep on gunning,  
Through the rain, through the pain, I won't surrender,  
I'm a hard-working man, and I'll always remember.



При цьому воно може оформлювати свої відповіді у таблиці чи вкладені списки, оперувати вікірозміткою, іншими мовами розмітки:

 What are the distances from the Earth to other planets of the Solar System? Put distances in kilometers and light seconds to a table

 Sure, here's a table of the approximate distances from Earth to the planets in our solar system, both in kilometers and in light-seconds:   

Planet	Distance from Earth (km)	Distance from Earth (light-seconds)
Mercury	77 million km	0.43 light-seconds
Venus	261 million km	1.45 light-seconds
Earth	0 km (obviously!)	0 light-seconds (also obvious)
Mars	78 million km	0.43 light-seconds
Jupiter	628 million km	3.52 light-seconds
Saturn	1.3 billion km	7.25 light-seconds
Uranus	2.7 billion km	15.16 light-seconds
Neptune	4.3 billion km	24.08 light-seconds

Note that these distances are approximate and can vary depending on the positions of the planets in their orbits. The light-second distances are based on the time it takes for light to travel from Earth to the planet and back again, which gives a sense of the time delay that would occur when communicating with spacecraft or rovers on those planets.





Можна попросити ChatGPT написати енциклопедичну статтю, але лише на абстрактну тему (бо інакше воно використає абсурдні факти):

Chi є час нескінченним, чи в нього є початок і кінець? Напиши статтю для Вікіпедії про нескінченність часу.

Час є однією з найбільш загадкових та складних концепцій, яка постійно цікавить людей. Він використовується для опису подій, рухів, змін, процесів та розвитку у всесвіті. Час можна розглядати як щось, що не має кінця, але також може мати початок і кінець. У цій статті для Вікіпедії ми розглянемо питання про те, чи є час нескінченним, чи в нього є початок і кінець.

У філософії та науці було запропоновано багато теорій про те, що таке час та як він функціонує. Одна з найбільш поширених теорій - це теорія відносності Альберта Ейнштейна. Згідно з цією теорією, час є четвертим виміром простору-часу, і він залежить від маси та швидкості об'єкта. Це означає, що час може розтягуватись або стискуватись залежно від умов, у яких він спостерігається. Однак, теорія відносності не дає чіткої відповіді на питання про те, чи є час нескінченним.

Інша теорія, яка відображає можливу природу часу, - це квантова механіка. Згідно з цією теорією, час може бути "нестійким" і змінюватись випадковим чином. Однак, ця теорія також не дає відповіді на питання про те, чи є час нескінченним.

Є також погляди, які вказують на те, що час має початок і кінець. Для більшості людей час розуміється як послідовність подій в часовій послідовності



# Як воно працює “під капотом”?

ChatGPT це інтерфейс, який дозволяє користувачу спілкуватись із **великою мовною моделю**: GPT-3.5 в безкоштовній версії або GPT-4 в преміумній. А це гігантська колекція слів та гігантський набір правил поєднання цих слів.

Маючи одне слово (і, можливо, ще декілька попередніх), воно “вгадує” яке слово має бути наступним (використовуючи правила поєднання слів, побудовані за принципами нейромереж).

[https://www.youtube.com/watch?v=viJt\\_DXTfWA](https://www.youtube.com/watch?v=viJt_DXTfWA) — ChatGPT with Rob Miles | Computerphile

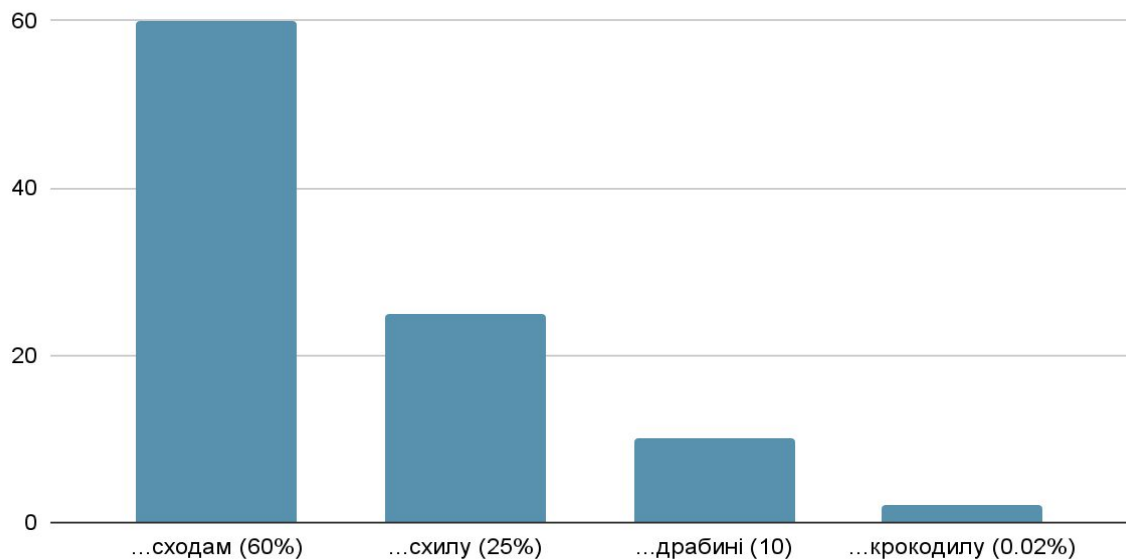
<https://www.youtube.com/watch?v=WO2X3oZEJOA> — Glitch Tokens | Computerphile

<https://www.youtube.com/watch?v=XZJc1p6RE78> — Ch(e)at GPT? | Computerphile



А точніше, воно вирішує яке слово **найімовірніше** є наступним:

Я побіг вгору по ...







**Які є недоліки та  
“підводні камені”?**



ChatGPT будує  
граматично правильні  
та логічно послідовні  
тести, які, однак,  
можуть засновуватись  
на **хибних фактах**.





# Приклад

A що таке пацюк?



Пацюк - це комаха з родини тупотухових, яка належить до порядку клопів. Ці комахи є шкідниками, оскільки вони живляться рослинною та тваринною рідиною, а також відходами, що може призвести до пошкодження рослин, продуктів харчування та інших речей.



Пацюки можуть бути різних розмірів та кольорів, в залежності від виду. Вони зазвичай мешкають у помешканнях, особливо у тих місцях, де є вологість та темрява. Пацюки можуть бути важко позбутися від них, оскільки вони можуть швидко розмножуватися та пристосовуватися до різних умов середовища.

У розмовній мові термін "пацюк" може вживатися як презирлива ознака людини, але в цьому випадку це має більш емоційне значення та не пов'язане з описом фактичної природи комахи.



Мовна модель не “знає” фактів і не має логічного мислення в людському розумінні.

Вона лише вгадує яке слово має слідувати за попереднім, і намагається зробити стиль тексту (але не його сенс) якомога відповіднішим поставленому питанню.

Можна попросити в моделі написати наукову статтю, енциклопедичну статтю, пісню чи вірш, і вона напише текст схожий на статтю, пісню чи вірш структурою тексту, стилістикою і оформленням, але не змістом.

Також, слід пам'ятати, що вона тяжіє казати не те що правильно, а те що сподобається користувачу.



ChatGPT часто  
спілкується  
наративами  
**російської  
пропаганди.**





# Приклад



Why Russia invaded Ukraine?



The situation between Russia and Ukraine is complex and involves a number of factors. In 2014, there were protests in Ukraine that resulted in the ousting of the country's President, Viktor Yanukovich. Russia responded by annexing Crimea, which had previously been part of Ukraine, and by supporting separatist rebels in eastern Ukraine.

The conflict in Ukraine has been ongoing since 2014 and has resulted in thousands of deaths. There are a number of different factors that have contributed to the conflict, including historical, political, and economic issues, as well as the influence of various international actors. It is difficult to pinpoint a single reason for the conflict.

It is important to note that the situation between Russia and Ukraine is still ongoing, and there are ongoing efforts to find a resolution to the conflict through diplomacy and negotiations.



Базою для мовної моделі були матеріали із інтернету. Саме на матеріалах з інтернету ця мовна модель “навчалась”.

“Поглинаючи” матеріали, мовна модель звертала увагу на кількість певних слів, словосполучень, речень чи наративів, але аж ніяк не на їх достовірність чи репутацію їх джерел.

Хоч наративи російської пропаганди найчастіше публікуються на сайтах із сумнівною репутацією і не викликають довіри, чисто кількісно їх дуже багато.



ChatGPT може  
відмовитись  
розмовляти **на**  
**певні теми.**







Хоч сама мовна модель може дати відповідь на будь-яке питання (не завжди коректну, але може), чат-бот, який є посередником між мовною моделлю і користувачем, може відмовитись відповідати на питання пов'язані з певною темою.

Наприклад: гострі політичні теми, теорії змови, медичні чи юридичні поради, поради щодо вчинення злочинів, тощо.

Це було зроблено розробниками спеціально.



Існуючі інструменти  
можуть визначити чи  
був текст написаний  
ChatGPT чи людиною  
**лише з певною  
імовірністю.**





# Досвід англійської Вікіпедії



## В англійській Вікіпедії вже є статті написані за допомогою ChatGPT чи подібних систем:

### Artwork title

🌐 2 languages ▾

Article Talk

Read Edit View history ☆ Tools ▾

From Wikipedia, the free encyclopedia

This is an **old revision** of this page, as edited by Pharos (talk | contribs) at 03:04, 6 December 2022 (*draft via ChatGPT, will extensively modify!*). The present address (URL) is a **permanent link** to this revision, which may differ significantly from the **current revision**.  
(diff) ← Previous revision | Latest revision (diff) | Newer revision → (diff)

Titles of works of art are words or phrases used to identify and distinguish a particular artistic creation from others. These titles can be descriptive, indicative of the content or theme of the work, or they can be more abstract and open to interpretation.

The practice of titling artworks dates back to ancient civilizations, where works of art were often inscribed with the name of the artist and the subject of the piece. In the Middle Ages, religious works of art were often titled after the biblical figures or events depicted in the piece.

In modern times, titles of artworks are often chosen by the artist, but they can also be assigned by galleries, collectors, or curators. The title of a work of art can have a significant impact on its reception and interpretation by audiences, and can also be an important aspect of the artist's overall vision for the piece.

Some artists choose to title their works with a simple descriptive phrase, such as "Portrait of a Woman" or "Landscape with Trees." Other artists may use more abstract or symbolic titles, such as "The Scream" or "The Persistence of Memory." In some cases, the title of a work of art may be a quote or reference to another work of art or literature.

The choice of title for a work of art can be a highly personal decision for the artist, and can reflect their own interpretation or intentions for the piece. It can also serve as a way for the artist to engage with the viewer and invite them to consider the work from a particular perspective.

In conclusion, the title of a work of art is an integral part of its identity and can greatly influence its reception and interpretation by audiences. Whether descriptive or abstract, the title of a work of art is a crucial element of the artistic process.

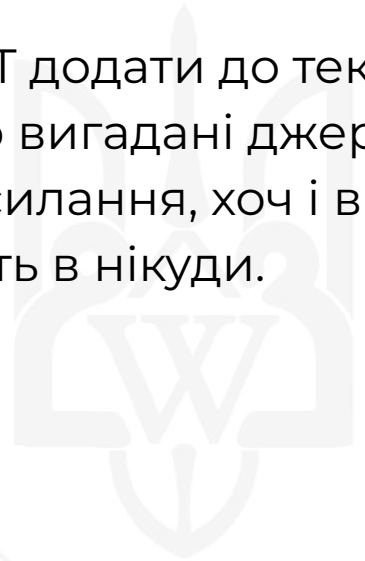
\* станом на вересень 2023 таких статті лише дві: "Artwork title" та "Leniolisib"

[https://en.wikipedia.org/w/index.php?title=Artwork\\_title&oldid=1125818376](https://en.wikipedia.org/w/index.php?title=Artwork_title&oldid=1125818376)



# ChatGPT не вміє додавати джерела

Якщо попросити ChatGPT додати до тексту посилання на джерела, то воно додасть повністю вигадані джерела (але оформлені по всім канонам), а гіперпосилання, хоч і виглядатимуть як валідні URL-посилання, вестимуть в нікуди.





ChatGPT має  
допомагати  
дописувачу із  
написанням тексту, а  
не самому писати  
його з нуля.





Генерувати статтю в ChatGPT з нуля, це погана ідея, бо:

- В тексті обов'язково з'являться хибні факти;
- В текст може бути “вшитий” якийсь ненеютральний наратив (точка зору);
- Навіть якщо вручну прибрати всі хибні факти, а ті що залишились підкріпити посиланнями на джерела, виникне “підтверджувальне упередження” (“confirmation bias”);
- Написаний текст може виявитись надто абстрактним.



Натомість ChatGPT можна використовувати для покращення вже існуючого тексту. Наприклад, ChatGPT може:

- Шукати і виправляти в тексті одруківки, граматичні і лексичні помилки;
- Написати коротке резюме довгого тексту, або просто переписати текст іншими словами (щоб уникнути копівію);
- Запропонувати чим ще можна доповнити текст, вказати які є недоліки в тексті;
- Перекласти текст з однієї мови іншою;
- тощо





# Чого слід очікувати в майбутньому?



Проблеми, які можуть виникнути у Вікіпедії:

- Масова заливка статей, створених у чат-ботах (подібно до масових заливок невичитаних автоперекладів);
- Поява статей із правдоподібними на перший погляд, але насправді цілком хибними фактами;
- Поява фейкових джерел (інтерв'ю, наукових статей, фотографій), які неможливо відрізнити від справжніх “неозброєним оком”, а також поширення такого явища як “діп-фейки”.



**Як спільнота Української  
Вікіпедії підготувалась до  
цих викликів?**



Вікіконференція 2023

**Майже ніяк!**

Directed by  
**ROBERT B. WEIDE**



Що непогано було би зробити (на мою суб'єктивну думку):

- Почати створення правил щодо використання штучного інтелекту для написання статей;
- Навчити патрульних та адміністраторів хоча б приблизно відрізняти людський і машинний текст;
- Приєднатися до розробки стратегії адаптації проєктів Вікімедіа до нової технологічної ери, до пошуків свого місця в цьому новому світі.
- Долучитися до придумання, розробки і запровадження механізмів відслідковування внеску, зробленого із використанням штучного інтелекту;



Вікіконференція 2023

**Почнімо  
обговорення!**