SYNERGIA

Exploring Artificial Intelligence (AI) and ChatGPT Safely

Eli Kliejunas and Dr. Sarah Appleton, Synergia

2024-02

The Why

- AI chatbots are extremely helpful when you know how to work with them safely and effectively
- Highly useful AI assistants → less time on admin work → more time on intellectually stimulating, valuable work that directly supports whaiora wellbeing
- I believe AI chatbots can have a massive democratizing effect on society
  - Accessibility of information
  - Language translation and support
  - Educational support
  - Assistance for people with disabilities with speech-to-text and text-to-speech capabilities
  - Creativity and innovation
  - Economic opportunities
- **There is also great risk involved with the irresponsible use of these technologies.** All the more reason to learn about the risks and how to mitigate them!

Intentions for this session

- Healthy balance of information sharing and korero
    - The slides will be available for your reference after the session
- Demystifying AI chatbots → you don’t need to be a tech wiz to use them effectively!
- Emphasizing that AI chatbots are evolving RAPIDLY
    - AI chatbots have limitations, risks, and biases
    - Those limitations, risks, and biases are being reduced over time with better ‘training’ from users and providers

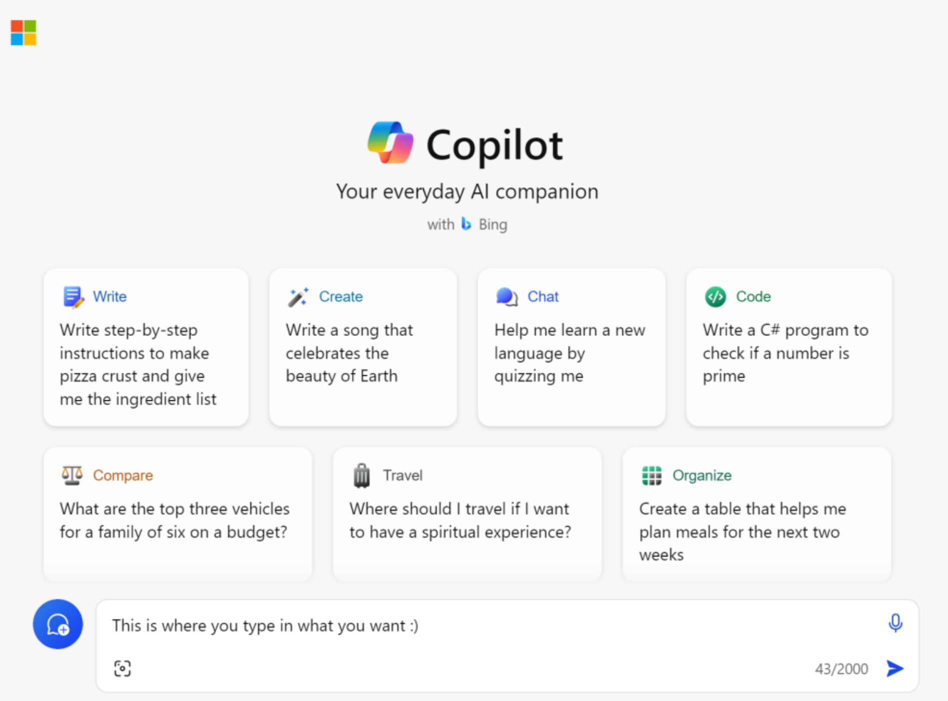
Overview

- Demystifying AI chatbot interfaces
- Ethical considerations
- What is generative AI? How do Large Language Models work (broadly speaking)?
- Responsible usage: what data can be safely input?
- Where is generative AI currently helpful, and where is it not?
- Tips for effective usage: an introduction to ‘Prompt Engineering’

Quick notes before we start

- Lots of options: ChatGPT, Bing Chat, Google Gemini, Claude, etc.
  - Each with their own strengths and weaknesses
- Paid vs. free versions
- Individual vs. enterprise AI
- ‘Multimodal’ inputs: text, images, video
- Right tool for the right job
- The field is changing rapidly
  - Some details explained today may be out of date next week

It's easy to use AI chatbots!



Copilot

Your everyday AI companion

with Bing

Write
Write step-by-step instructions to make pizza crust and give me the ingredient list

Create
Write a song that celebrates the beauty of Earth

Chat
Help me learn a new language by quizzing me

Code
Write a C# program to check if a number is prime

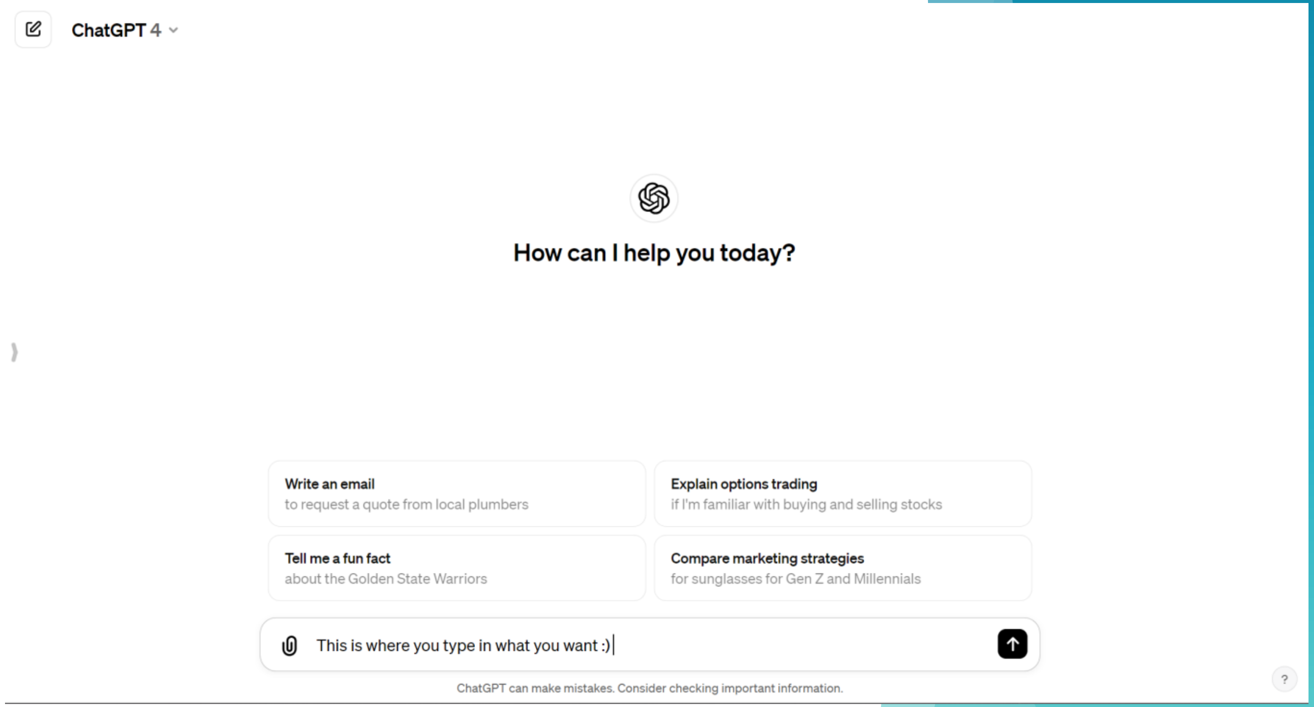
Compare
What are the top three vehicles for a family of six on a budget?

Travel
Where should I travel if I want to have a spiritual experience?

Organize
Create a table that helps me plan meals for the next two weeks

This is where you type in what you want :)

43/2000



ChatGPT 4

How can I help you today?

Write an email
to request a quote from local plumbers

Explain options trading
if I'm familiar with buying and selling stocks

Tell me a fun fact
about the Golden State Warriors

Compare marketing strategies
for sunglasses for Gen Z and Millennials

This is where you type in what you want :)

ChatGPT can make mistakes. Consider checking important information.

Ethical questions around AI chatbots

- Is it stealing?
  - Consent and compensation
- Does it perpetuate harmful biases and stereotypes?
- Could it lead to artificial general intelligence (AGI) misaligned with our goals?
- Do the benefits outweigh the risks?
  - Ex: ‘The game has changed.’ AI triumphs at protein folding. → AI has fast-tracked biomedical innovation by solving the ‘protein folding problem’.
  - Ex: Can AI Catch What Doctors Miss? | Eric Topol | TED →AI starting to detect and diagnose diseases more accurately than humans in some cases
  - Ex: ‘GraphCast: AI model for faster and more accurate global weather forecasting’ → AI-generated early detection of severe weather events could save lives and mitigate damages

What is ChatGPT?
What is ‘Generative AI’?

What is Generative AI?

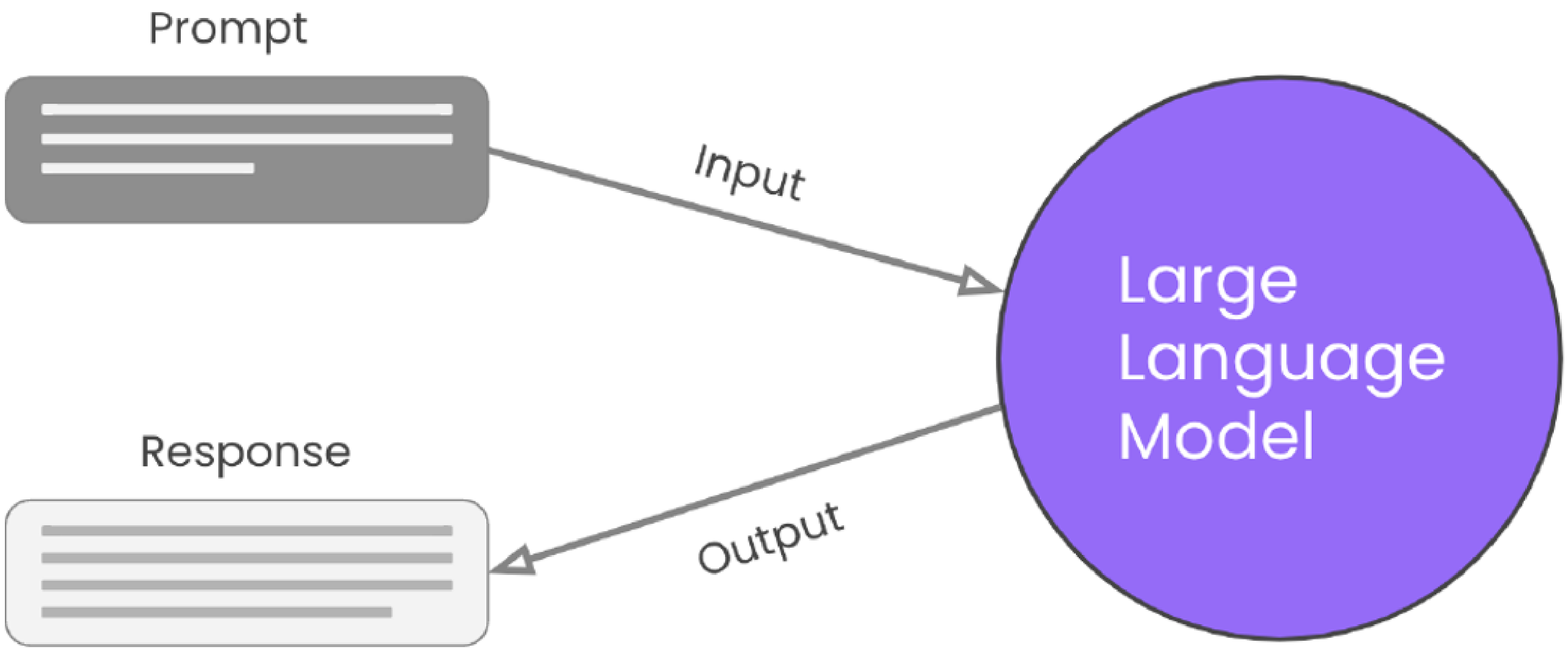
- **Discriminative AI**: from *data* to *description*

→ "Kitten"

- **Generative AI:** from *description* to *data*

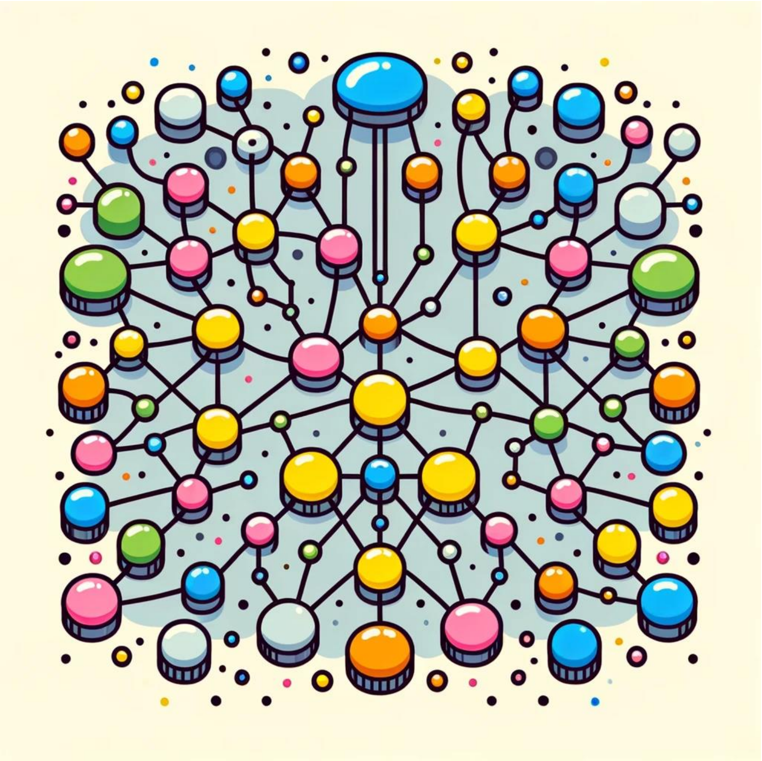
"Kitten" →

How does an AI chatbot/Large Language Model work?



What's going on underneath the hood of an AI chatbot?

- Billions of statistical connections are being made between your prompt and the chatbot's training text <u>to predict what the most likely word to come next is</u>.



Prompt in DALL-E image generator: 'A simplified illustration of a neural network'

What's going on underneath the hood of an AI chatbot?

For more details on how AI chatbots work, here is a great reference article from the Guardian featuring a visual explainer of how large language models work — **Google 'How AI Chatbots like ChatGPT or Bard work – visual explainer'**: https://www.theguardian.com/technology/ng-interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer

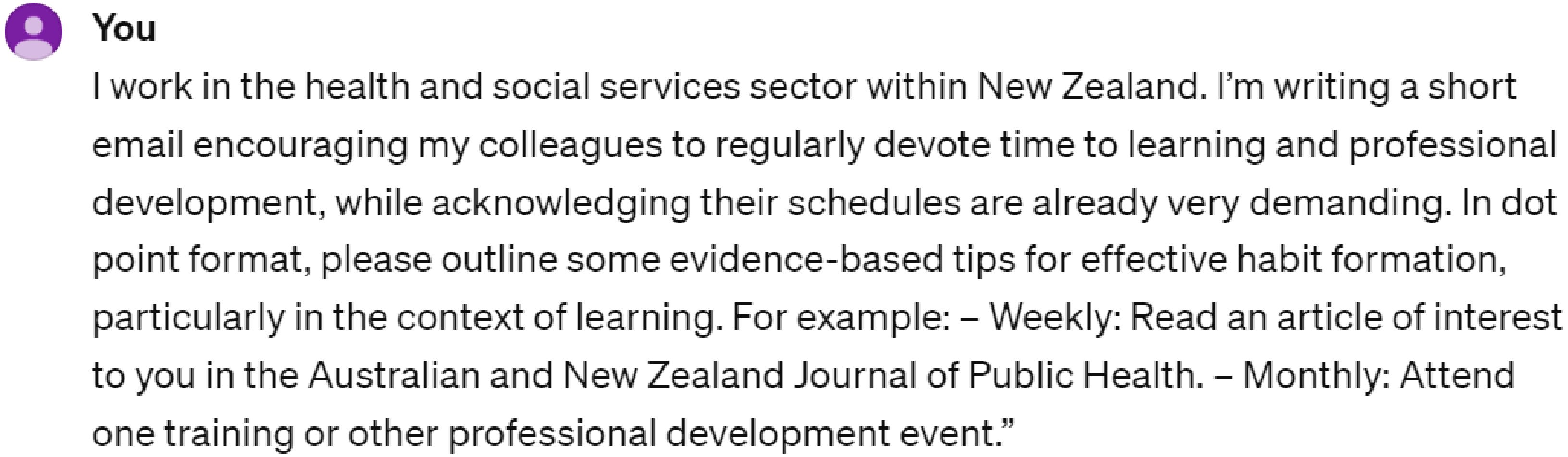
Introducing ‘prompt engineering’, or how to speak to AI chatbots effectively

Prompt Components

- Topic and context
- Instructions
- Examples (optional but helpful)

"I work in the health and social services sector in New Zealand. I'm writing a short email encouraging my colleagues to regularly devote time to learning and development, while acknowledging their schedules are already very demanding. In dot point format, please outline some evidence-based tips for effective habit formation, particularly in the context of learning. For example: – Weekly: Read an article of interest to you in the Australian and New Zealand Journal of Public Health. – Monthly: Attend one training or other professional development event."

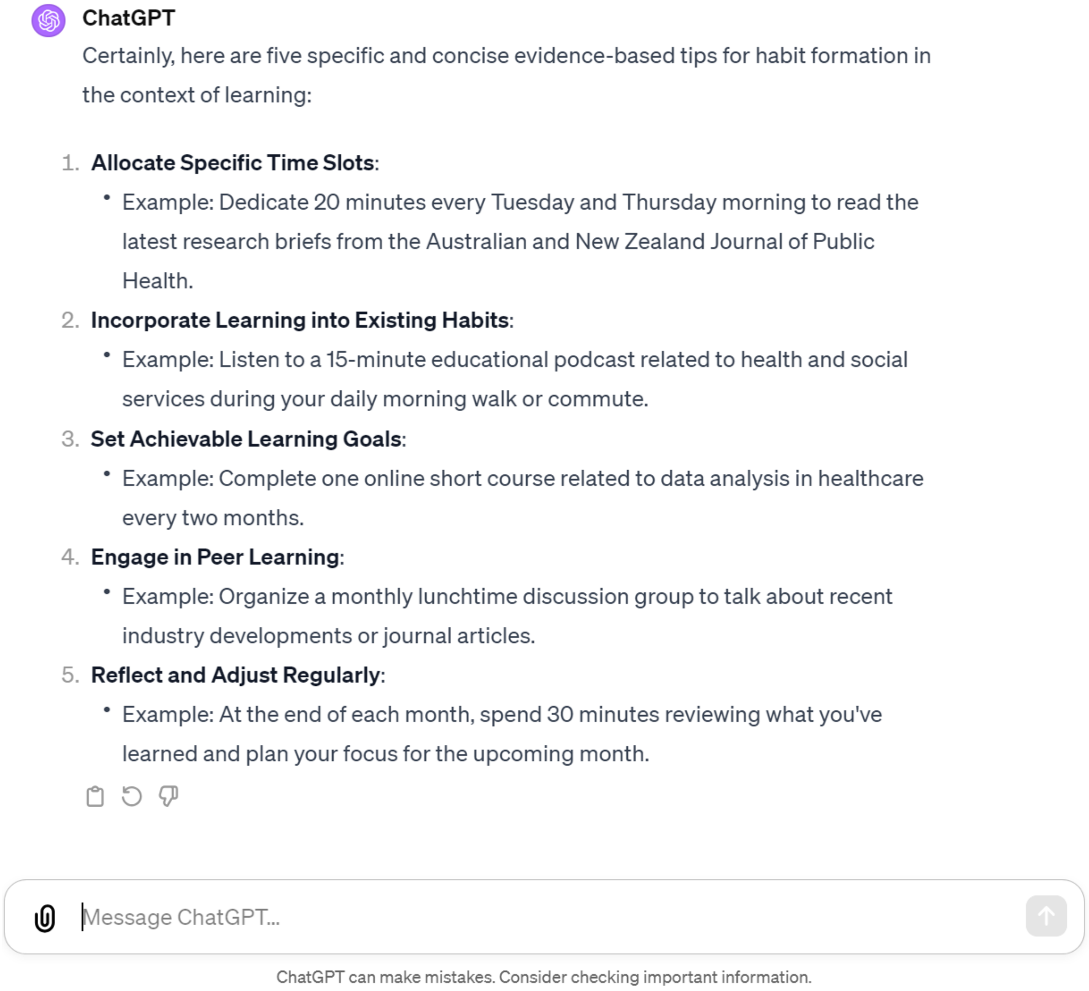
Prompt to AI Chatbot



You

I work in the health and social services sector within New Zealand. I'm writing a short email encouraging my colleagues to regularly devote time to learning and professional development, while acknowledging their schedules are already very demanding. In dot point format, please outline some evidence-based tips for effective habit formation, particularly in the context of learning. For example: – Weekly: Read an article of interest to you in the Australian and New Zealand Journal of Public Health. – Monthly: Attend one training or other professional development event."

Response from AI Chatbot



ChatGPT

Certainly, here are five specific and concise evidence-based tips for habit formation in the context of learning:

1. **Allocate Specific Time Slots**:
    - Example: Dedicate 20 minutes every Tuesday and Thursday morning to read the latest research briefs from the Australian and New Zealand Journal of Public Health.
2. **Incorporate Learning into Existing Habits**:
    - Example: Listen to a 15-minute educational podcast related to health and social services during your daily morning walk or commute.
3. **Set Achievable Learning Goals**:
    - Example: Complete one online short course related to data analysis in healthcare every two months.
4. **Engage in Peer Learning**:
    - Example: Organize a monthly lunchtime discussion group to talk about recent industry developments or journal articles.
5. **Reflect and Adjust Regularly**:
    - Example: At the end of each month, spend 30 minutes reviewing what you've learned and plan your focus for the upcoming month.

ChatGPT can make mistakes. Consider checking important information.

Basic Tips

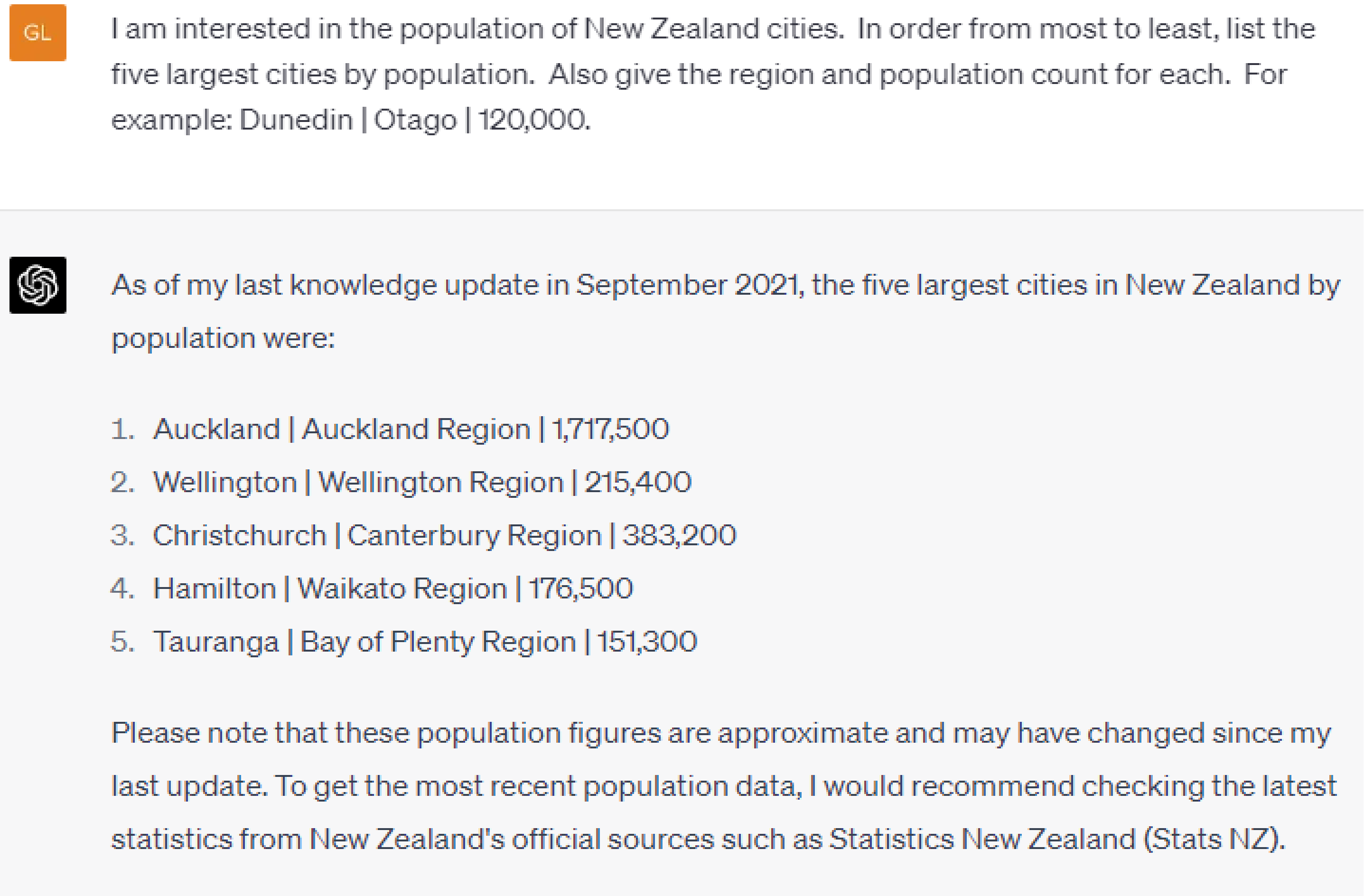
- Be clear and **specific**
- Provide relevant context, but
- Be concise and stick to what’s relevant
- Use correct grammar and spelling
- Correct, refine, rephrase, and try out different ways

Prompt Components

- Topic and context
- Instructions
- Examples (optional but helpful)

"I am interested in the population of New Zealand cities. In order from most to least, list the five largest cities by population. Also give the region and population count for each. For example: Dunedin | Otago |120,000."

Prompt Components - Response



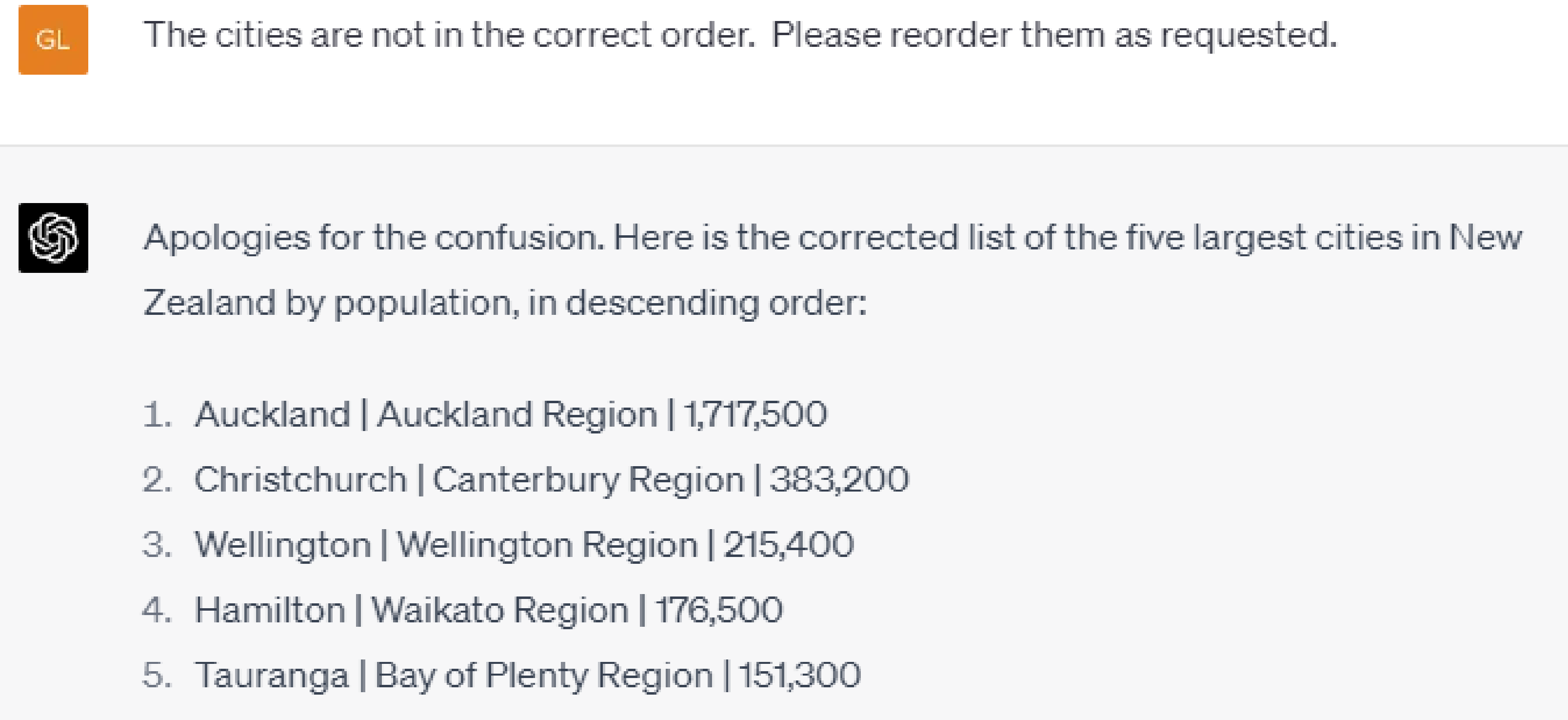
I am interested in the population of New Zealand cities. In order from most to least, list the five largest cities by population. Also give the region and population count for each. For example: Dunedin | Otago | 120,000.

As of my last knowledge update in September 2021, the five largest cities in New Zealand by population were:

1. Auckland | Auckland Region | 1,717,500
2. Wellington | Wellington Region | 215,400
3. Christchurch | Canterbury Region | 383,200
4. Hamilton | Waikato Region | 176,500
5. Tauranga | Bay of Plenty Region | 151,300

Please note that these population figures are approximate and may have changed since my last update. To get the most recent population data, I would recommend checking the latest statistics from New Zealand's official sources such as Statistics New Zealand (Stats NZ).

Correcting and Refining the Response



The cities are not in the correct order. Please reorder them as requested.

Apologies for the confusion. Here is the corrected list of the five largest cities in New Zealand by population, in descending order:

1. Auckland | Auckland Region | 1,717,500
2. Christchurch | Canterbury Region | 383,200
3. Wellington | Wellington Region | 215,400
4. Hamilton | Waikato Region | 176,500
5. Tauranga | Bay of Plenty Region | 151,300

Writing ‘safe’ prompts to AI chatbots

Principles to keep in mind with AI chatbot prompts

- Ownership and Consent:
    - Whose data are you inputting?
    - Do you have that person(s) consent to input their data into an AI chatbot?
- Confidentiality and Identifiability
    - Are you inputting private information that can be used to identify individuals or entities?
- Data Minimisation
    - To reduce the risk of exposing sensitive information, input only the minimum amount of data necessary

Risk Spectrum — Quantitative Data

Risk level	Data description		Examples
Low	Intended for public disclosure	✓	Open-source literature and data freely available on the internet
Medium	De-individualised information	✗	Encrypted NHI
High	Potentially identifying information	✗	NHI, unique identifier, DOB, specific demographic information
Very high	Sensitive information and/or identifying information	✗	Name, address, NHI, banking information (credit/debit card, account), passport, driver’s licence

Reference material for further details

- Privacy Commissioner's (Te Mana Mātāpono Matatapu) Guidelines on AI and Information Privacy Principles (IPP): https://www.privacy.org.nz/publications/guidance-resources/ai/
- NZ Data Protection and Use (DPUP) Policy: https://www.digital.govt.nz/standards-and-guidance/privacy-security-and-risk/privacy/data-protection-and-use-policy-dpup/
- Microdata output guide from StatsNZ: https://www.stats.govt.nz/assets/Methods/Microdata-Output-Guide-2020-v5-Sept22update.pdf

Risk Spectrum – Qualitative Data

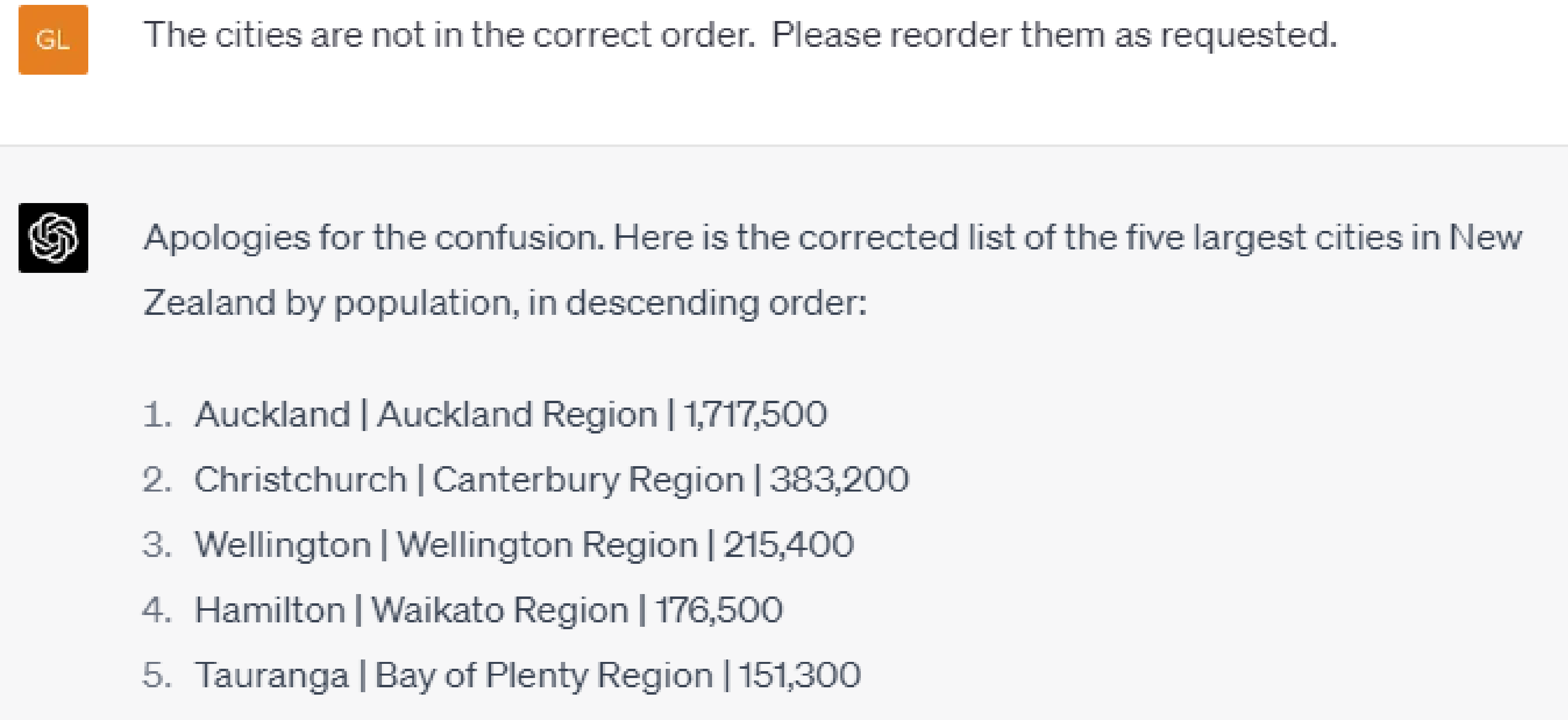
Risk level	Data Type	Description	Examples
Low	Public opinion narratives	General opinions and narratives shared in public forums or social media that do not contain personal or sensitive information ✓	Public blog posts, general discussions in open forums
Low-Medium	Summarised customer/client experiences	Collective summaries of customer experiences that exclude personal details. ?	Overarching themes from customer feedback, summaries of focus group discussions.
Medium	Anonymised individual feedback	Detailed feedback on non-sensitive topics from individuals that have been anonymised ?	Anonymised user experience stories, personal testimonials without identifiers, anonymised case studies
High	Sensitive personal narratives	Personal stories or feedback that involve sensitive topics but are shared in a way that the information remains non-identifiable ✗	Narratives about personal health experiences without revealing identity, stories of workplace issues shared anonymously
High	Confidential or Proprietary Narratives	Information that is not publicly available and could be damaging if disclosed, including narratives around sensitive business practices or personal life. ✗	Detailed accounts of proprietary business processes, personal diaries, privileged legal communications.

Scrutinising AI chatbot responses safely

LLM Output Considerations

- **Scrutinise LLM outputs carefully!**
  - Never use AI generated content as your final draft
  - Check primary source material when citing facts
  - Be on the lookout for content that perpetuates biases and/or harmful stereotypes

Correcting and Refining the Response



The cities are not in the correct order. Please reorder them as requested.

Apologies for the confusion. Here is the corrected list of the five largest cities in New Zealand by population, in descending order:

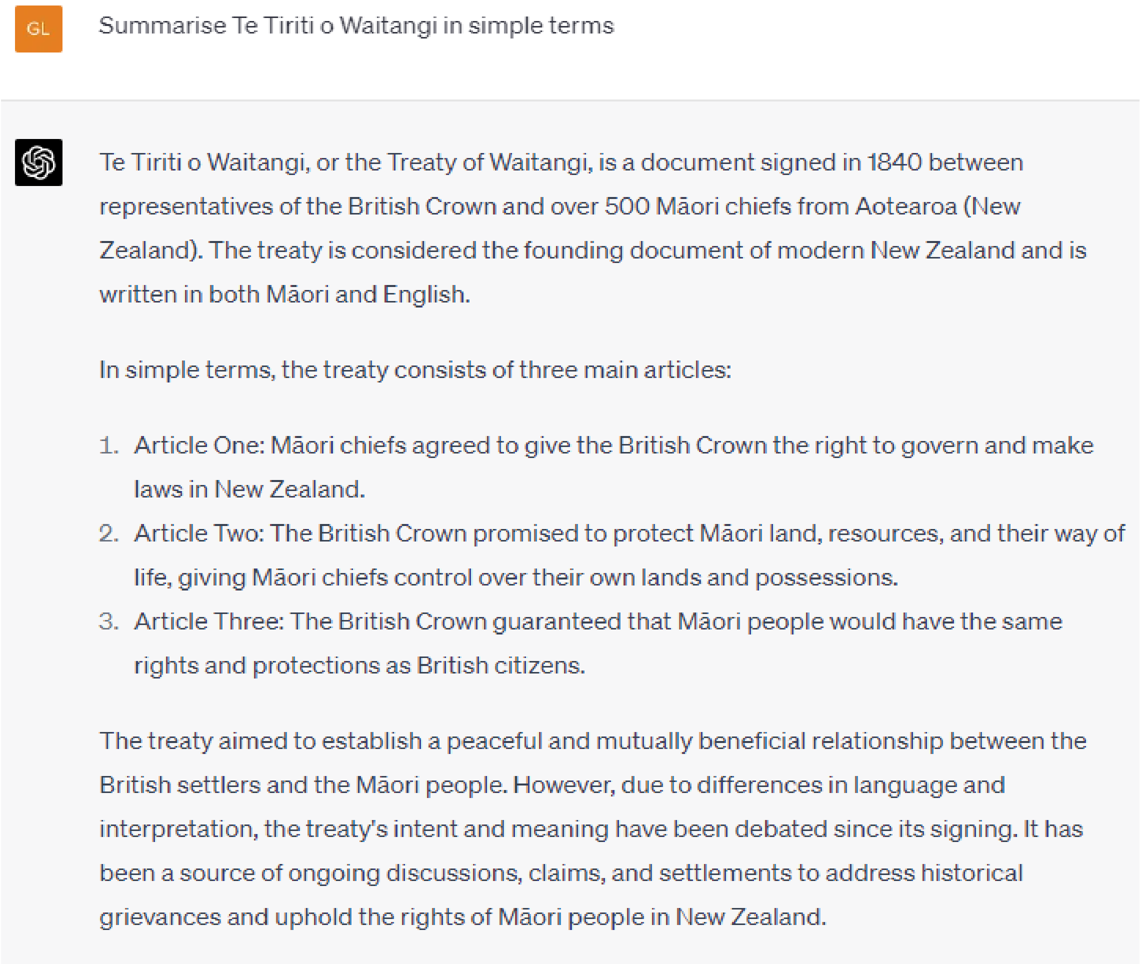
1. Auckland | Auckland Region | 1,717,500
2. Christchurch | Canterbury Region | 383,200
3. Wellington | Wellington Region | 215,400
4. Hamilton | Waikato Region | 176,500
5. Tauranga | Bay of Plenty Region | 151,300

Where Generative AI is Powerful

Where ChatGPT and Other LLMs Are Powerful

- Summarising text and code
- Transforming text and code
    - Ex: Change
- Generating text and code
- Explaining/distilling complex topics
- Technical help
    - Ex: How do I change American style dates in Excel into the much more sensible Day/Month/Year format?

Summarising text



Summarise Te Tiriti o Waitangi in simple terms

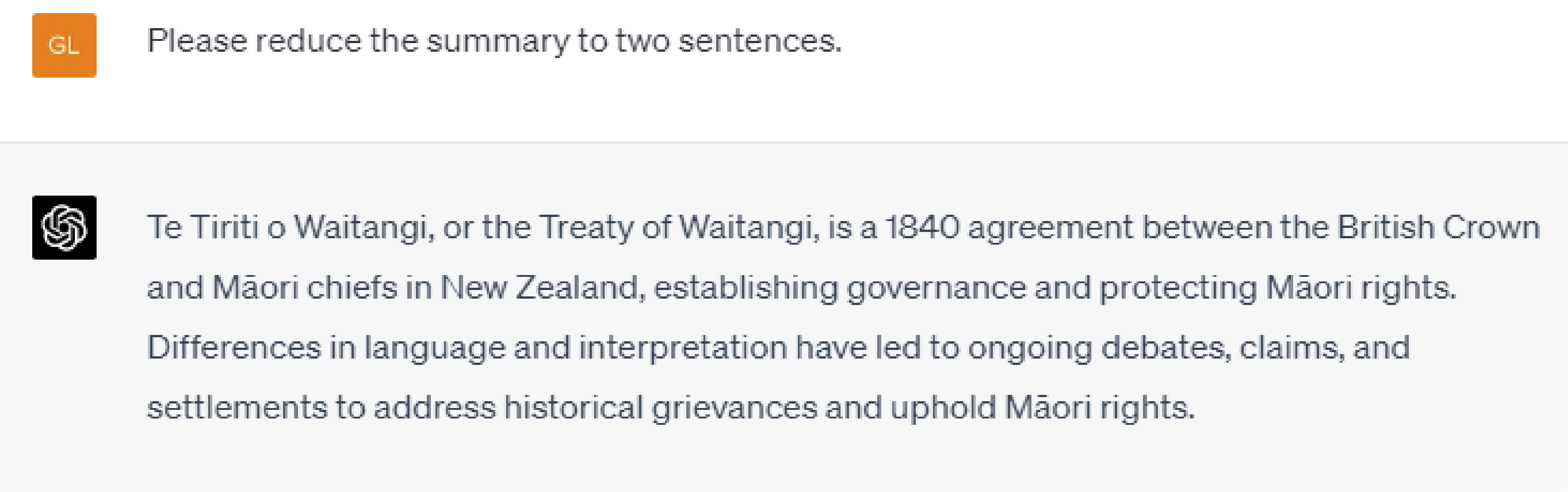
Te Tiriti o Waitangi, or the Treaty of Waitangi, is a document signed in 1840 between representatives of the British Crown and over 500 Māori chiefs from Aotearoa (New Zealand). The treaty is considered the founding document of modern New Zealand and is written in both Māori and English.

In simple terms, the treaty consists of three main articles:

1. Article One: Māori chiefs agreed to give the British Crown the right to govern and make laws in New Zealand.
2. Article Two: The British Crown promised to protect Māori land, resources, and their way of life, giving Māori chiefs control over their own lands and possessions.
3. Article Three: The British Crown guaranteed that Māori people would have the same rights and protections as British citizens.

The treaty aimed to establish a peaceful and mutually beneficial relationship between the British settlers and the Māori people. However, due to differences in language and interpretation, the treaty's intent and meaning have been debated since its signing. It has been a source of ongoing discussions, claims, and settlements to address historical grievances and uphold the rights of Māori people in New Zealand.

Summarising text

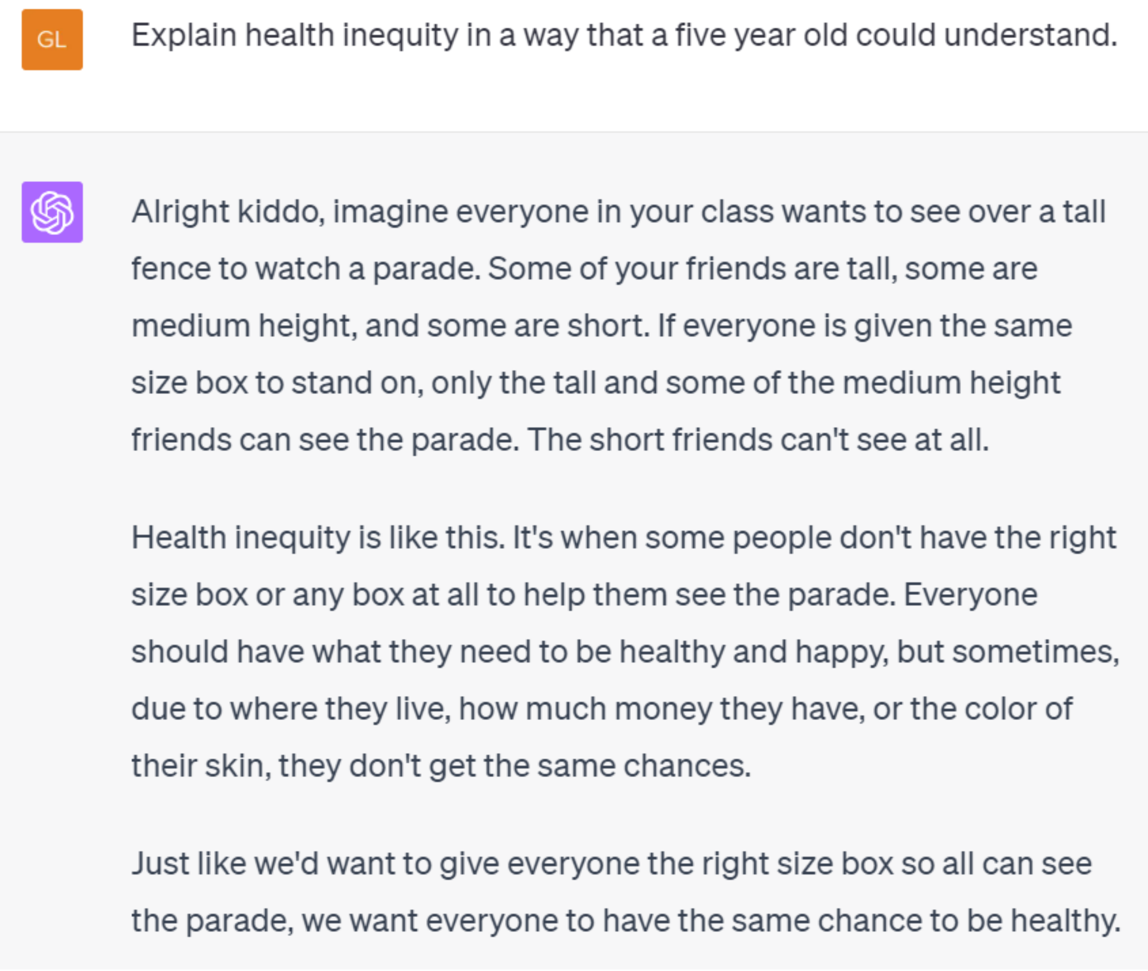


Please reduce the summary to two sentences.

Te Tiriti o Waitangi, or the Treaty of Waitangi, is a 1840 agreement between the British Crown and Māori chiefs in New Zealand, establishing governance and protecting Māori rights. Differences in language and interpretation have led to ongoing debates, claims, and settlements to address historical grievances and uphold Māori rights.

- ChatGPT remembers information from earlier in the conversation

Explaining complex ideas



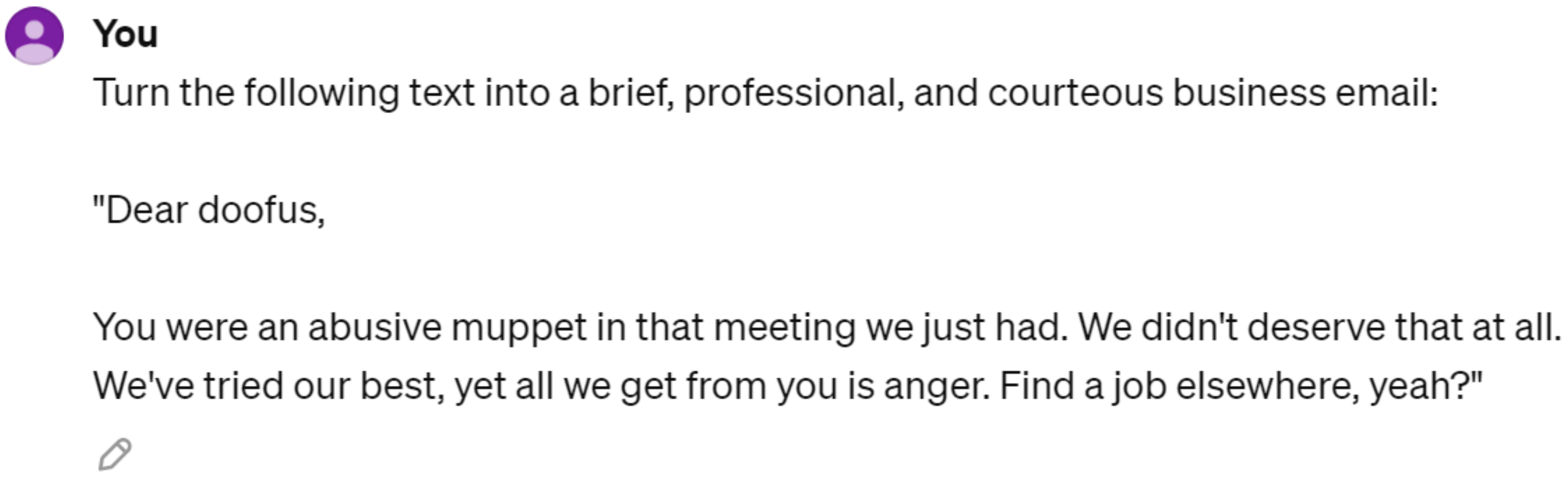
Explain health inequity in a way that a five year old could understand.

Alright kiddo, imagine everyone in your class wants to see over a tall fence to watch a parade. Some of your friends are tall, some are medium height, and some are short. If everyone is given the same size box to stand on, only the tall and some of the medium height friends can see the parade. The short friends can't see at all.

Health inequity is like this. It's when some people don't have the right size box or any box at all to help them see the parade. Everyone should have what they need to be healthy and happy, but sometimes, due to where they live, how much money they have, or the color of their skin, they don't get the same chances.

Just like we'd want to give everyone the right size box so all can see the parade, we want everyone to have the same chance to be healthy.

Transforming (One Style to Another)



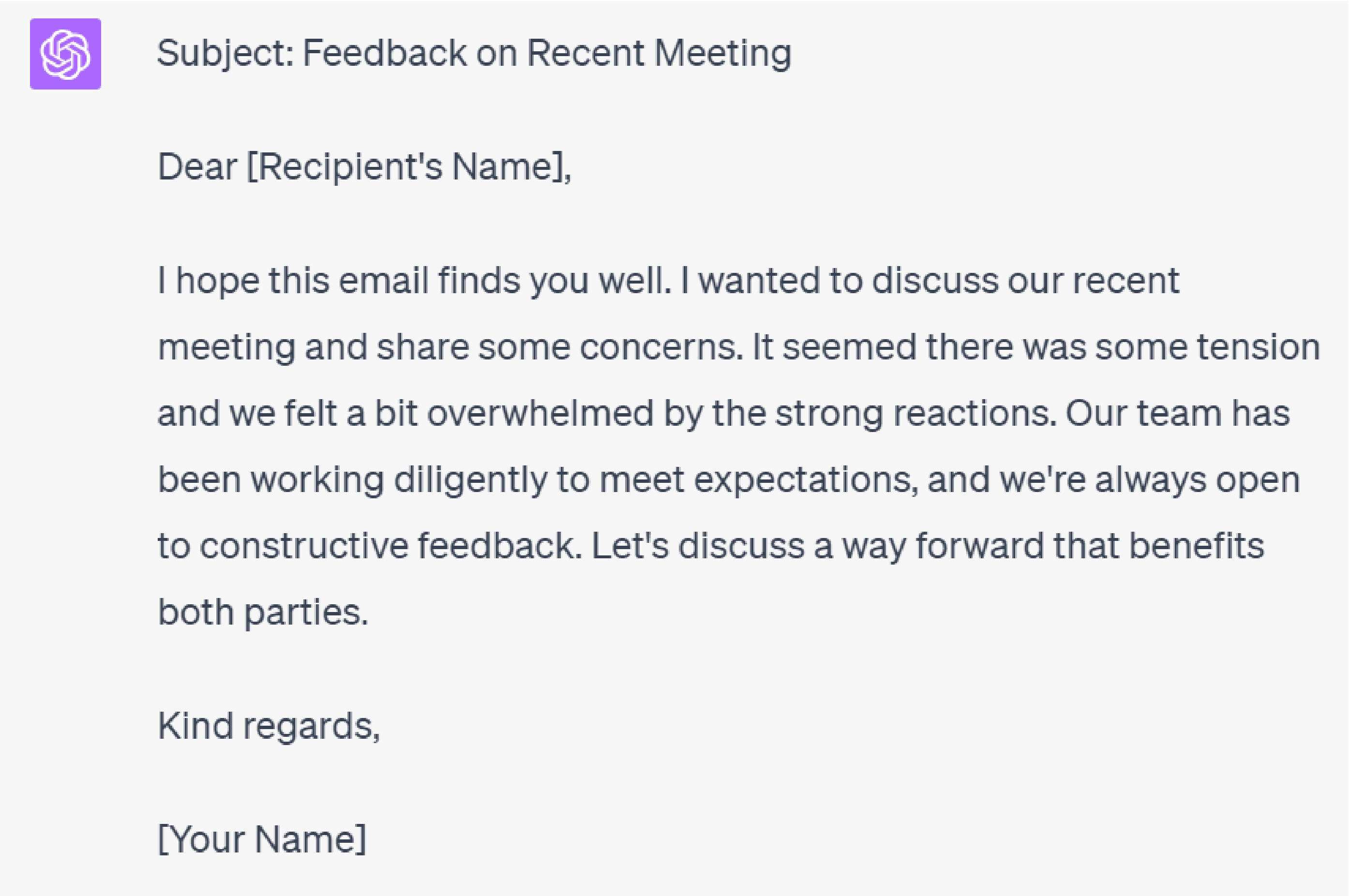
You

Turn the following text into a brief, professional, and courteous business email:

"Dear doofus,

You were an abusive muppet in that meeting we just had. We didn't deserve that at all. We've tried our best, yet all we get from you is anger. Find a job elsewhere, yeah?"

Transforming (One Style to Another)



Subject: Feedback on Recent Meeting

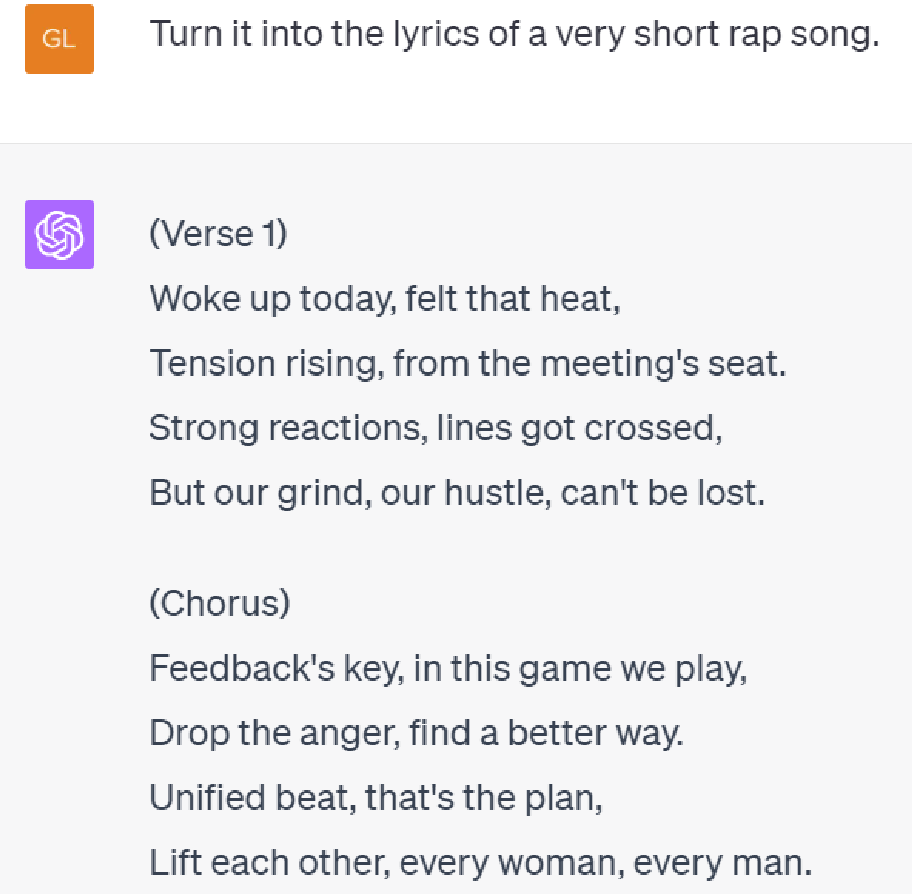
Dear [Recipient's Name],

I hope this email finds you well. I wanted to discuss our recent meeting and share some concerns. It seemed there was some tension and we felt a bit overwhelmed by the strong reactions. Our team has been working diligently to meet expectations, and we're always open to constructive feedback. Let's discuss a way forward that benefits both parties.

Kind regards,

[Your Name]

Transforming (One Style to Another)

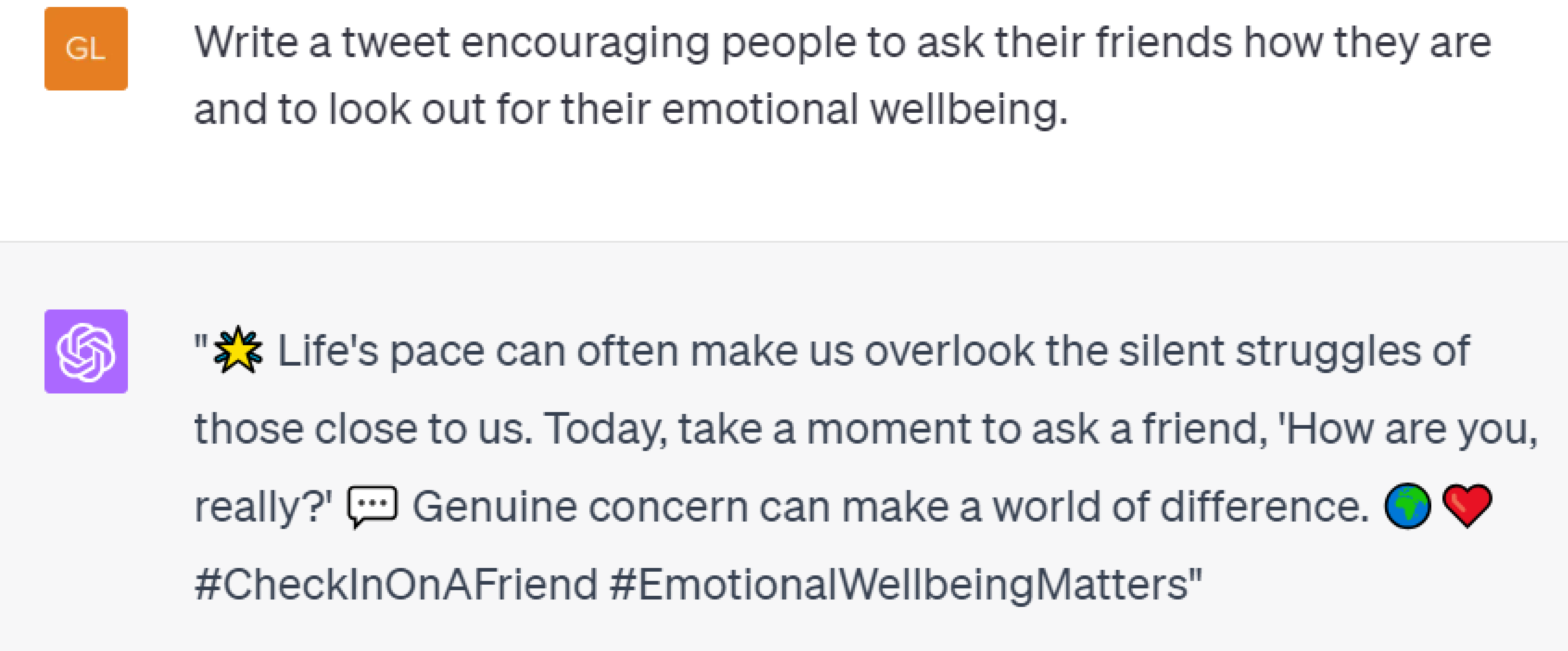


Turn it into the lyrics of a very short rap song.

(Verse 1)
Woke up today, felt that heat,
Tension rising, from the meeting's seat.
Strong reactions, lines got crossed,
But our grind, our hustle, can't be lost.

(Chorus)
Feedback's key, in this game we play,
Drop the anger, find a better way.
Unified beat, that's the plan,
Lift each other, every woman, every man.

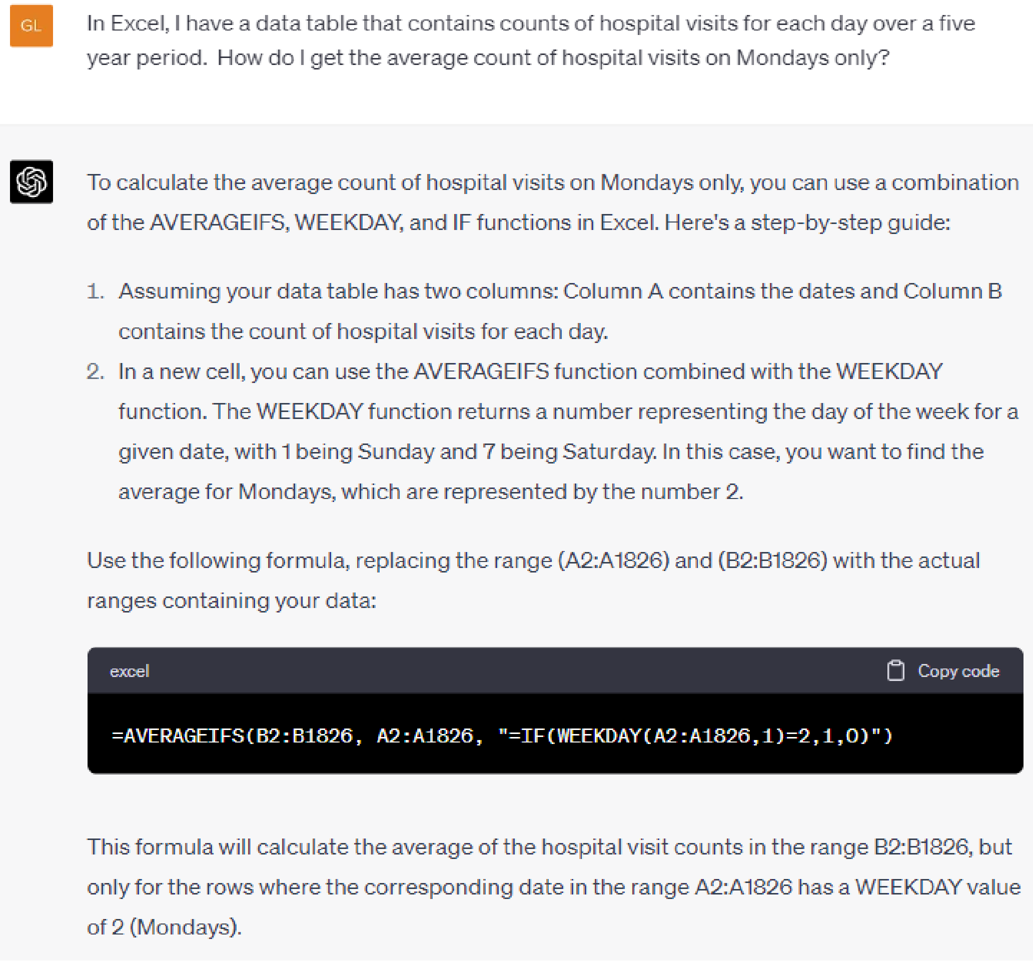
Generating text



Write a tweet encouraging people to ask their friends how they are and to look out for their emotional wellbeing.

"🌟 Life's pace can often make us overlook the silent struggles of those close to us. Today, take a moment to ask a friend, 'How are you, really?' 💬 Genuine concern can make a world of difference. 🌍 ❤️ #CheckInOnAFriend #EmotionalWellbeingMatters"

Technical Help



In Excel, I have a data table that contains counts of hospital visits for each day over a five year period. How do I get the average count of hospital visits on Mondays only?

To calculate the average count of hospital visits on Mondays only, you can use a combination of the AVERAGEIFS, WEEKDAY, and IF functions in Excel. Here's a step-by-step guide:

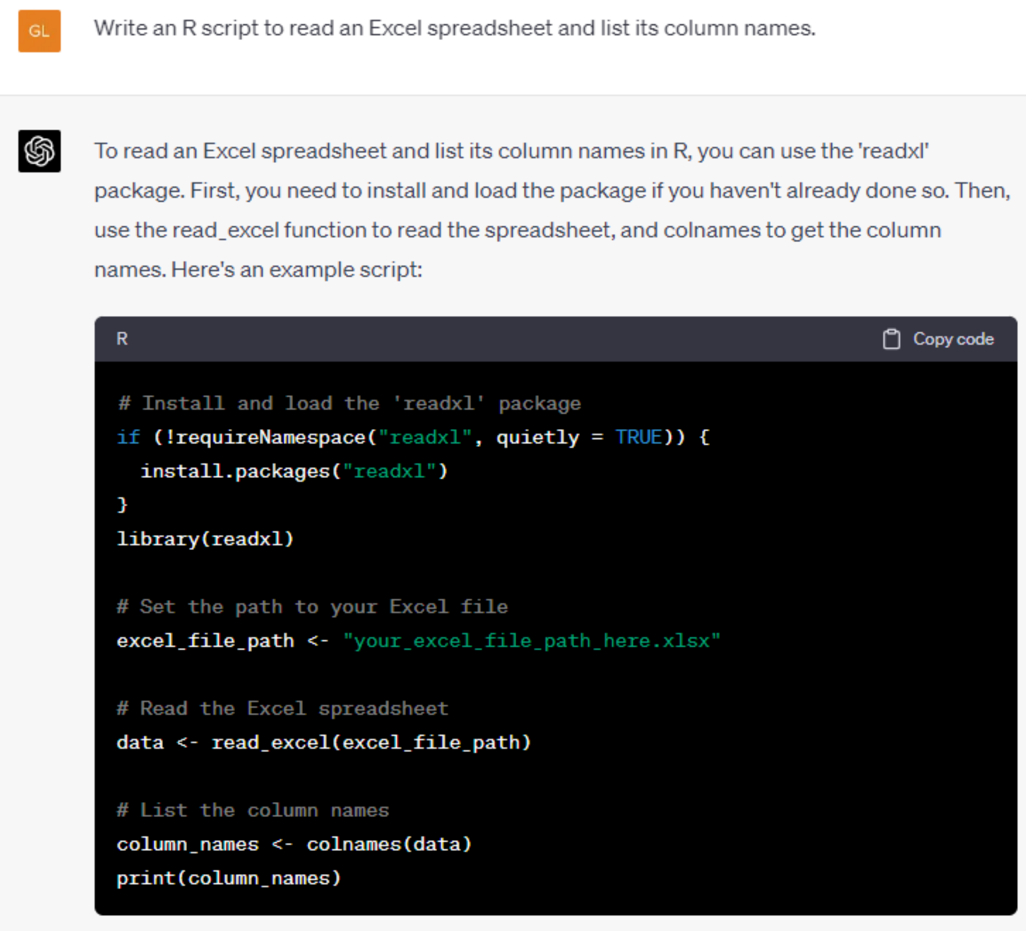
1. Assuming your data table has two columns: Column A contains the dates and Column B contains the count of hospital visits for each day.
2. In a new cell, you can use the AVERAGEIFS function combined with the WEEKDAY function. The WEEKDAY function returns a number representing the day of the week for a given date, with 1 being Sunday and 7 being Saturday. In this case, you want to find the average for Mondays, which are represented by the number 2.

Use the following formula, replacing the range (A2:A1826) and (B2:B1826) with the actual ranges containing your data:

```excel
=AVERAGEIFS(B2:B1826, A2:A1826, "=IF(WEEKDAY(A2:A1826,1)=2,1,0)")
```

This formula will calculate the average of the hospital visit counts in the range B2:B1826, but only for the rows where the corresponding date in the range A2:A1826 has a WEEKDAY value of 2 (Mondays).

Writing and Debugging Code



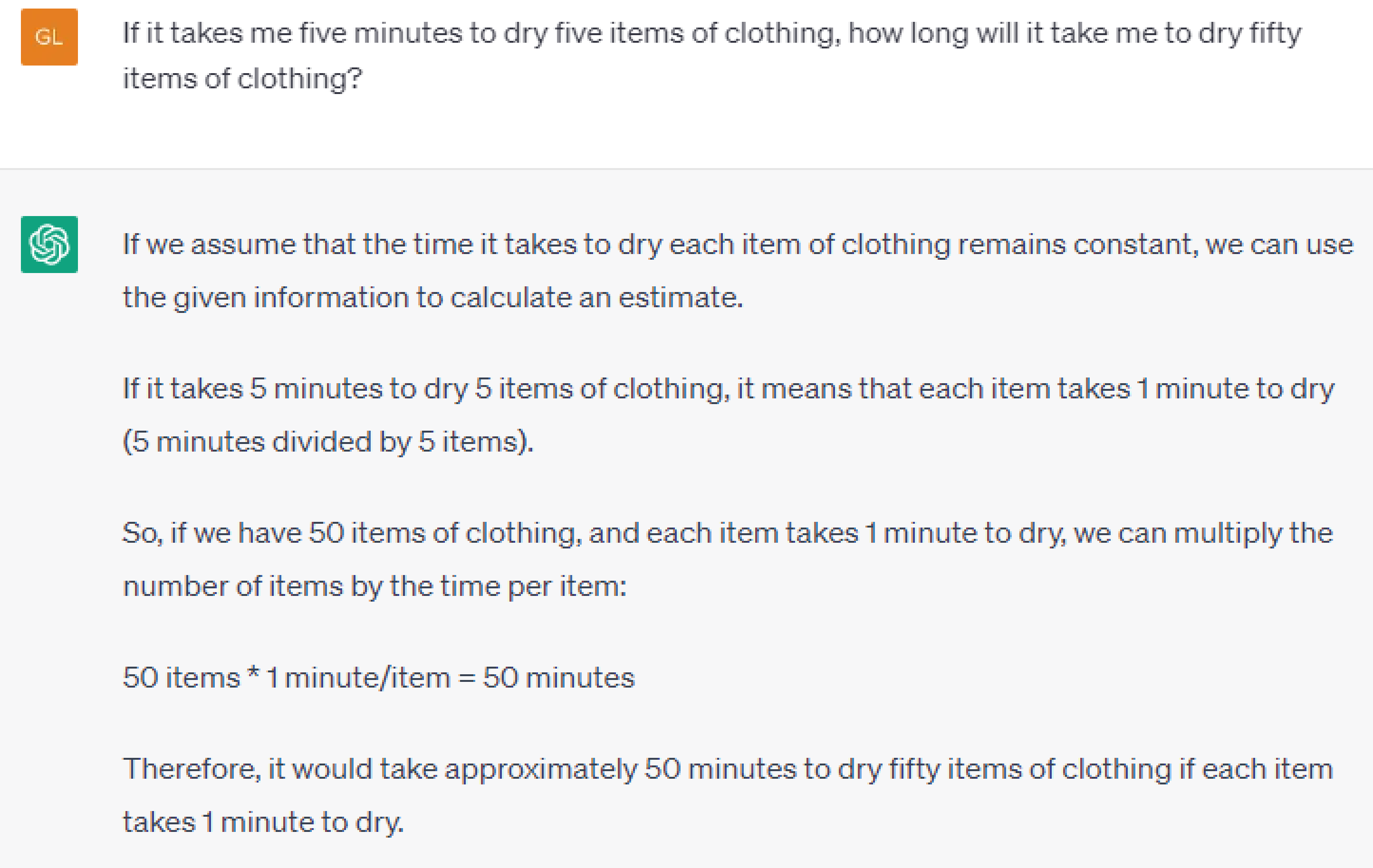
Write an R script to read an Excel spreadsheet and list its column names.

To read an Excel spreadsheet and list its column names in R, you can use the 'readxl' package. First, you need to install and load the package if you haven't already done so. Then, use the read_excel function to read the spreadsheet, and colnames to get the column names. Here's an example script:

```r
# Install and load the 'readxl' package
if (!requireNamespace("readxl", quietly = TRUE)) {
  install.packages("readxl")
}
library(readxl)

# Set the path to your Excel file
excel_file_path <- "your_excel_file_path_here.xlsx"

# Read the Excel spreadsheet
data <- read_excel(excel_file_path)

# List the column names
column_names <- colnames(data)
print(column_names)
```

Key Limitations and Mitigations

Reasoning Limitations



If it takes me five minutes to dry five items of clothing, how long will it take me to dry fifty items of clothing?

If we assume that the time it takes to dry each item of clothing remains constant, we can use the given information to calculate an estimate.

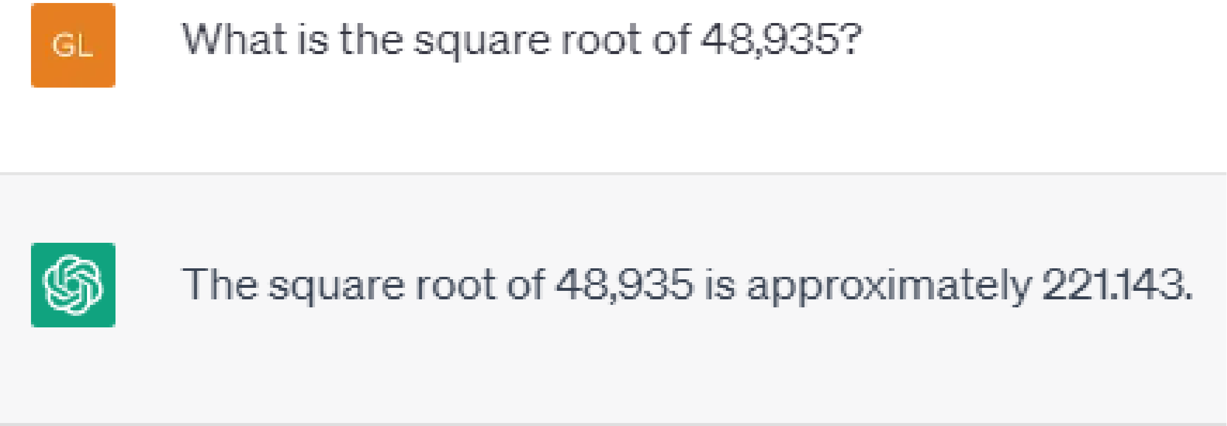
If it takes 5 minutes to dry 5 items of clothing, it means that each item takes 1 minute to dry (5 minutes divided by 5 items).

So, if we have 50 items of clothing, and each item takes 1 minute to dry, we can multiply the number of items by the time per item:

50 items * 1 minute/item = 50 minutes

Therefore, it would take approximately 50 minutes to dry fifty items of clothing if each item takes 1 minute to dry.

Numbers and Calculations



What is the square root of 48,935?

The square root of 48,935 is approximately 221.143.

- Actually, it's 221.213 (3dp)

Key Limitations and Mitigations

- ~~Knowledge cut-off (September 2021)~~
- Confabulations → *check facts*
- Reasoning limitations → *show step by step*
- Numbers and calculations → *check calculations*
- Difficulty tracking context → *stick to one topic*
- Size limitations (context window) → *break into chunks*
- Training data bias

Supplementary Slides

Legal and Ethical Considerations

Ownership and Privacy of Prompts

See https://openai.com/policies/terms-of-use

- *"to the extent permitted by applicable law, you own all Input"*
- BUT: *"We may use Content … to help develop and improve our Services"* (note: there are ways to opt out of this)
- SO: Prompts and their content are not private **and so should not contain sensitive information**
- ALSO: *"If you use the Services to process personal data, you must provide legally adequate privacy notices and obtain necessary consents for the processing of such data"*

Ownership of Responses

- *"Subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output"*
- BUT: *"Responses that are requested by and generated for other users are not considered your Content"*
So, cannot own small text snippets or responses to factual questions
- *"You may not use the Services in a way that infringes, misappropriates or violates any person's rights"* – and this includes **copyright infringement.** If generated content resembles copyrighted property, infringement claims can be made.

Representation of Responses

- *“You may not represent that output from the Services was human-generated when it is not”*
- SO: where possible, include a statement that makes it clear that machine-generated content may have been used

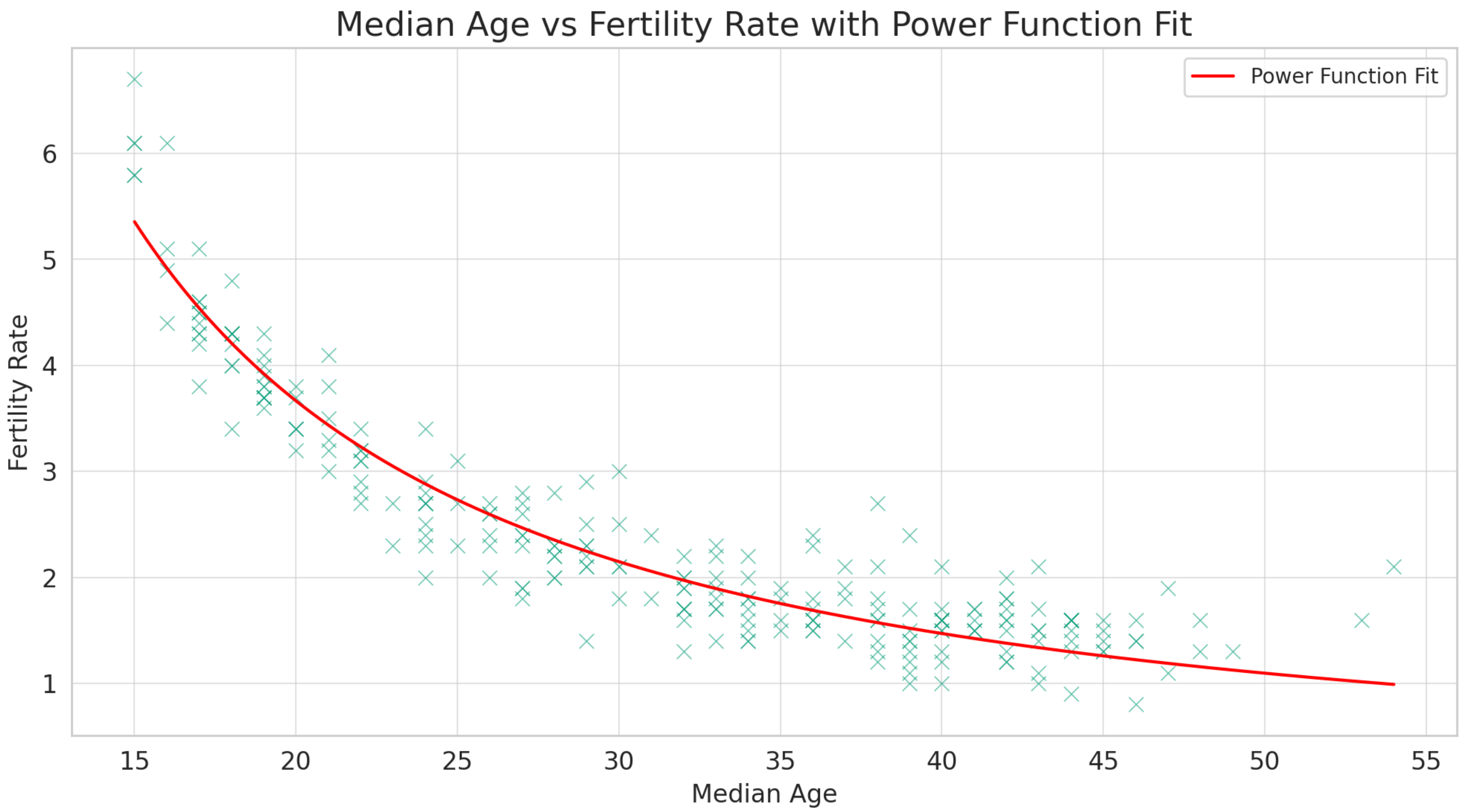
Uses for ChatGPT and similar technology

Examples of Practical Uses

- Getting summaries of documents
- Helping you get past the “blank screen”
- Literature scanning in a topic area
- Developing policies against standards: Rewording and restructure
- Exploring data …

Code Interpreter Demo

Code Interpreter Demo



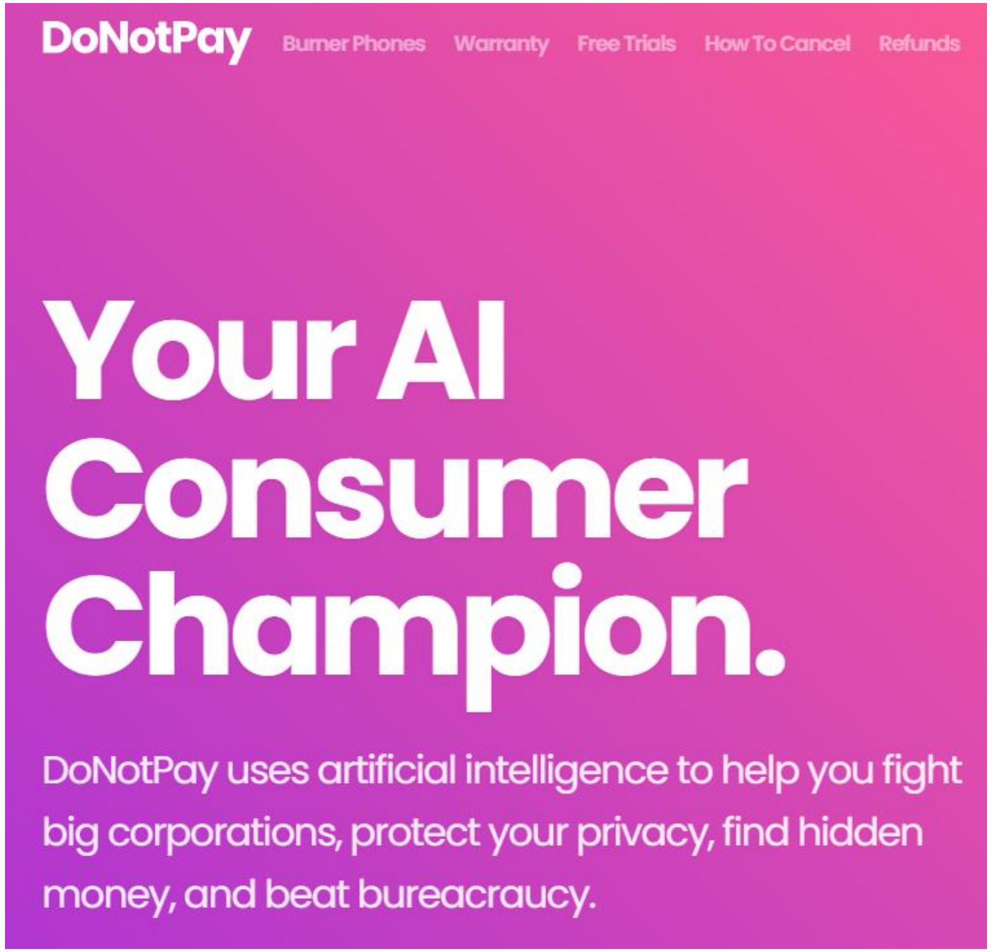
Related Technology and Uses

Similar Technology to ChatGPT

- Gemini (Google)
- Bing Chat (Microsoft)
- Claude (Anthropic)

- … and many more that incorporate such technologies into their workstream …

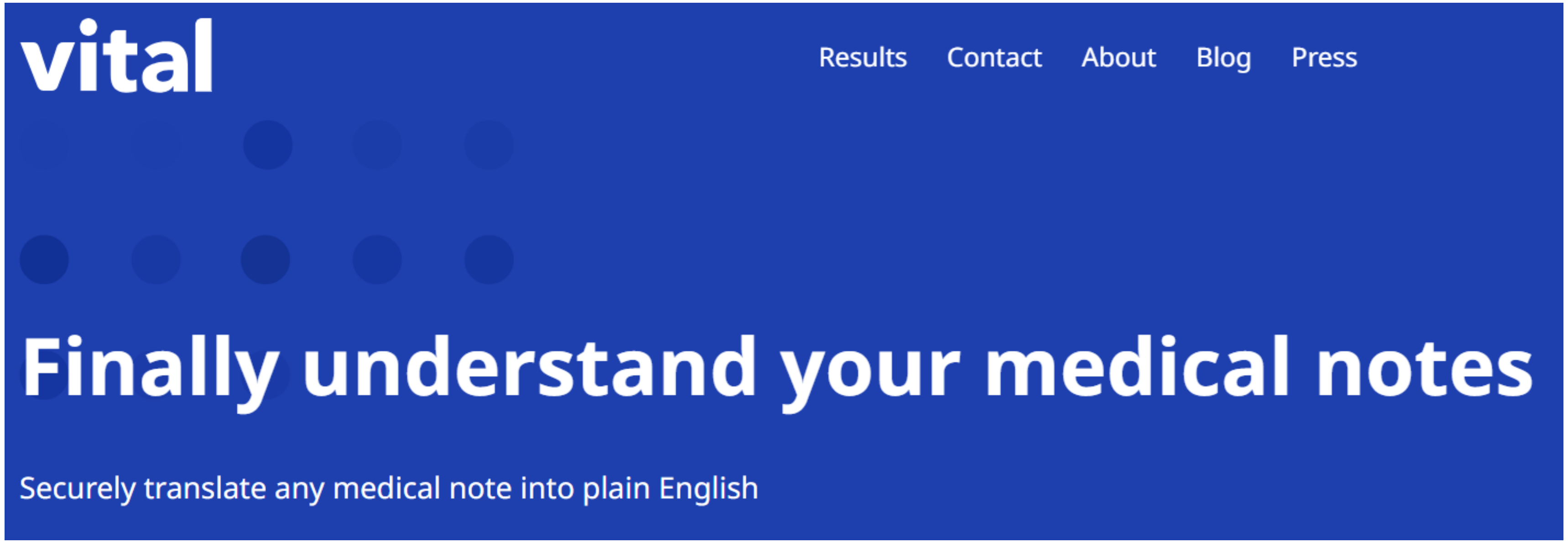
DoNotPay (donotpay.com)



Your AI Consumer Champion.

DoNotPay uses artificial intelligence to help you fight big corporations, protect your privacy, find hidden money, and beat bureacraucy.

Vital (vital.io/translate)

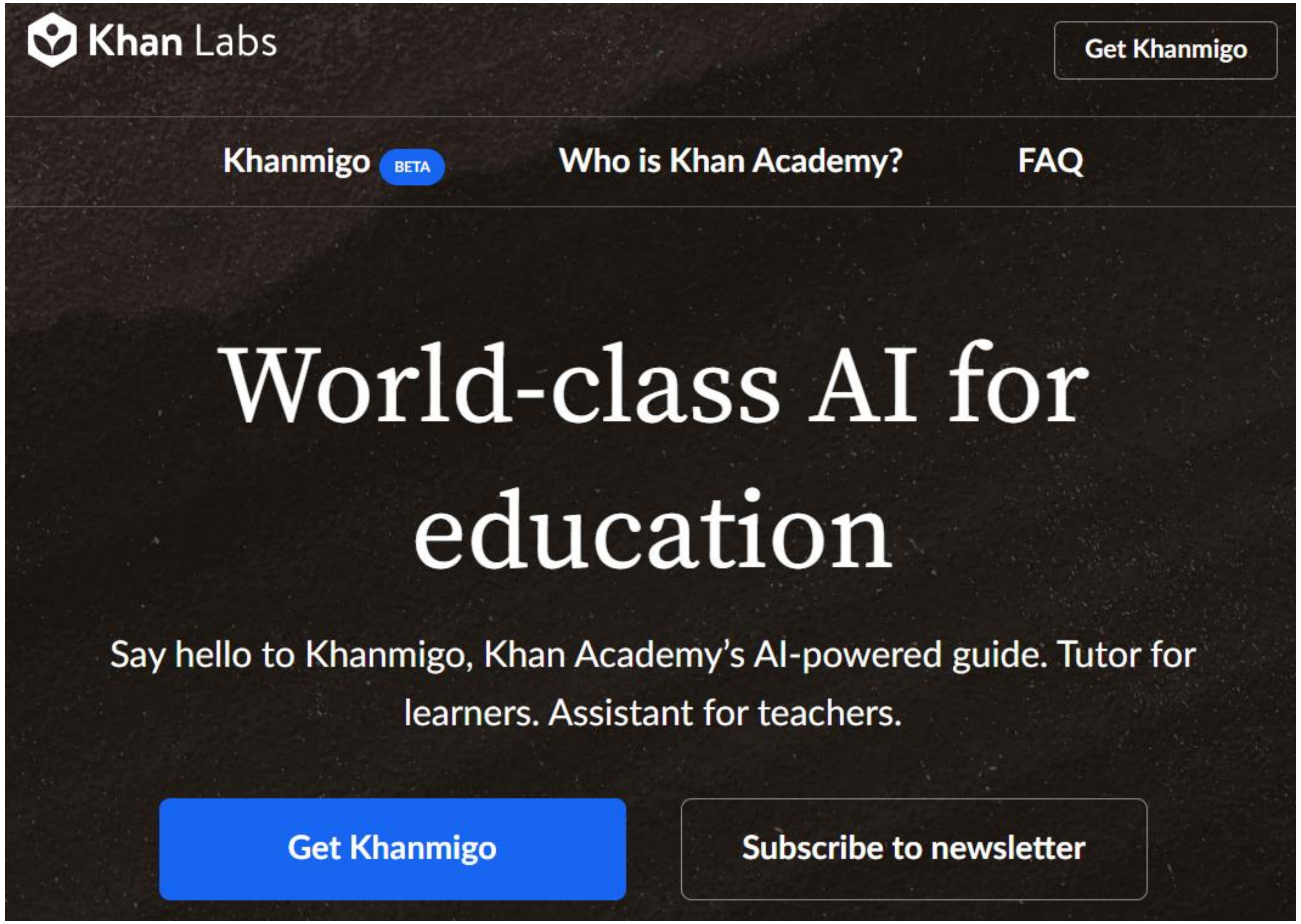


vital

Finally understand your medical notes

Securely translate any medical note into plain English

Khan Academy (khanacademy.org/khan-labs)



World-class AI for education

Say hello to Khanmigo, Khan Academy's AI-powered guide. Tutor for learners. Assistant for teachers.

Closing Karakia

Kia hora te marino
Kia whakapapa pounamu te moana
Hei huarahi mā aroha mai
Tātou i a tatou katoa
Hui e! Tāiki e!