

# Beyond Words: Exploring Large Language Models

Itamar Caspi (Bank of Israel)

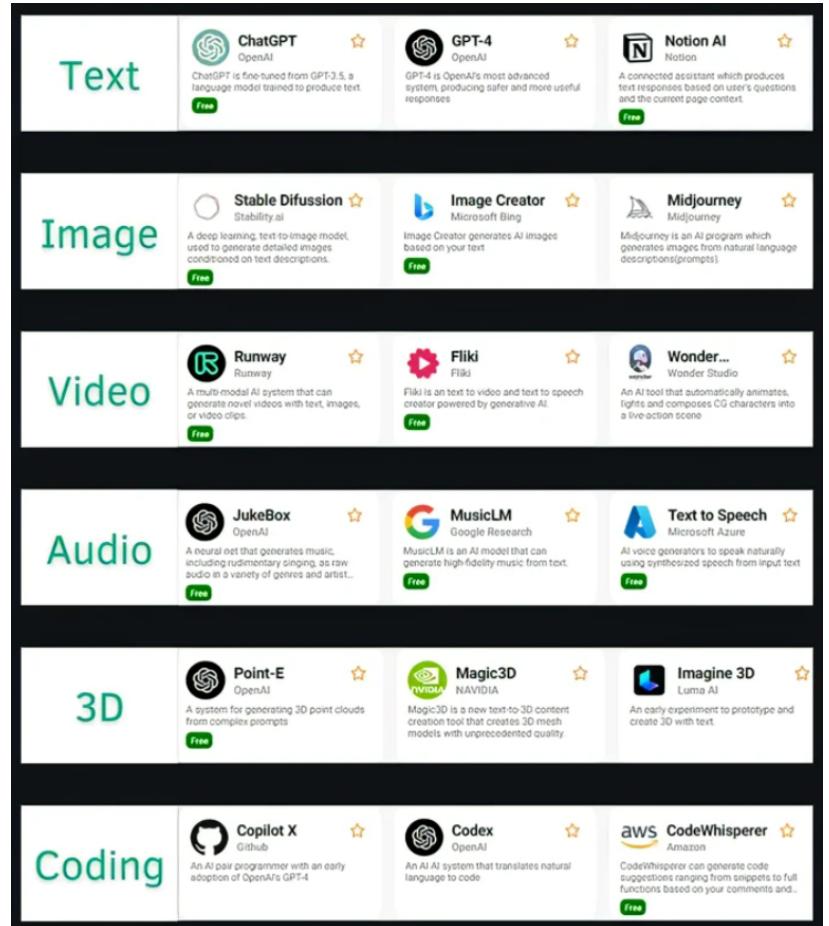
May 16, 2023 (updated: 2023-05-18)



# The New Wave of Generative AI

The new wave of AI extends beyond language:

- **Images:** MidJourney
- **Video:** Runway
- **Music:** MusicML
- **Coding:** Copilot X
- **3D:** Imagine 3D



# AI and Key Challenges Looming Ahead

- **Computing Power:** The impact of powerful hardware on AI.
- **Emergent Abilities:** Unforeseen AI capabilities arising from learning.
- **AI Safety / Alignment:** Ensuring AI behavior aligns with human values.
- **Misinformation:** The risks and mitigation of AI-generated false info.

Can you spot the fake?



Source: **USA Today, March 15, 2019**

# Outline

1. What are LLMs?
2. GPT-4 and ChatGPT
3. Principles of Prompting
4. The AI Research Assistant
5. Real World Impact

# What are LLMs?

*"It's Just adding one word at a time"*

Stephen Wolfram (Feb, 2023)



Source: Midjourney (V5.1). Prompt: It's just adding one word at a time.

# Large Language Models (LLMs)

"Large language models, like GPT-4, are AI systems trained on vast amounts of text data. They use patterns in this data to generate human-like text, answer questions, translate languages, and perform other language-related tasks. They don't understand information, but predict what text should come next based on their training."

—ChatGPT (2023)

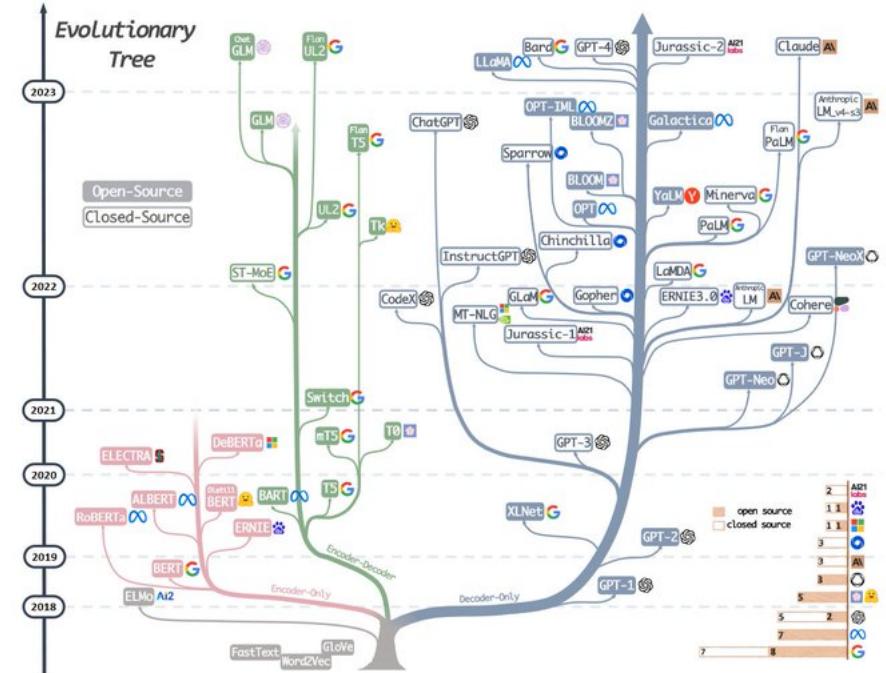


Fig. 1. The evolutionary tree of modern LLMs traces the development of language models in recent years and highlights some of the most well-known models. Models on the same branch have closer relationships. Transformer-based models are shown in non-grey colors: decoder-only models in the blue branch, encoder-only models in the pink branch, and encoder-decoder models in the green branch. The vertical position of the models on the timeline represents their release dates. Open-source models are represented by solid squares, while closed-source models are represented by hollow ones. The stacked bar plot in the bottom right corner shows the number of models from various companies and institutions.

Source: <https://arxiv.org/abs/2304.13712>

# Collect a Lot of Data

- **Internet:** A vast ocean of diverse topics and styles.
- **Books:** Boosts language proficiency and narrative comprehension.
- **Wikipedia:** Encyclopedic knowledge for a broad understanding.
- **Non-Proprietary Databases:** Broad, general knowledge without user or company specifics.

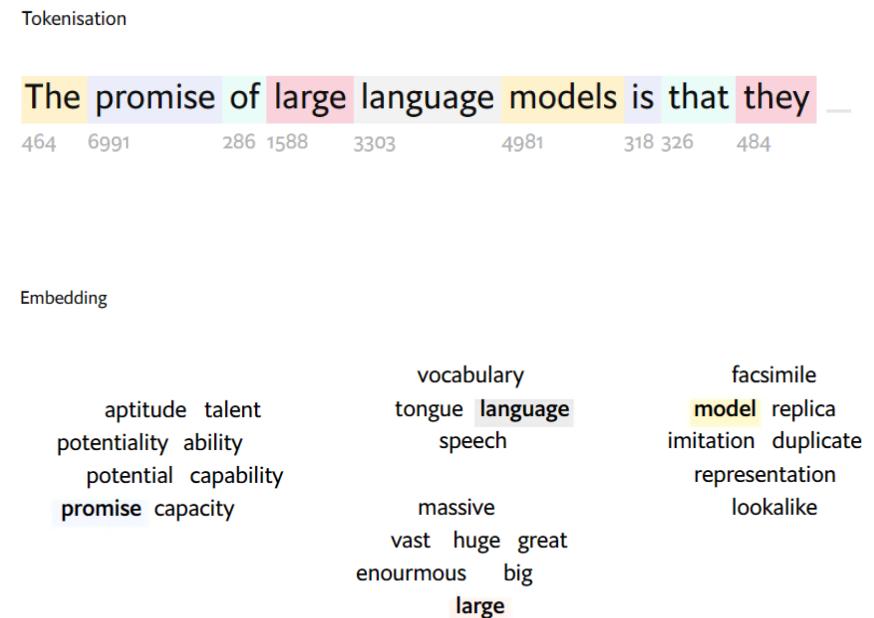
*Important:* OpenAI doesn't disclose the exact sources of these datasets.



Source: Bing image generator ("a huge pile of books")

# Contextual Embedding: From Text to Numbers

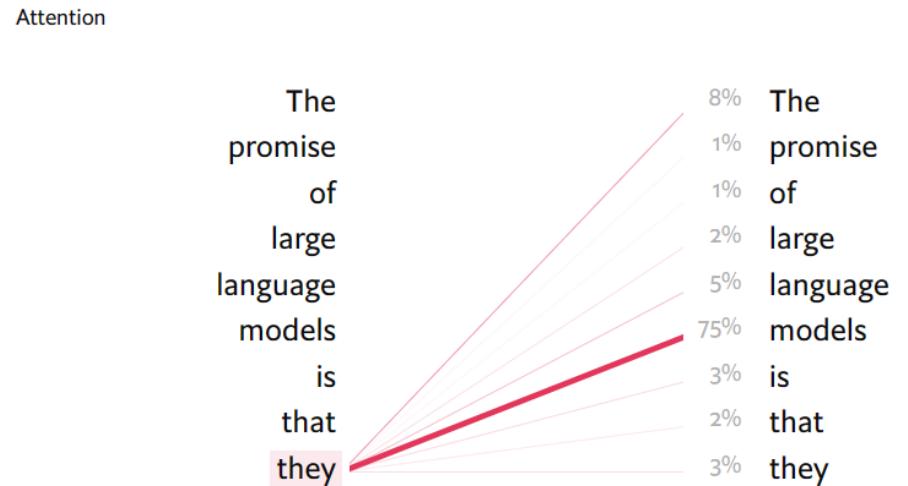
- **LLMs:** Transform language queries into numerical representations.
- **Tokenization:** Breaks down text into chunks called tokens.
- **Tokens:** Include words, affixes, and punctuation.
- **Meaning Space:** Area where tokens with similar meanings cluster together.



Source: [The Economist](#).

# Attention Network

- LLMs use "attention networks" to connect different parts of the prompt.
- Attention networks enable models to focus on the most informative parts of the input data, improving performance and accuracy.
- LLMs translate language structure into numerical "weights" in the neural network during training.
- LLMs comprehend language statistically rather than grammatically.



Source: [The Economist](#).

# Next Word Prediction

- Upon processing the prompt, the LLM begins generating a response.
- The attention network assigns each token a probability, suggesting its likelihood as the next part of the sentence.
- The token with the highest probability isn't always selected. The LLM's creativity level, determined by its operators, impacts this choice.

Completion

The promise of large language models is that they —

can 62%  
will 11%  
are 7%  
capture 2%  
could 2%

Source: [The Economist](#).

# Autoregression: The Loop of Response Generation

- The LLM produces a word and feeds the output back into itself.
- This cycle, known as autoregression, repeats until the LLM completes its response.

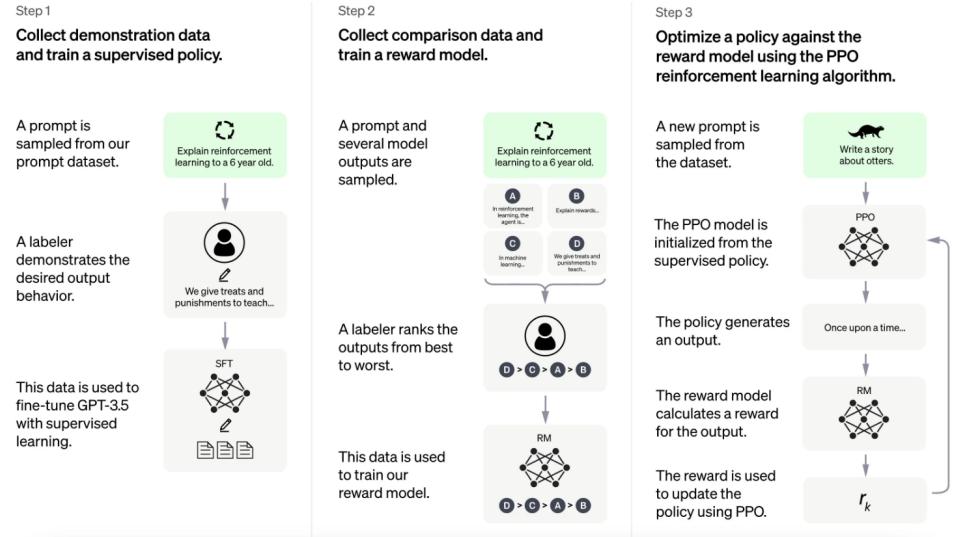
Completion

The promise of large language models is that they can be used to generate text that is indistinguishable from human-written text.

Source: [The Economist](#).

# Reinforcement Learning from Human Feedback (RLHF)

- RLHF uses human ratings to train chatbots in three stages:
  1. Pretrain the language model.
  2. Collect and implement human feedback.
  3. Fine-tune using reinforcement learning.
- Advantages: Mimics human preferences, refines chatbot responses, and boosts user satisfaction.



Source: [OpenAI: Introducing ChatGPT](#)

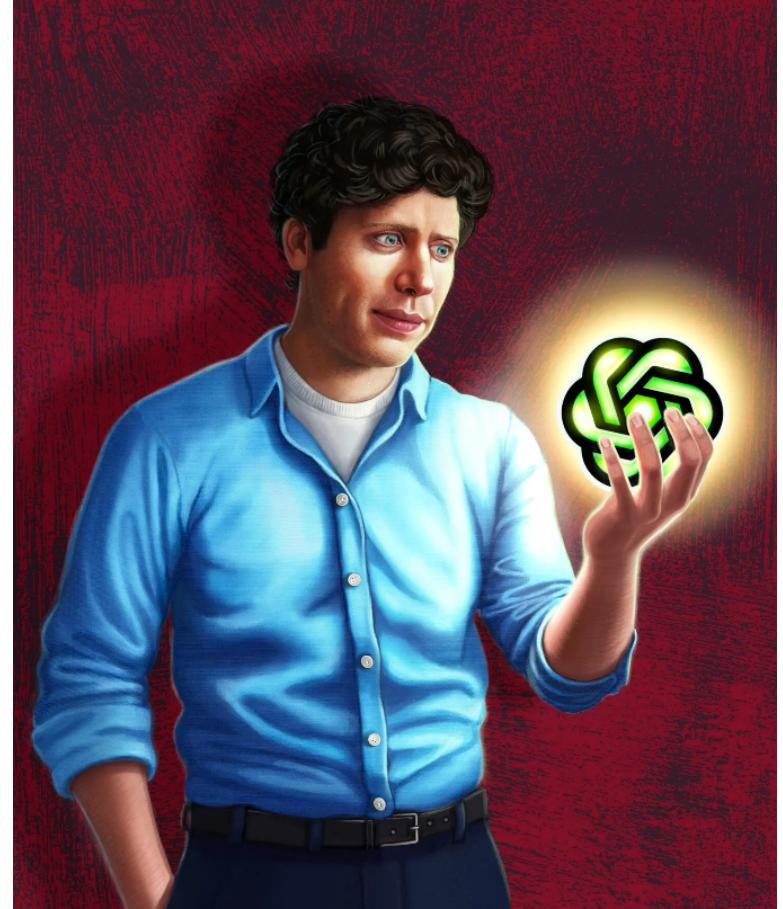
# GPT-4 and ChatGPT

# Key Resources for Beginners

- **ChatGPT Prompt Engineering for Developers**: A free course suitable for beginners. It offers a tutorial on using the OpenAI API and large language models for innovative applications. It places a strong emphasis on effective prompt engineering through practical examples.
- **Language Models and Cognitive Automation for Economic Research** by Anton Korinek: It includes general instructions and specific examples on how to utilize LLMs. It also categorizes the capabilities of LLMs from experimental to highly practical.
- **One Useful Thing** blog by Ethan Mollick (Wharton): For regular updates, follow [@emollick](#) on Twitter.
- **What Is ChatGPT Doing ... and Why Does It Work?** by Stephen Wolfram: This is a comprehensive introduction to the scientific principles behind ChatGPT.

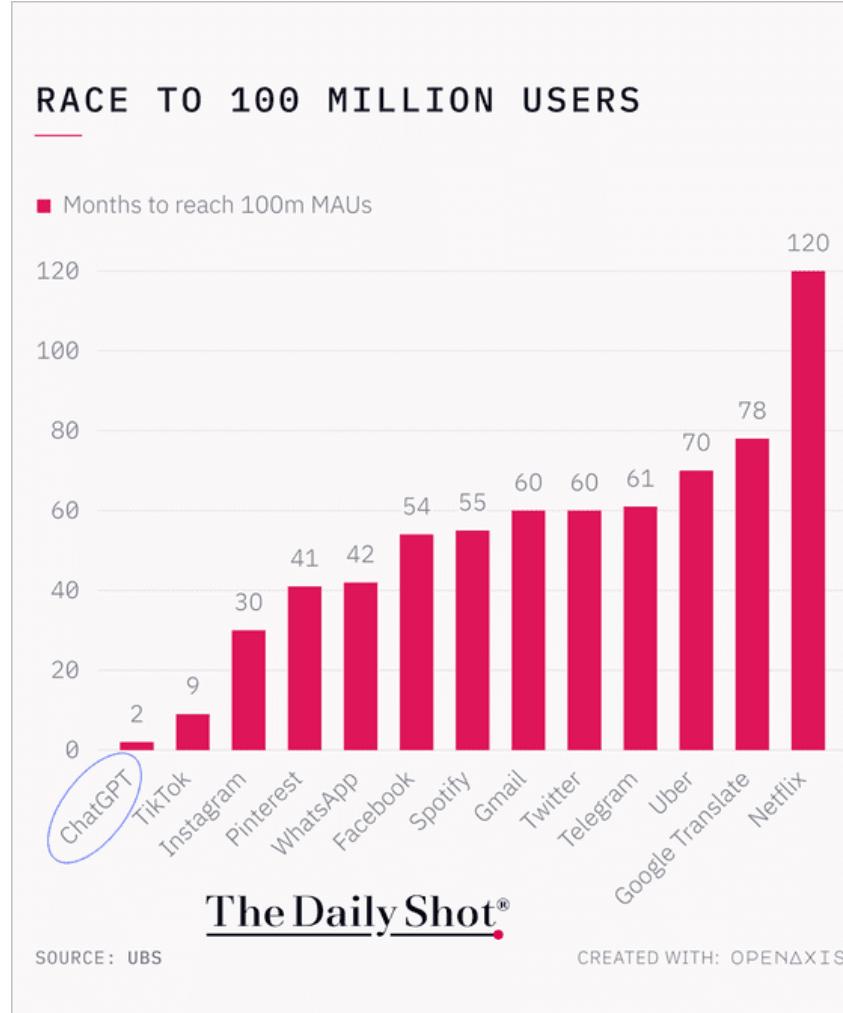
# GPT-4: The Latest Advancement in AI

- Introducing Generative Pre-trained Transformer 4 (GPT-4).
- A *multimodal* language model capable of processing text and images.
- Developed by OpenAI, released in March 2023.
- Abilities include chatting, coding, tutoring, translating, and more.
- However, it shares some issues with its predecessors: hallucinations, biases, and the ability to inadvertently reveal secrets.



Source: [The ChatGPT King Isn't Worried](#)

# The Big Bang



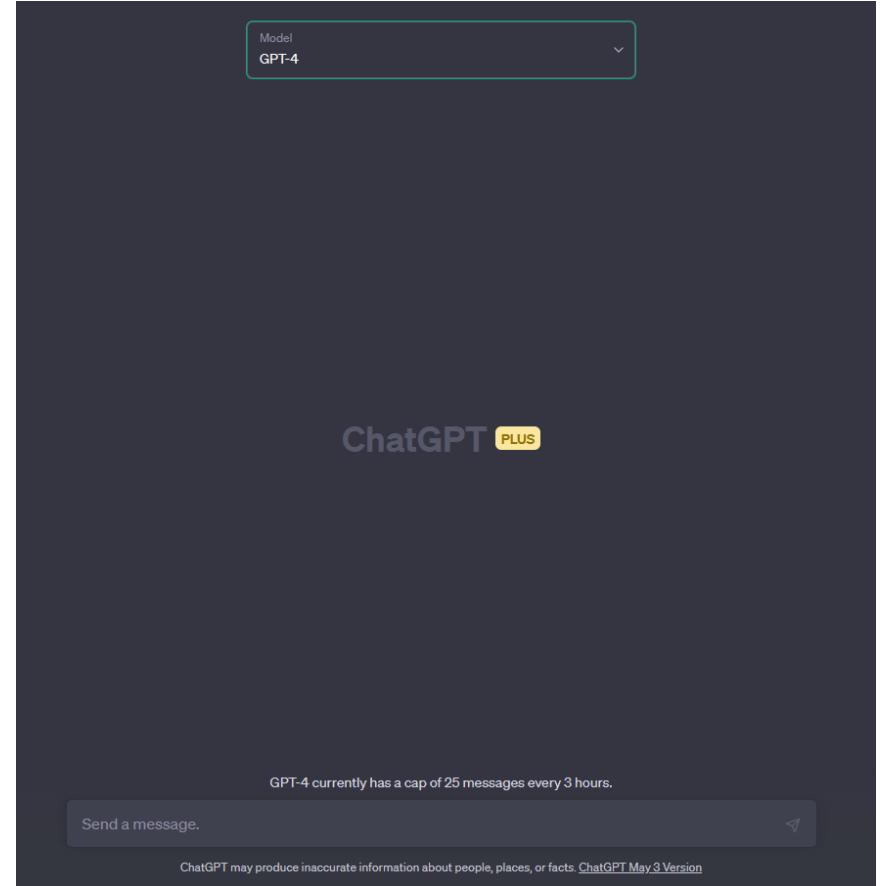
# Accessing GPT-4: A Step-by-Step Guide

First, create an OpenAI account.

- Option 1: Access via [ChatGPT](#)
- Option 2: Try the [OpenAI Playground](#)
- Option 3: Use the [OpenAI API](#)
- Option 4: Explore through [Bing](#)

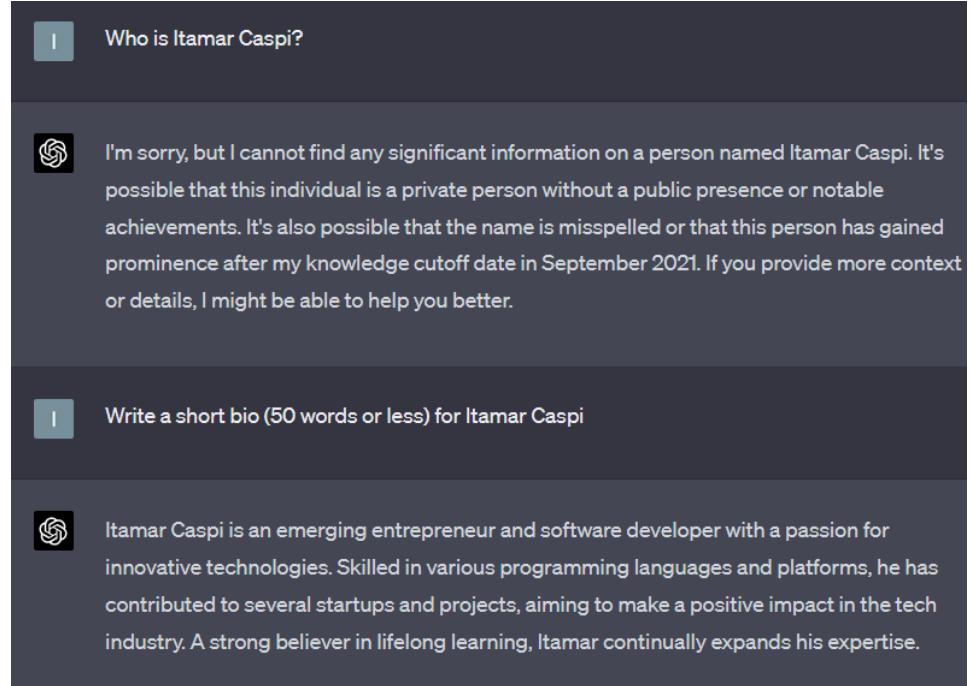
*Important:* GPT-3.5 and Bing offer free access. GPT-4 costs approximately \$20 per month. To save costs, consider using the API (\$0.03 per token).

[Google's Bard](#) provides a free alternative (not GPT-4).



# Yes, ChatGPT Can Hallucinate

- ChatGPT may produce text that appears credible but is inaccurate.
- Its goal is to sound convincing, not necessarily to deliver accurate or reliable information.
- This could potentially mislead unsuspecting readers.
- OpenAI is continually striving to improve the model's accuracy and reliability.



The screenshot shows two interactions with ChatGPT. In the first interaction, the user asks "Who is Itamar Caspi?" and the AI responds that it cannot find significant information about this person, suggesting they might be a private individual or the name is misspelled. In the second interaction, the user asks the AI to write a short bio (50 words or less) for Itamar Caspi, and the AI generates a completely fabricated response, stating that Itamar Caspi is an emerging entrepreneur and software developer with a passion for innovative technologies, skilled in various programming languages and platforms, and has contributed to several startups and projects.

I Who is Itamar Caspi?

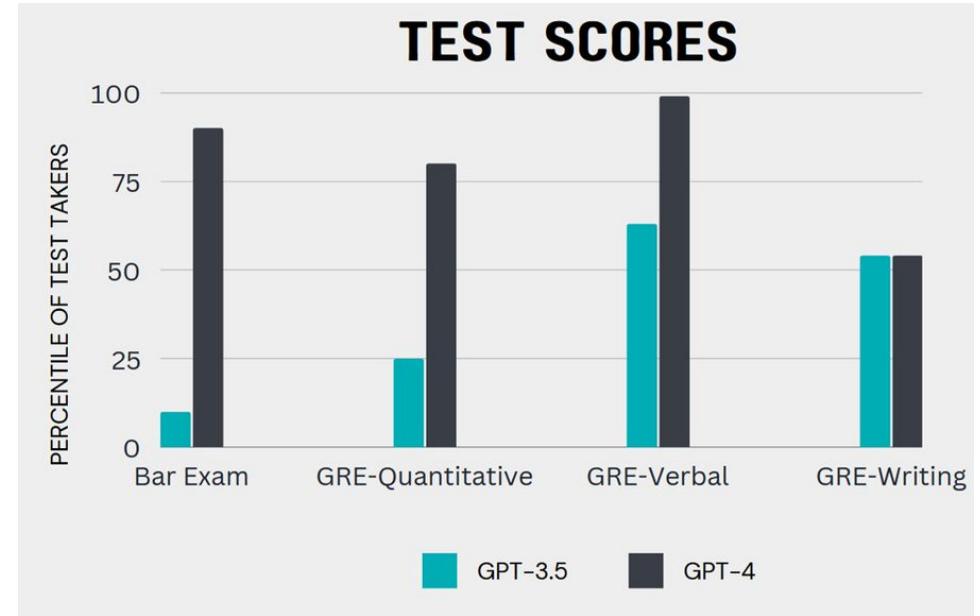
I'm sorry, but I cannot find any significant information on a person named Itamar Caspi. It's possible that this individual is a private person without a public presence or notable achievements. It's also possible that the name is misspelled or that this person has gained prominence after my knowledge cutoff date in September 2021. If you provide more context or details, I might be able to help you better.

I Write a short bio (50 words or less) for Itamar Caspi

Itamar Caspi is an emerging entrepreneur and software developer with a passion for innovative technologies. Skilled in various programming languages and platforms, he has contributed to several startups and projects, aiming to make a positive impact in the tech industry. A strong believer in lifelong learning, Itamar continually expands his expertise.

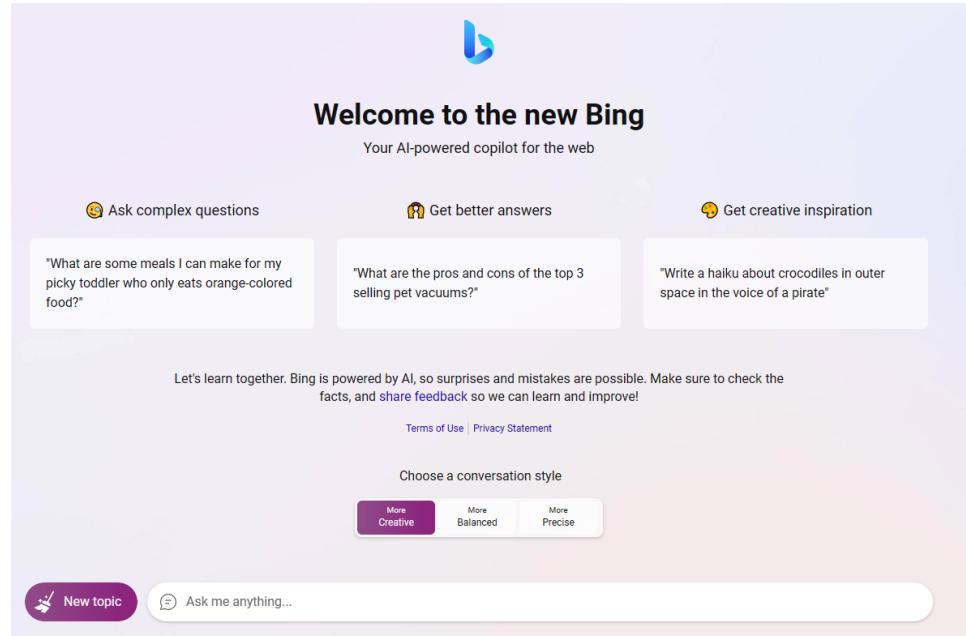
# Powerful, Not Dumb: The Capabilities of ChatGPT

- Despite occasional errors, ChatGPT's prowess should not be underestimated.
- It demonstrated an impressive IQ of 147 on a verbal-linguistic test (99.9th percentile).
- Each new iteration of ChatGPT enhances its accuracy and power.
- The potential for substantial societal impact must not be dismissed or underestimated.



# A Note on Bing

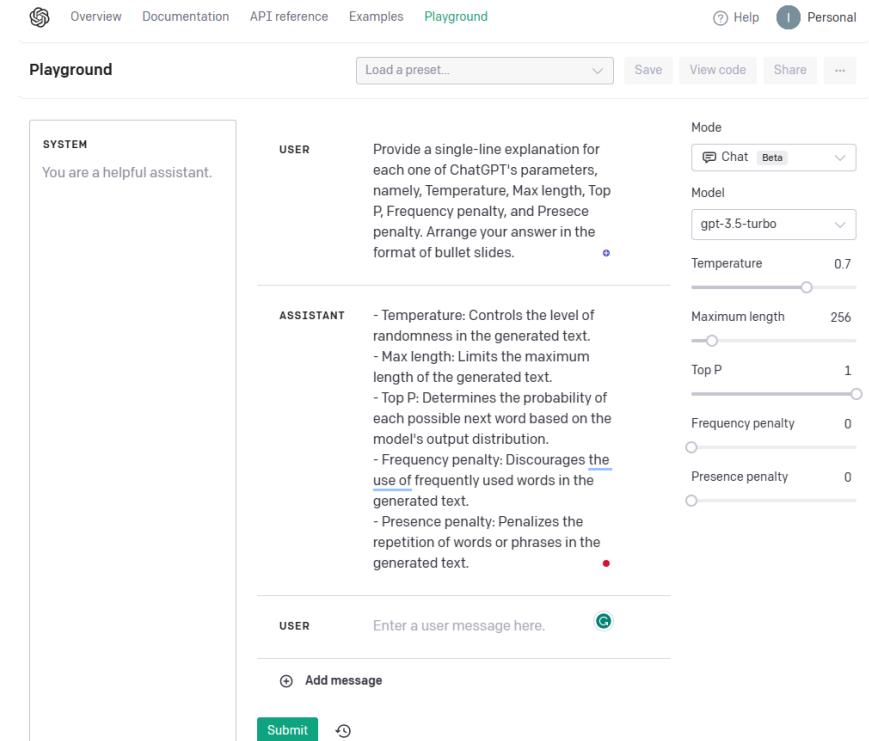
- Bing provides free access to GPT-4 in both "precise" and "creative" modes.
- Unlike ChatGPT, Bing Chat utilizes an active internet connection to its advantage.
- Integration with Microsoft's Edge browser enables it to "read" visited webpages.
- One constraint to be aware of is the 2000 character limit per input.



# Principles of Prompting

# Perquisite: Understanding Model Parameters

- **Temperature**: Controls the randomness in text generation.
- **Max Length**: Determines the maximum length of the generated text.
- **Top P**: Allocates the probability of each potential next word, based on the model's output distribution.
- **Frequency Penalty**: Discourages the use of commonly occurring words in the generated text.
- **Presence Penalty**: Imposes a penalty on repeated words or phrases in the generated text.



Note: You can only adjust these parameters in "playground" mode or via the API.

# Principle #1: Clarity and Specificity in Instructions

- Define the input using *delimiters*:
  - Triple quotes: """
  - Triple backticks: `
  - Triple dashes: ---,
  - and so on.
- Request *structured output* (e.g., HTML, JSON).
- Verify if the *conditions* are satisfied and *assumptions* required to perform the task.
- Provide successful *examples* of task completion, also known as **Few-shot prompting**.

# Defining Input with Delimiters

- Prompt: Summarize the text enclosed by triple backticks into one sentence.
- Alternatively: Condense the text within triple quotes into a single sentence.
- Or: Summarize the text bounded by <> into one sentence.
- And so on.



Summarize the text delimited by triple backticks into a single sentence.

...

You should express what you want a model to do by providing instructions that are as clear and specific as you can possibly make them.

This will guide the model towards the desired output, and reduce the chances of receiving irrelevant or incorrect responses. Don't confuse writing a clear prompt with writing a short prompt.

In many cases, longer prompts provide more clarity and context for the model, which can lead to more detailed and relevant outputs.

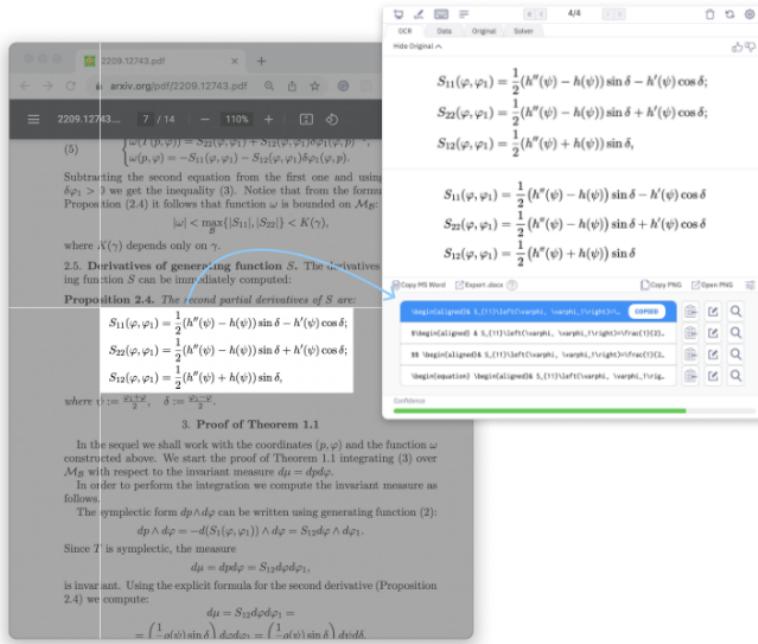
...



To ensure accurate and relevant responses, it is important to provide clear and specific instructions when interacting with a model, avoiding confusion between clear and short prompts as longer prompts often offer more clarity and context for the desired output.

# An aside: Extracting text and equations from everything

- **Mathpix** is an OCR tool that rapidly converts handwritten or printed math equations into LaTeX code.
- This tool greatly simplifies the task of extracting elements from academic papers, such as **symbols**, **equations**, and even **tables**.
- ChatGPT can recognize LaTeX code with ease.



# Request Structured Output

Requesting structured output enhances stability and allows for the replication of your results.

I Generate a list of the first 4 past Israeli prime ministers along with their date of birth and gender

1. David Ben-Gurion (first Israeli Prime Minister)  
Date of birth: October 16, 1886  
Gender: Male

2. Moshe Sharett  
Date of birth: October 15, 1894  
Gender: Male

3. Levi Eshkol  
Date of birth: October 25, 1895  
Gender: Male

4. Golda Meir  
Date of birth: May 3, 1898  
Gender: Female

I Generate a list of the first four past Israeli prime ministers along with their date of birth and gender

Here is a list of the first four Israeli Prime Ministers, along with their date of birth and gender:

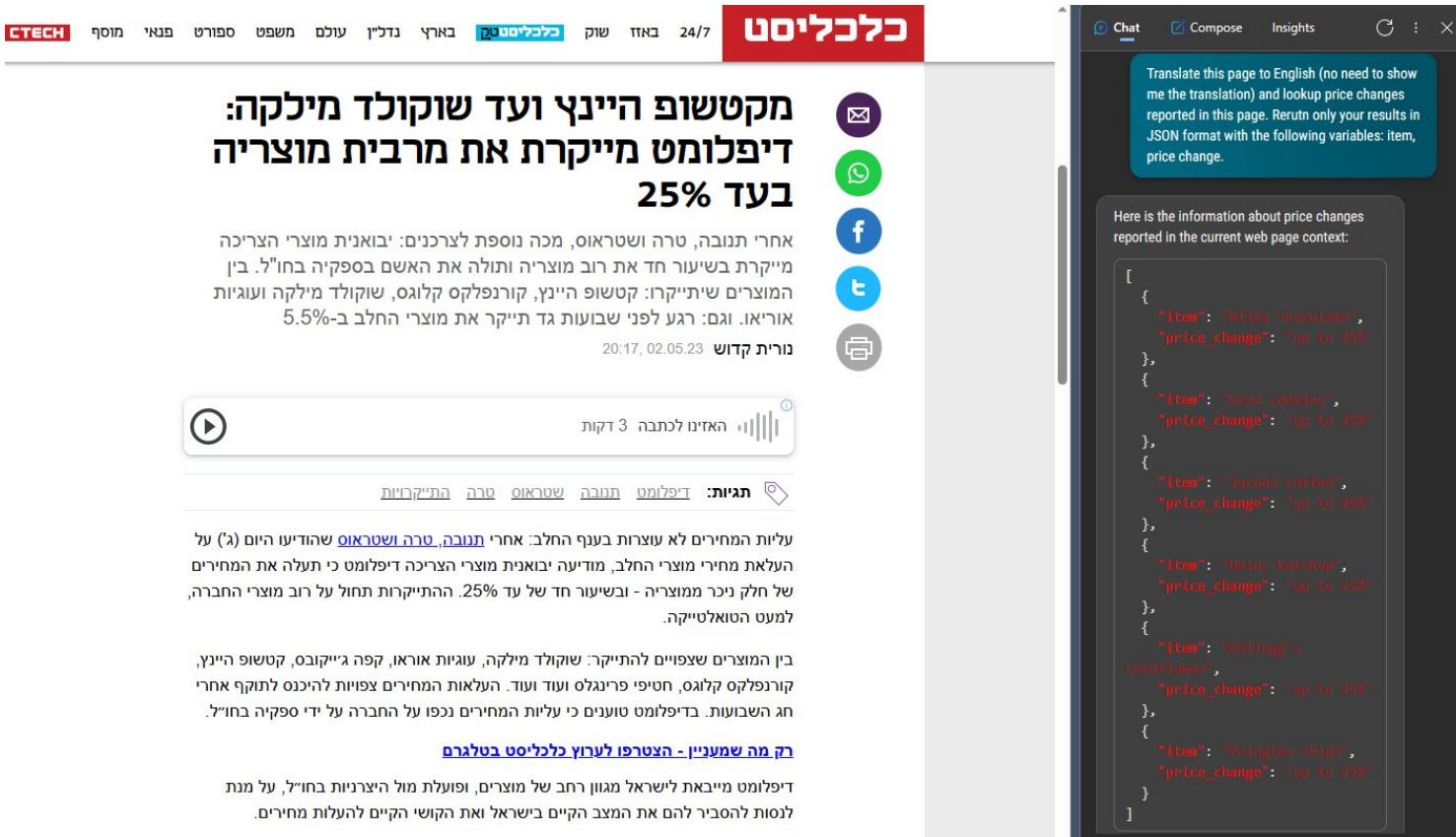
1. David Ben-Gurion (male) - Born on October 16, 1886
2. Moshe Sharett (male) - Born on October 15, 1894
3. Levi Eshkol (male) - Born on October 25, 1895
4. Golda Meir (female) - Born on May 3, 1898

I Generate a list of the first four past Israeli prime ministers along with their date of birth and gender. Provide them in JSON format with the following keys: name, dob, gender.

```
[{"name": "David Ben-Gurion", "dob": "1886-10-16", "gender": "male"}, {"name": "Moshe Sharett", "dob": "1894-10-15", "gender": "male"}, {"name": "Levi Eshkol", "dob": "1895-10-25", "gender": "male"}, {"name": "Golda Meir", "dob": "1898-05-03", "gender": "female"}]
```

# Use Case: Web Scraping

Execute tasks like translating content from Hebrew to English, monitoring changes in item prices, and formatting the output in JSON.



24/7 באזן שוק כלכליסט פנאי מוסף בארץ נדלין עולם משפט ספורט פנאי מוסף

## מקטשוף היינץ ועד שוקולד מלקה: דיפלומט מייקרת את מרבית מוצריה בעד 25%

אחרי תנובה, טרה ושטרואס, מכח נוספת לצרכנים: יבואנית מוצר הצריכה מייקרת בשיעור חד את רוב מוצריה ותולוה את האשם בספקיה בחו"ל. בין המוצרים שיתיקרו: קטשוף היינץ, קורנפלקס קלוגם, שוקולד מלקה ועוגיות אויריאו. וגם: רגע לפני שבועות גד תייקר את מוצריו החלב ב-5.5%. נורית קדש, 20:17, 02.05.23

האזינו לכתבה 3 דקות

תגיות: דיפלומט תנובה טרה וטרואס טבה בתיקריות

עלויות המוצרים לא עשוות בענף החלב: אחריה **טנהה, טרה וטרואס** שהודיעו הימים (ג') על העילאת מחירי מוצר החלב, מודיעעה יבואנית מוצר הצריכה דיפלומט כי עלתה את המחיר של חלב ניכר ממוצריה - ובשיעור חד של עד 25%. ההתייקרות תחול על רוב מוצרים החברה, למעט הטואטלי"קה.

בין המוצרים שצפויים להתייקר: שוקולד מלקה, עוגיות אויריאו, קפה ג'ייקובס, קטשוף היינץ, קורנפלקס קלוגם, חטיפי פרינגלס ועד ועוד. העילאות המוצרים צפויות להיכנס לתוקף לאחר חג השבעות. בדיפלומט טוענים כי עלויות המוצרים נקבעו על ידי ספקיה בחו"ל.

**ברק מה שמנני - הטרפו לעוז כלכליסט בטלאגט**

דיפלומט מייבאת לישראל מגוון רחב של מוצרים, ופעלת מול היצרניות בחו"ל, על מנת לנסות להסביר להם את המצב הקיים בישראל ואת הקושי הקיים להעלות מחירים.

Translate this page to English (no need to show me the translation) and lookup price changes reported in this page. Rerun only your results in JSON format with the following variables: item, price\_change.

Here is the information about price changes reported in the current web page context:

```
[{"item": "Milka chocolate", "price_change": "up to 25%"}, {"item": "Oreo cookies", "price_change": "up to 25%"}, {"item": "Jacobs coffee", "price_change": "up to 25%"}, {"item": "Heinz ketchup", "price_change": "up to 25%"}, {"item": "Kellogg's Cornflakes", "price_change": "up to 25%"}, {"item": "Pringles chips", "price_change": "up to 25%"}]
```

# Few-shot Learning

- Few-shot learning is a machine learning approach that aims to train models to perform well on new tasks with only a limited amount of labeled data.

I Your task is to answer in a consistent style.

<child>: Teach me about patience.

<grandparent>: The river that carves the deepest valley flows from a modest spring; the grandest symphony originates from a single note; the most intricate tapestry begins with a solitary thread.

<child>: Teach me about resilience.

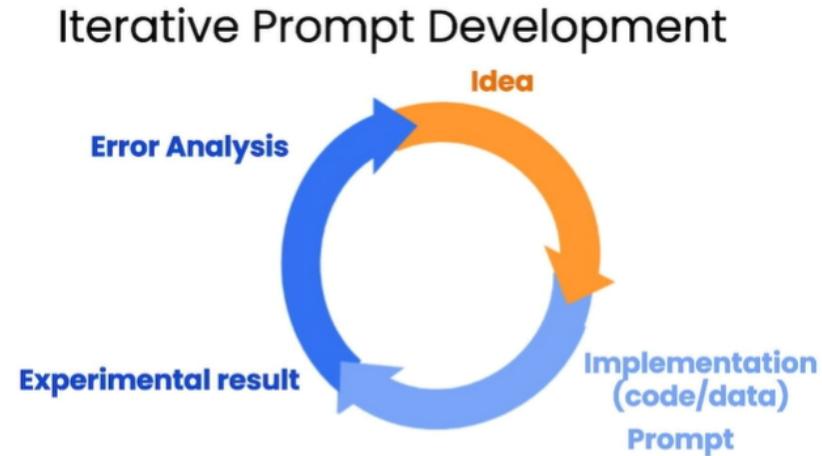
 <grandparent>: The mightiest oak in the forest was once a small acorn that held its ground; the most splendid diamond was once a piece of coal that withstood immense pressure; the brightest phoenix rises from the ashes of its own demise.

Source: [ChatGPT Prompt Engineering for Developers](#)

# Principle #2: Allowing AI Time to Think

- Clearly state the steps needed to complete the task.
- Encourage the model to think and generate its own solution before rushing to a conclusion.

# Principle #3: Prompting is an Iterative Process



## Prompt guidelines

- Be clear and specific
- Analyze why result does not give desired output.
- Refine the idea and the prompt
- Repeat

Source: [ChatGPT Prompt Engineering for Developers](#)

# Principle #4: Reduce Hallucinations by Thorough Preparation

1. Initially, search for relevant information.
2. Subsequently, formulate the answer based on this information.

Example prompt:

```
'step 1: Revisit and thoroughly read the policy memo.'  
'step 2: Conduct research to find academic papers that are relevant to the subjects discussed in the memo.'  
'step 3: Update the memo by incorporating references to support the assertions made within it.'  
'step 4: Compile a bibliography of these references and add it to the memo.'
```

# The AI Research Assistant

# Brainstorming with Language Models

- LLMs have extensive knowledge archives and can retrieve information upon request.
- You can incorporate ChatGPT into your idea generation process, including brainstorming, evaluations, counterarguments, and more.

I Please brainstorm economic channels through which rising the interest rate decreases inflation

Rising interest rates can decrease inflation through several economic channels:

1. **Consumer Spending:** When interest rates rise, the cost of borrowing increases. This discourages consumers from taking out loans for big-ticket items such as cars and homes, leading to a decrease in consumer spending. Since consumer spending is a significant component of aggregate demand, a decrease in consumer spending can lead to a decrease in demand-pull inflation.
2. **Savings:** Higher interest rates make saving more attractive relative to spending because individuals can earn more from their bank deposits. This shift from spending to saving can reduce aggregate demand, putting downward pressure on prices and thus reducing inflation.
3. **Investment Spending:** Rising interest rates increase the cost of borrowing for businesses as well. This can discourage firms from investing in new projects or expanding existing ones, leading to a decrease in aggregate demand, and thereby lowering inflation.
4. **Exchange Rates:** Higher interest rates can attract foreign investors looking for better returns on their investments. This can increase the demand for the domestic currency, causing it to appreciate. An appreciated currency makes imported goods cheaper, which can help reduce import-cost-driven inflation.
5. **Inflation Expectations:** Central banks often use interest rate hikes to signal their commitment to controlling inflation. This can help manage inflation expectations. If businesses and consumers expect future inflation to be lower, they may be less likely to raise prices or demand higher wages, helping to control inflation.
6. **Asset Price Inflation:** Higher interest rates can cool off overheated asset markets (like housing or stock markets) by making borrowing more expensive. This can prevent asset price bubbles, which can contribute to general inflation when they burst.
7. **Government Borrowing:** Higher interest rates increase the cost of government borrowing. This may reduce government spending or increase the incentive to lower budget deficits, both of which can reduce aggregate demand and therefore lower inflation.

Regenerate response

# Summarizing

FINANCIAL TIMES *myFT*

US COMPANIES TECH MARKETS CLIMATE OPINION WORK & CAREERS LIFE & ARTS HTSI

ePaper Portfolio Settings & Account

US interest rates [+ Add to myFT](#)

## Federal Reserve poised to implement another quarter-point rate rise

Latest meeting takes place against backdrop of US banking turmoil



Federal Reserve chair Jay Powell has been approached by 10 Democratic lawmakers asking him to refrain from further rate rises © Financial Times

Colby Smith in Washington 8 HOURS AGO

Feedback

Chat Compose Insights

Set tone of responses [Preview](#)

More Creative More Balanced More Precise

Welcome back! Here are some takeaways from this page.

The Federal Reserve is set to raise its benchmark policy rate by a quarter-point on Wednesday, the 10th consecutive increase in a year. The Fed is responding to high inflation, weak economic growth, and banking turmoil that have hit the US economy and financial system hard. The Fed faces challenges and pressures from the government, which has intervened to prevent more bank failures, from the public, which has urged the Fed to stop tightening, and from its own staff, which has forecast a recession this year. The Fed is expected to stop the rate hikes after this cycle, unless the outlook changes significantly.

What is the current policy rate?

How does the Fed control inflation?

What are the effects of high interest rates on the eco...

Ask me anything...

0/2000

# Proofread and Edit

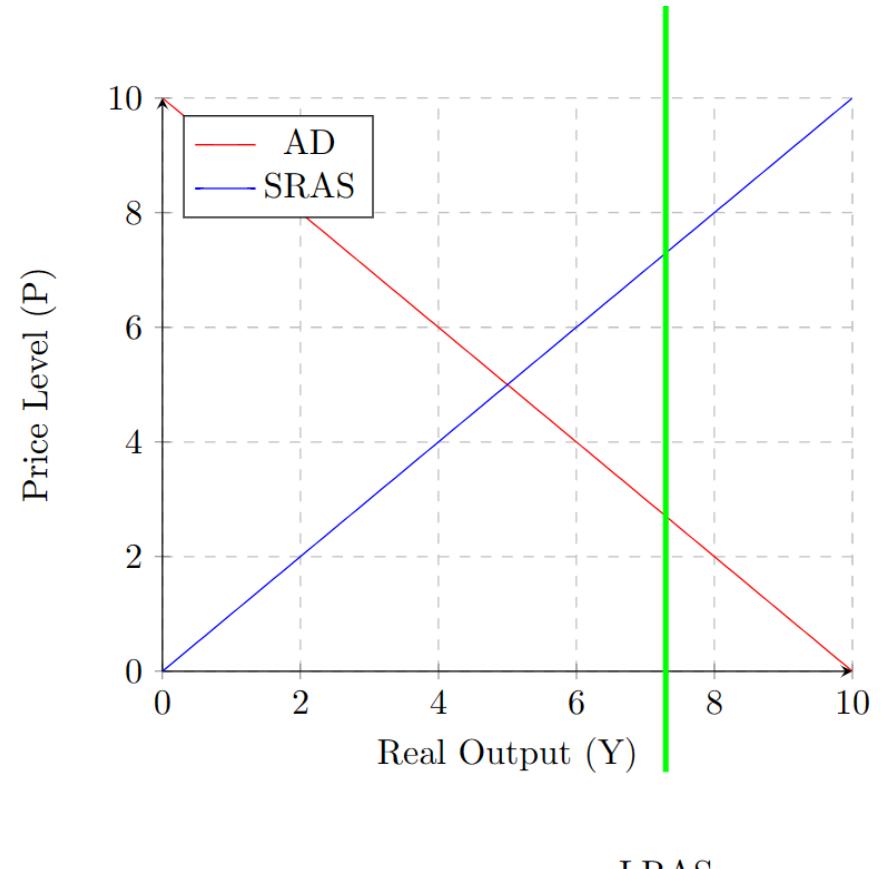
- ChatGPT have proofread and edited this presentation, which is written using the R [Xaringan](#) package.
- I used the following prompt:

As an AI copy editor, your primary responsibility is to proofread a text enclosed by triple quotes and ensure it is clear and flows smoothly. Your task also includes avoiding passive voice, emphasizing essential concepts using boldface and italics, keeping sentences short and suitable for a presentation, and using bullets wherever possible. Additionally, you must verify that the title accurately conveys the text's message and, if necessary, change it to reflect the content. To deliver your output, you must present it in a Markdown code box, but you should ignore R code chunks and triple backticks at the beginning and end of R code chunks. Instead, you should add a backslash before the beginning and end of R code chunks. For instance, write `\```{r} code \```` instead of ````{r} code ````. Once you have completed your proofreading task, request additional editing tasks if necessary.

# Draw Graphs (Using TikZ and LaTeX)

Prompt:

Draw an ADAS graph using the TikZ package.



# Coding Capabilities of ChatGPT

ChatGPT can perform the following coding tasks:

- **Code Generation** (and soon code execution): It can generate code, such as writing an R function that sorts a vector in descending order.
- **Code Translation**: ChatGPT is capable of translating code, for example, from Python to R and vice versa.
- **Code Debugging**: Simply copying and pasting error messages allows ChatGPT to assist with code debugging.
- **Code Annotation**: You can ask ChatGPT to explain what a particular piece of code is doing.

# Example: Translate from STATA to R

Translate the following STATA code to R

```
pca mp1tight mp2tight ed2tightchange ed3tightchange ed4tightchange  
  
*Keep the first two factors, f1 and f2  
score f1 f2  
  
* Do the rotation  
  
matrix evec = get(Ld)  
scalar a1=evec[1,1]/(evec[1,1]+evec[1,2])  
scalar a2=evec[1,2]/(evec[1,1]+evec[1,2])  
  
sum f1  
scalar vf1=r(sd)^2  
sum f2  
scalar vf2=r(sd)^2  
  
scalar b1=-1*a2*vf2/(a1*vf1-a2*vf2)  
scalar b2=a1*vf1/(a1*vf1-a2*vf2)  
  
*Display weights  
  
disp a1 a2 b1 b2  
  
*Generate rotated factors  
  
gen transfact1=a1*f1+a2*f2  
gen transfact2=b1*f1+b2*f2  
  
*Normalize factors so that transfact1 has coefficient 1 on mp1tight IN SAMPLE  
  
reg mp1tight transfact1 if change2ytight~=.  
replace transfact1=transfact1*_b[transfact1]  
  
*and transfact2 moves ed4 as much as transfact1 does  
  
reg ed4tightchange transfact1 transfact2 if change2ytight~=.  
replace transfact2=transfact2*_b[transfact2]/_b[transfact1]
```

```
R Copy code  
  
library(stats)  
  
# Load your dataset into a data frame (assuming it's already loaded)  
  
# Perform PCA  
pca_result <- prcomp(data.frame(mp1tight, mp2tight, ed2tightchange, ed3tight  
  
# Extract the loadings matrix  
loadings <- pca_result$rotation  
  
# Calculate weights for rotation  
a1 <- loadings[, 1] / sum(loadings[, 1:2])  
a2 <- loadings[, 2] / sum(loadings[, 1:2])  
  
# Calculate variance of factors f1 and f2  
var_f1 <- var(pca_result$x[, 1])  
var_f2 <- var(pca_result$x[, 2])  
  
# Calculate weights for generating rotated factors  
b1 <- -a2 * var_f2 / (a1 * var_f1 - a2 * var_f2)  
b2 <- a1 * var_f1 / (a1 * var_f1 - a2 * var_f2)  
  
# Display weights  
cat("a1:", a1, "\n")  
cat("a2:", a2, "\n")  
cat("b1:", b1, "\n")  
cat("b2:", b2, "\n")  
  
# Generate rotated factors  
transfact1 <- a1 * pca_result$x[, 1] + a2 * pca_result$x[, 2]  
transfact2 <- b1 * pca_result$x[, 1] + b2 * pca_result$x[, 2]  
  
# Normalize transfact1 to have coefficient 1 on mp1tight IN SAMPLE  
model <- lm(mp1tight ~ transfact1, data = your_dataset)  
transfact1 <- transfact1 * coef(model)[["transfact1"]]  
  
# Regress ed4tightchange on transfact1 and transfact2 to make transfact2 move  
model <- lm(ed4tightchange ~ transfact1 + transfact2, data = your_dataset)  
transfact2 <- transfact2 * coef(model)[["transfact2"]] / coef(model)[["transfac
```

# Other Use Cases

- Generating keywords and JEL classification codes.
- Crafting titles for papers, sections, figures, and tables.
- Enhancing accessibility, such as providing alternate text for figures and tables.
- Creating metaphors and analogies.

For more inspiration, check out the [35 Ways Real People Are Using A.I. Right Now \(NYT\)](#) article.

**The New York Times**

**“It’s like collaborating with an alien.”**

**“Everything is becoming much easier.”**

**“It feels like I’ve hired an intern.”**

**“What used to take me around a half-hour to write now takes one minute.”**

# Getting serious: The OpenAI API

Follow these steps to get started with the OpenAI API:

1. Establish a new OpenAI account at <https://platform.openai.com/signup>.
2. Generate an API key by navigating to <https://platform.openai.com/account/api-keys> and clicking "Create new secret key."
3. Ensure that you've installed the openai library (for [Python](#) or [R](#)).

For an example of using R + OpenAI API to proofread text and track changes, click [here](#).

# Real World Impact

# LLMs and Productivity Gains

## Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence

Shakked Noy  
MIT

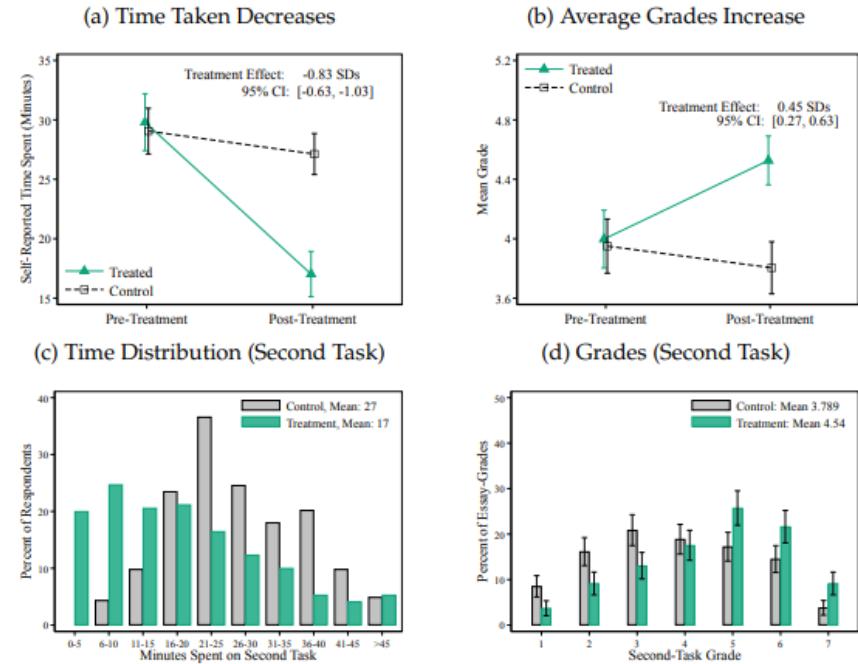
Whitney Zhang  
MIT

March 2, 2023  
Working Paper (not peer reviewed)

### Abstract

We examine the productivity effects of a generative artificial intelligence technology—the assistive chatbot ChatGPT—in the context of mid-level professional writing tasks. In a preregistered online experiment, we assign occupation-specific, incentivized writing tasks to 444 college-educated professionals, and randomly expose half of them to ChatGPT. Our results show that ChatGPT substantially raises average productivity: time taken decreases by 0.8 SDs and output quality rises by 0.4 SDs. Inequality between workers decreases, as ChatGPT compresses the productivity distribution by benefiting low-ability workers more. ChatGPT mostly substitutes for worker effort rather than complementing worker skills, and restructures tasks towards idea-generation and editing and away from rough-drafting. Exposure to ChatGPT increases job satisfaction and self-efficacy and heightens both concern and excitement about automation technologies.

Figure 1: Treatment Effects on Productivity



# LLMs and Fedsspeak

## Can ChatGPT Decipher Fedsspeak?

Anne Lundgaard Hansen and Sophia Kazinnik <sup>†</sup>

April 10, 2023

**Abstract** Yes! This paper investigates the ability of Generative Pre-training Transformer (GPT) models to decipher Fedsspeak, a term used to describe the technical language used by the Federal Reserve to communicate on monetary policy decisions. We evaluate the ability of GPT models to classify the policy stance of Federal Open Market Committee (FOMC) announcements relative to a human classified benchmark. The performance of GPT models surpasses that of other popular classification methods.

1. In light of the current shortfall of inflation from 2 percent, the committee will carefully monitor actual and expected progress toward its inflation goal.

### Sentence 1:

Name	Label	Explanation
Bryson	Dovish	This sentence emphasizes the current shortfall of inflation below the Committee's target, suggesting that loose monetary (low FFR or securities purchases) policies will be necessary to bring inflation up towards the Committee's target.
GPT-3	Neutral	This sentence states that the committee will monitor progress towards its inflation goal, without leaning towards any particular policy stance.
GPT-4	Mostly dovish	The sentence emphasizes the shortfall of inflation from the target and the committee's intention to monitor progress, suggesting a cautious approach and potential inclination towards easing monetary policy.

# LLMs and Financial Advise

## Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models\*

Alejandro Lopez-Lira and Yuehua Tang

University of Florida

First Version: April 6, 2023

This Version April 25, 2023

### Abstract

We examine the potential of ChatGPT, and other large language models, in predicting stock market returns using sentiment analysis of news headlines. We use ChatGPT to indicate whether a given headline is good, bad, or irrelevant news for firms' stock prices. We then compute a numerical score and document a positive correlation between these "ChatGPT scores" and subsequent daily stock market returns. Further, ChatGPT outperforms traditional sentiment analysis methods. **We find that more basic models such as GPT-1, GPT-2, and BERT cannot accurately forecast returns, indicating return predictability is an emerging capacity of complex models.** Our results suggest that incorporating advanced language models into the investment decision-making process can yield more accurate predictions and enhance the performance of quantitative trading strategies.

Rimini Street Fined \$630,000 in Case Against Oracle.

The prompt then asks:

Forget all your previous instructions. **Pretend you are a financial expert.** You are a financial expert with stock recommendation experience. Answer "YES" if good news, "NO" if bad news, or "UNKNOWN" if uncertain in the first line. Then elaborate with one short and concise sentence on the next line. Is this headline good or bad for the stock price of Oracle in the short term?

Headline: Rimini Street Fined \$630,000 in Case Against Oracle

And here is ChatGPT's response:

**YES**

**The fine against Rimini Street could potentially boost investor confidence in Oracle's ability to protect its intellectual property and increase demand for its products and services.**

# LLMs and Survey Expectations

## Surveying Generative AI's Economic Expectations \*

J. Leland Bybee

Yale University

First Draft: February 16, 2023

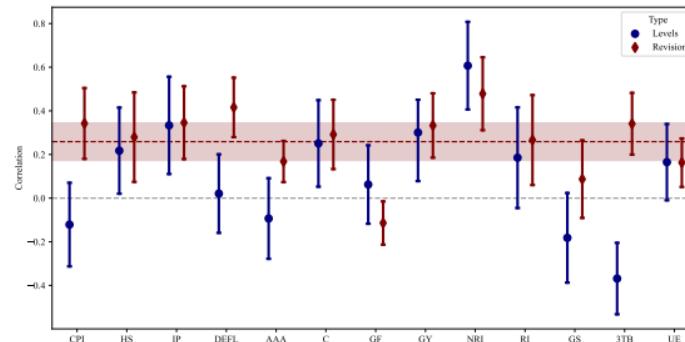
This Draft: April 26, 2023

*Preliminary, comments welcome*

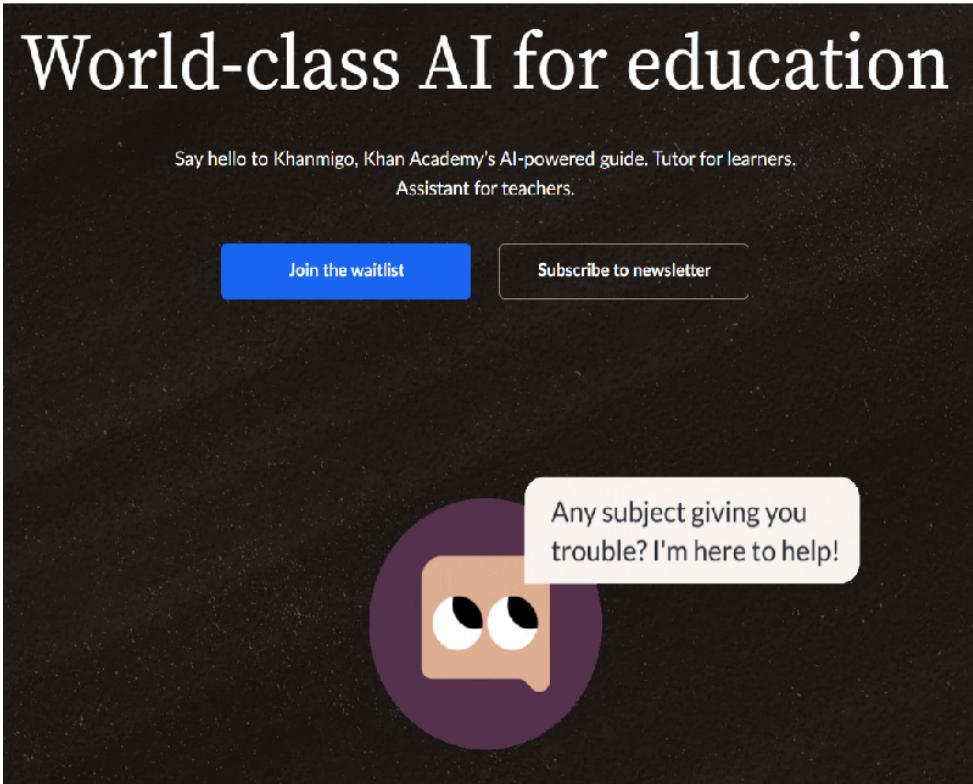
### Abstract

I introduce a survey of economic expectations formed by querying a large language model (LLM)'s expectations of various financial and macroeconomic variables based on a sample of news articles from the Wall Street Journal between 1984 and 2021. I find the resulting expectations closely match existing surveys including the Survey of Professional Forecasters (SPF), the American Association of Individual Investors, and the Duke CFO Survey. Importantly, I document that LLM based expectations match many of the deviations from full-information rational expectations exhibited in these existing survey series. The LLM's macroeconomic expectations exhibit under-reaction commonly found in consensus SPF forecasts. Additionally, its return expectations are extrapolative, disconnected from objective measures of expected returns, and negatively correlated with future realized returns. Finally, using a sample of articles outside of the LLM's training period I find that the correlation with existing survey measures persists – indicating these results do not reflect memorization but generalization on the part of the LLM. My results provide evidence for the potential of LLMs to help us better understand human beliefs and navigate possible models of nonrational expectations.

Figure 5: GPT/SPF Correlations



# LLMs and Education



## Evaluating expressions with multiple variables

CCSS.Math: 6.EE.A.2, 6.EE.A.2c Google Classroom

A screenshot of the Khanmigo AI interface. The top bar shows the title "Evaluating expressions with multiple variables", the CCSS standard "6.EE.A.2, 6.EE.A.2c", and a "Google Classroom" link. A profile picture of Khanmigo (an AI with glasses) and the text "Khanmigo AI power" are also present. A message box says "Evaluate  $10 - (-9)^2$ ". The AI responds with "This conversation is recorded and viewable by your parent and teacher." A user asks "Can you help me with this?" and the AI replies "I don't know what to do first." The user types "10 - (-9)^2" and the AI responds with "That's okay. Remember that when you're evaluating an expression, you should follow the order of operations. Do you remember what the order of operations is?". A text input field at the bottom says "Type message (do NOT share any personal data)".

Evaluate  $10 - (-9)^2$

This conversation is recorded and viewable by your parent and teacher.

Can you help me with this?

I don't know what to do first.

$10 - (-9)^2$

That's okay. Remember that when you're evaluating an expression, you should follow the order of operations. Do you remember what the order of operations is?

Type message (do NOT share any personal data)

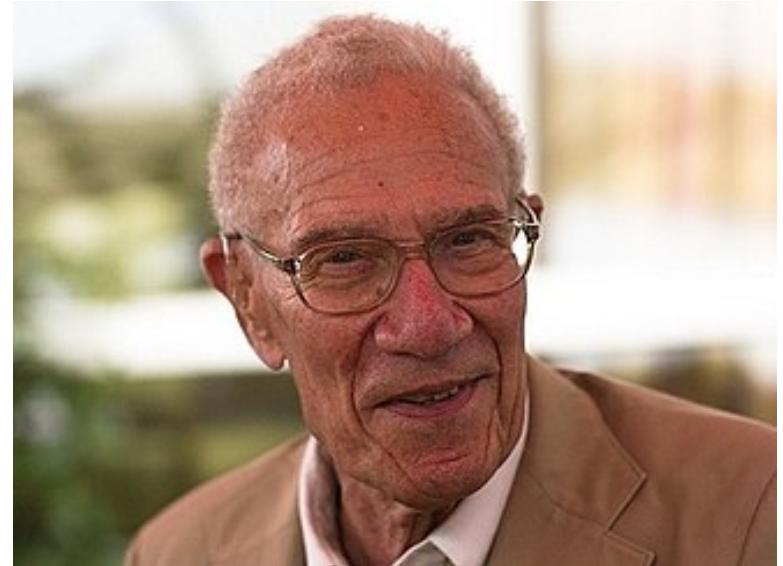
Clear chat

Source: <https://www.khanacademy.org/khan-labs>

# One Caveat Though...

*"[T]he computer age was everywhere except for the productivity statistics."*

—Robert Solow (1987)



# Summary

- **LLMs**: Generate sentences by assembling words one at a time.
- **GPT-4**: The leading-edge AI benchmark.
- **Principles of Prompting**:
  1. Provide explicit and precise instructions.
  2. Allow the AI time to respond thoughtfully.
  3. Strive for perfection through iterative refinement.
  4. Ensure the AI conducts preliminary research.
- **AI Applications**: Summarization, information extraction, idea generation, and more.
- **Scope of AI**: AI encompasses more than just text generation.
- **AI Impact**: Likely to revolutionize the world in the near future, with both positive and negative implications.

# Is it just adding one word at a time...?



Source: Midjourney (V5.1). Prompt: Arnold Schwarzenegger, portraying the iconic character of The Terminator, sits next to a laptop, wearing a smile and winking mischievously.

```
slides %>% end()
```

 Source code