



University of Antwerp
| Faculty of Arts

Generative AI in Academic Research

Opportunities and Pitfalls of using large language models

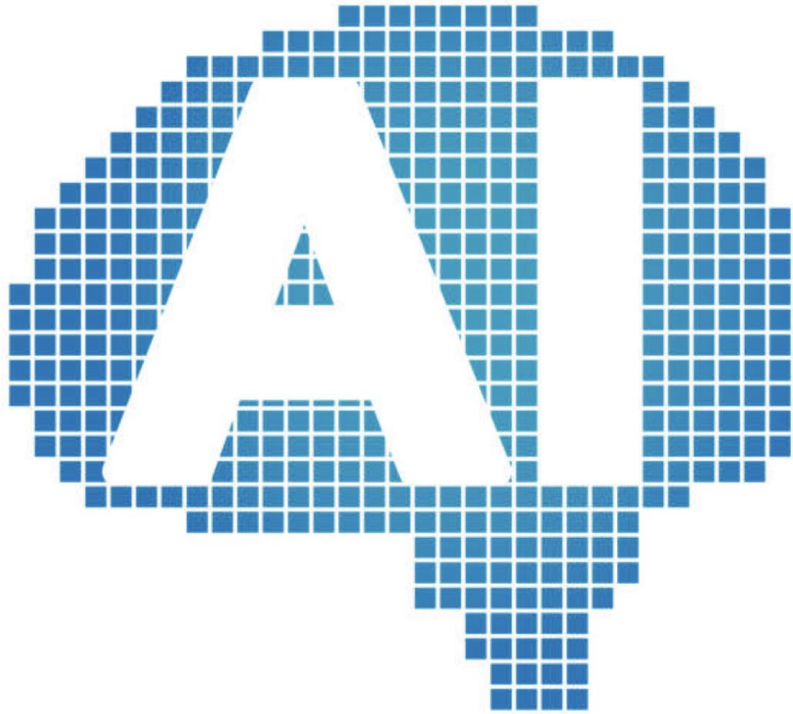
Luna De Bruyne, CLiPS

ADS Doctoral Day

16 & 17 October 2025

A few words on AI

Artificial intelligence



= computer systems that perform tasks which usually need human intelligence

→ more specifically: software that is very good in **recognizing patterns**

Artificial Intelligence



Not just large language models!

Image recognition

Modeling protein
structure

Forecasting

Machine translation

Autonomous systems

Artificial Intelligence > GenAI

- **GenAI:** AI for generating new content (text, images, audio, video, code, ...) that resembles human-created data
- First became popular through text-based models (ChatGPT)
→ **Large Language Models**

GenAI > LLMs

ChatGPT



- < Natural language processing
- < Computational linguistics

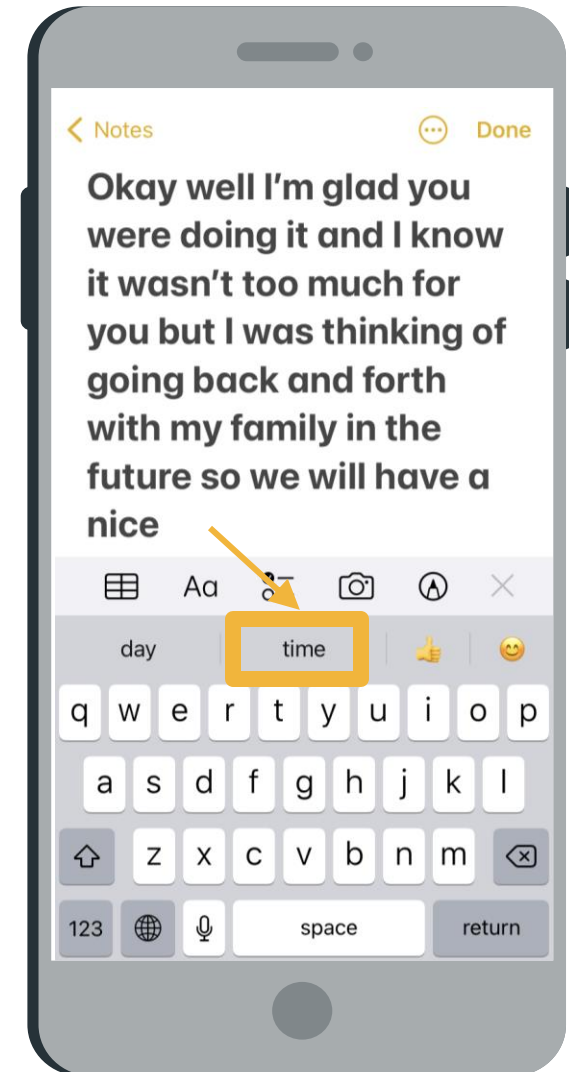
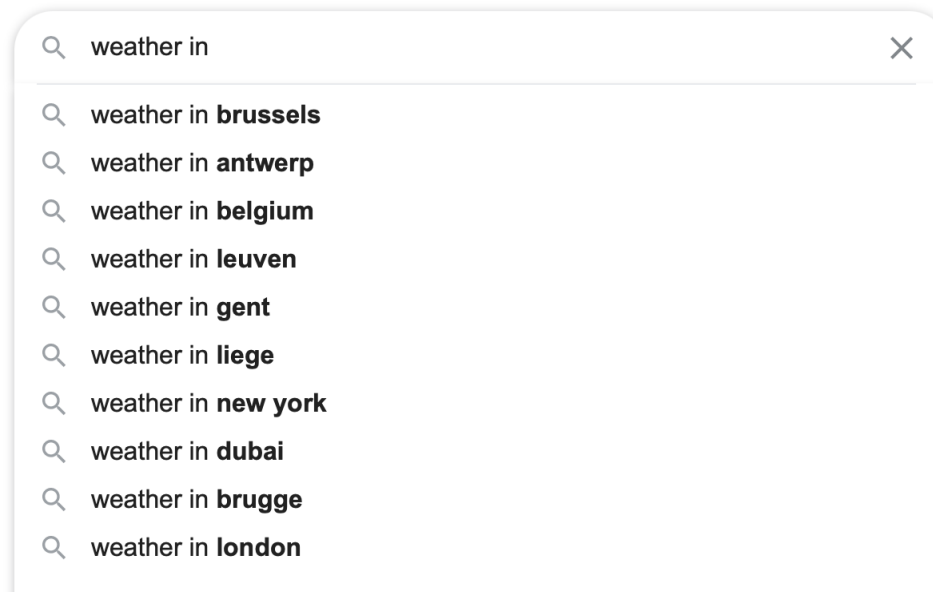


University of Antwerp
CLIPS | Centre for Computational
Linguistics and Psycholinguistics

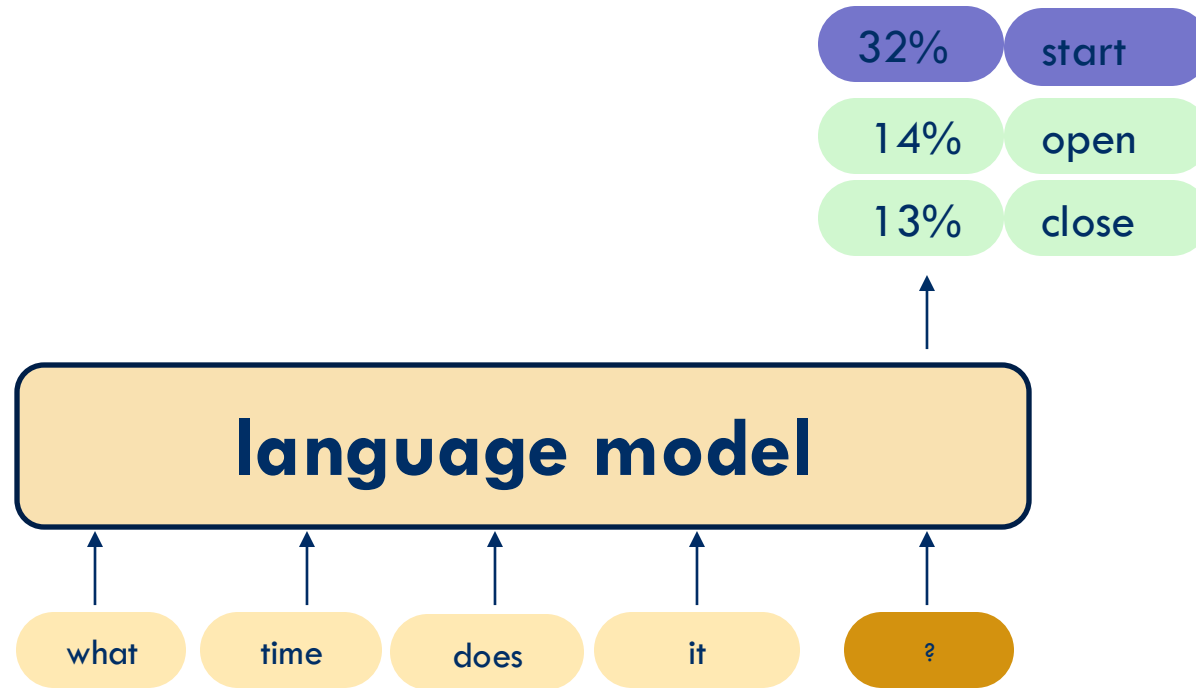
A few words on **LLMs**

Language model

- **Language model:** Models the probability of a word or sequences of words
→ predict words based on these probabilities

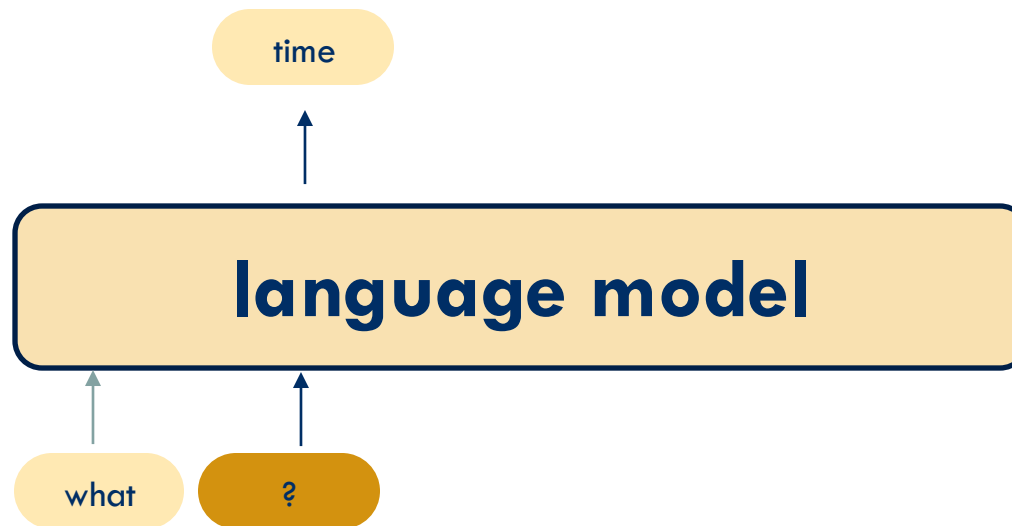


Language model



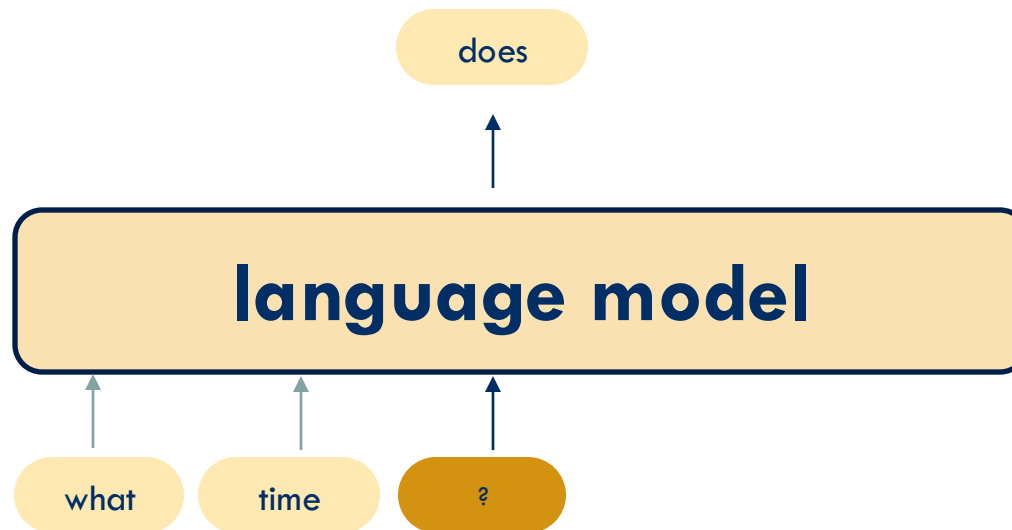
Language model: Autoregression

- Language models like ChatGPT are **autoregressive**: the model predicts one word at a time, based on all previous words (that it has generated)



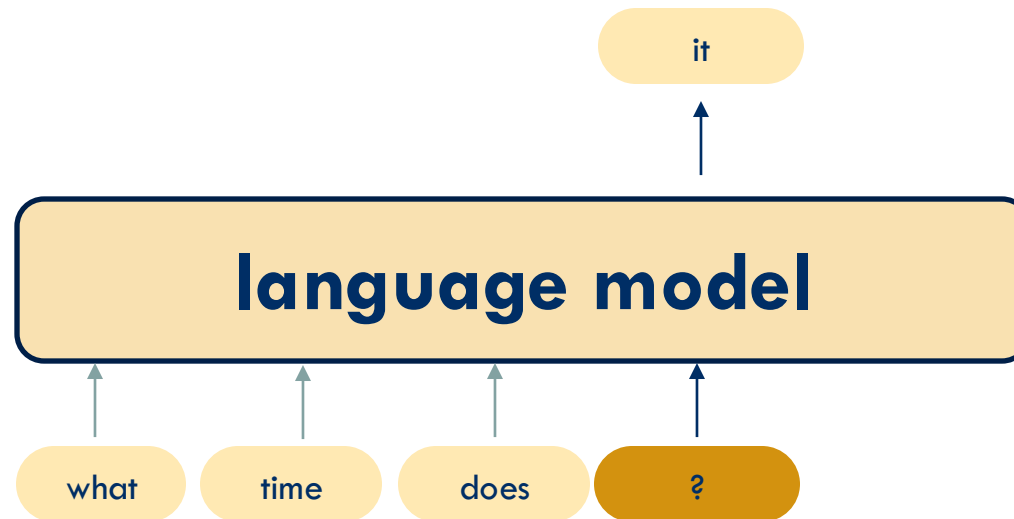
Language model: Autoregression

- Language models like ChatGPT are **autoregressive**: the model predicts one word at a time, based on all previous words (that it has generated)



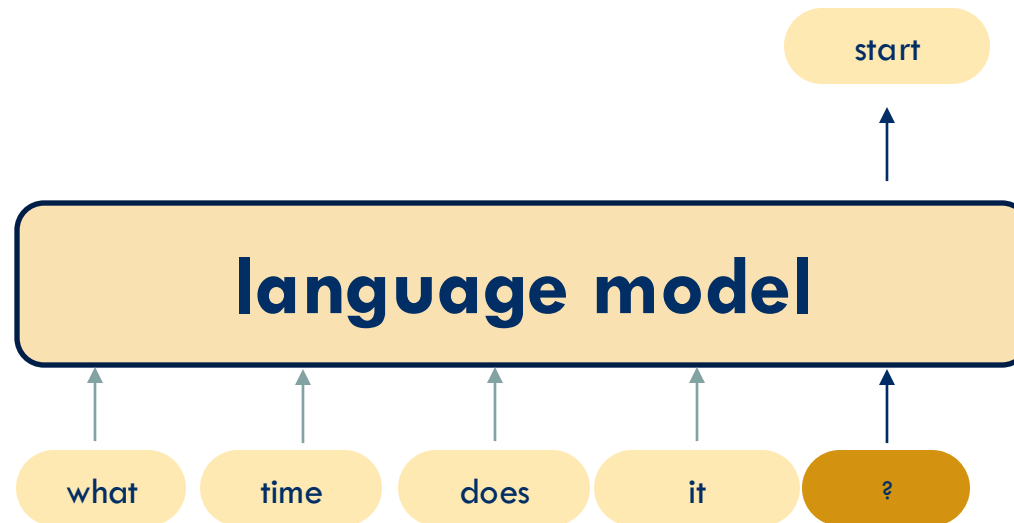
Language model: Autoregression

- Language models like ChatGPT are **autoregressive**: the model predicts one word at a time, based on all previous words (that it has generated)



Language model: Autoregression

- Language models like ChatGPT are **autoregressive**: the model predicts one word at a time, based on all previous words (that it has generated)

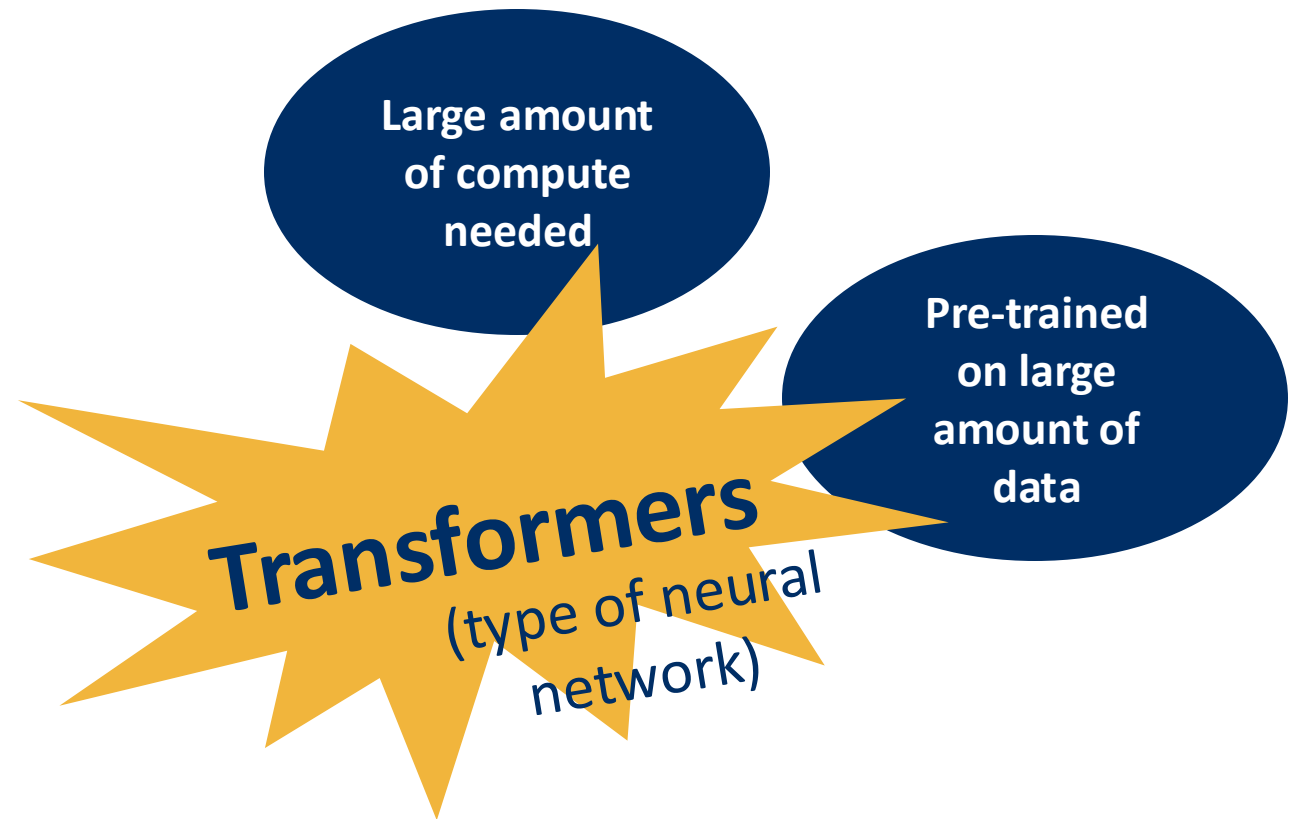


Language model: Autoregression

				what
			what	time
		what	time	does
	what	time	does	it
what	time	does	it	start

How does ChatGPT work?

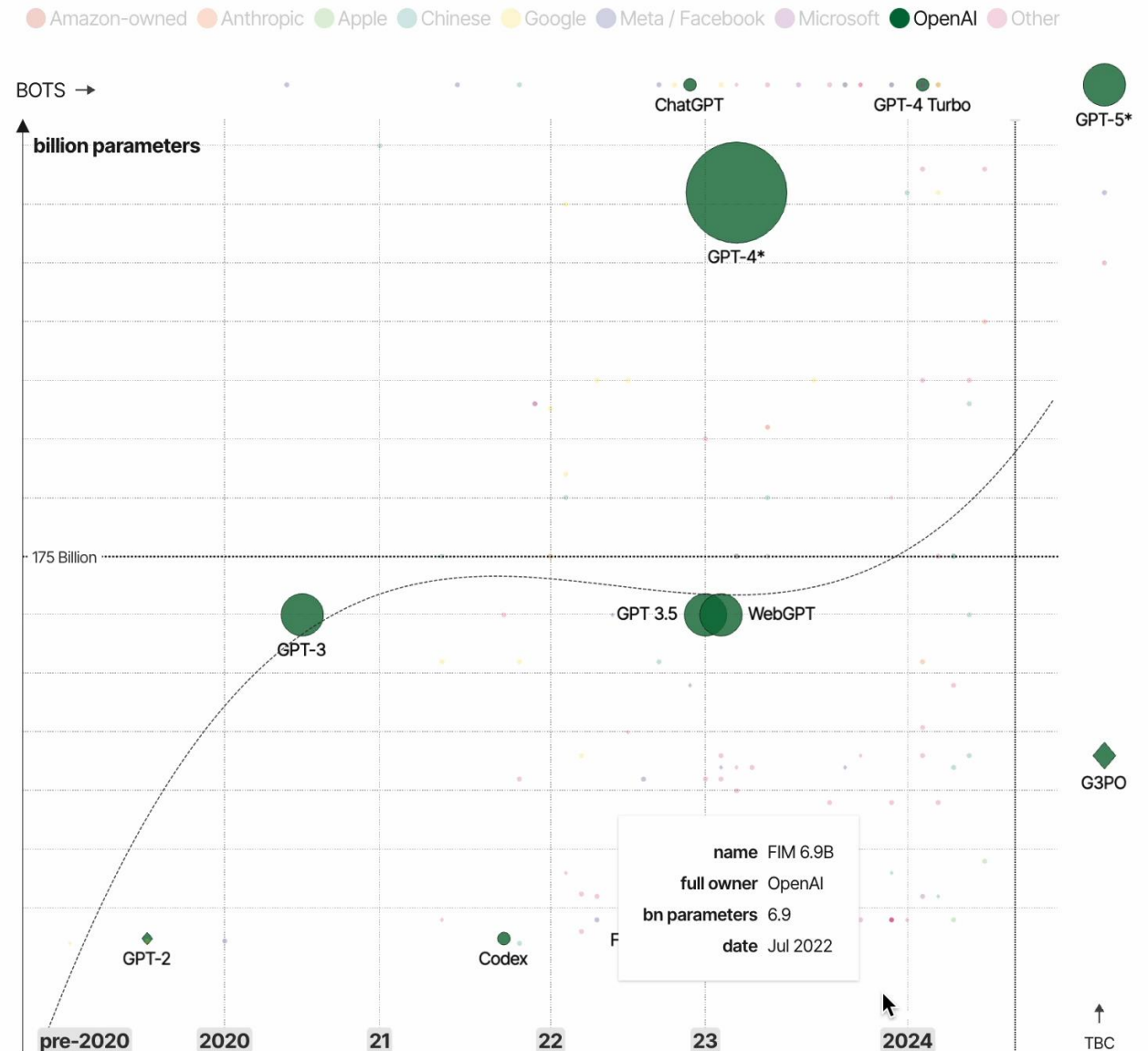
- ChatGPT < LLMs



Scale

- Bigger and bigger models
- Trained on more and more data
- Check <https://informationisbeautiful.net/visualizations/the-rise-of-generative-ai-large-language-models-llms-like-chatgpt/>

The Rise and Rise of A.I. Large Language Models (LLMs) & their associated bots like ChatGPT



David McCandless, Tom Evans, Paul Barton
Information is Beautiful // UPDATED 20th Mar 24

source: news reports, [LifeArchitect.ai](#)
* = parameters undisclosed // see [the data](#)

Scale

- LLMs are trained on much more data than a human could ever read in a lifetime

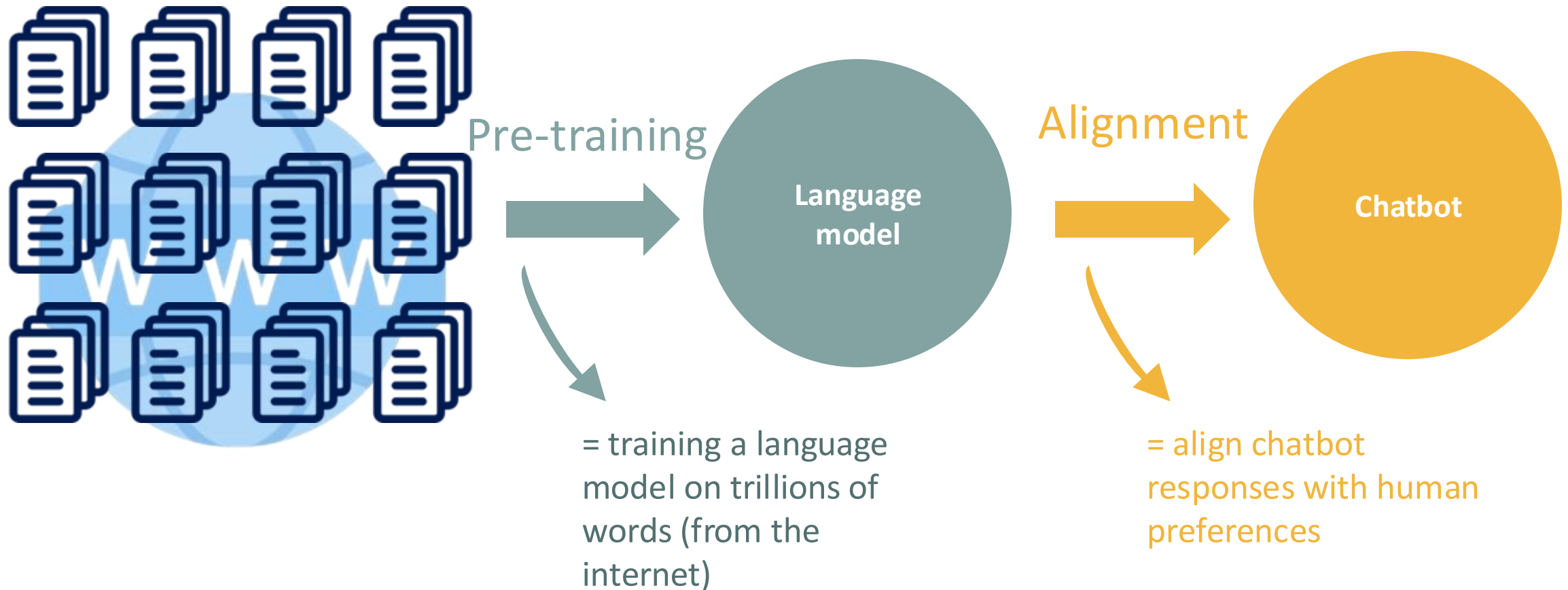
Llama 2:
trained on
2 trillion tokens



- Humans read 300 words per minute
- If human would read non-stop for 80 years
→ 0.06% of Llama 2



From language model to chatbot



Alignment

- Aligning chatbot with human preferences
- Based on
 - Supervised finetuning: humans response to prompts
 - Language model is trained further on pairs of prompt and response
 - Optimise based on human feedback: language model generated multiple responses and human ranks according to preference
 - Language model gets updated in order to generate preferred response

GenAI in **Research**

How NOT to do it

PROBLEMS OF LAND POLLUTION

ОА Брунич - E-Scio, 2023 - cyberleninka.ru

... **As an AI language model**, I am programmed to agree that stopping pollution is necessary to preserve the environment and protect human health. Here are some ways we can stop ...

☆ Enregistrer Citer Autres articles >>

Design and Implementation of Smart Hydroponics Farming for Growing Lettuce Plantation under Nutrient Film Technology

M Venkatraman, R Surendran - 2023 2nd International ..., 2023 - ieeexplore.ieee.org

... **As an AI language model**, there is no access to the specific database details of any particular research study. However, in general, a well-designed database for a hydroponics system ...

☆ Enregistrer Citer Autres articles

[PDF] ieee.org

Assessing the Impact of Climate Factors on Sea Ice Extent with ML Regression

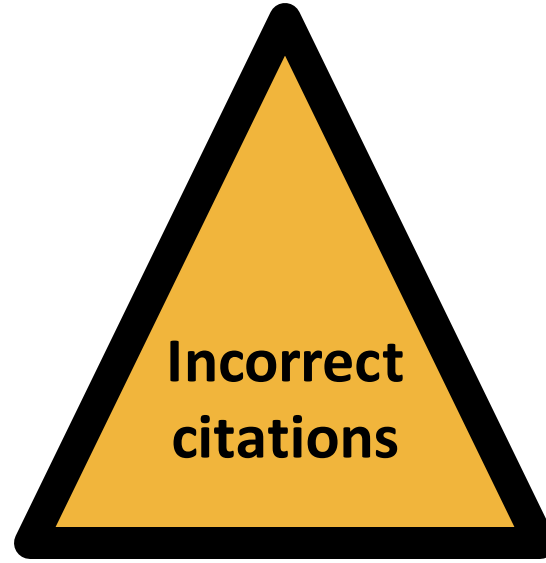
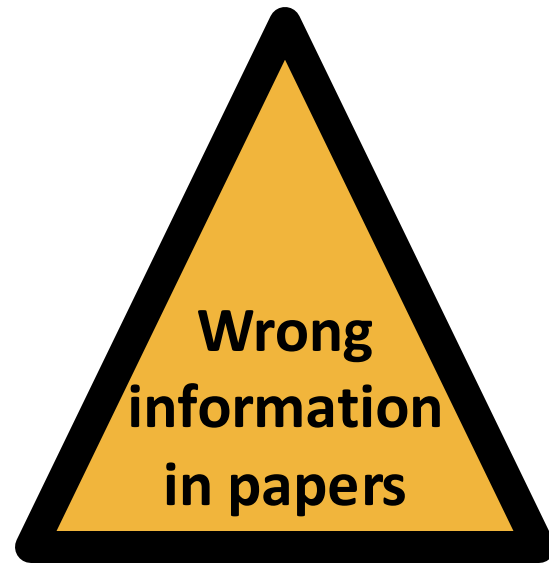
Z Ashraf, K Fatima - Multi-Disciplinary Journal for Early Stage ..., 2023 - nuzmsol.com

... **As an AI language model**, I don't have direct access to real-time data or databases, including specific datasets. However, I can provide information about commonly used Arctic sea ice ...

☆ Enregistrer Citer Les 2 versions >>

[PDF] nuzmsol.com

Pitfalls



Opportunities



**Focus more on
approach / process!**

Opportunities pt.1

(the usual suspects)

Opportunities (the usual suspects)

- Use ChatGPT

as a
brainstorm
partner

to
understand
papers

as a writing
assistant

ChatGPT as a brainstorm partner

- Be specific!
- Give context (research domain, what type of grant, ...)
- Build incrementally (engage in a back-and-forth way to refine ideas)
- Use it to explore multiple perspectives
- ...

I'm applying for a grant on how AI can enhance healthcare delivery in low-resource settings. I need help framing the research gaps and potential solutions.



Let's outline a strategy:

1. Identify Research Gaps:

- Are there limitations in current healthcare technologies that AI can address? For example, is there a lack of accurate diagnostic tools for certain diseases?
- What are the barriers to implementing AI in low-resource settings, such as infrastructure, training, or affordability?

2. Define Potential Solutions:

- AI tools that work offline or require minimal computational power.
- Algorithms designed for low-resolution imaging or noisy data.

3. Emphasize Impact:

- How would AI reduce disparities in healthcare delivery?
- Can AI reduce the workload on overburdened healthcare workers?

Would you like me to help draft specific research questions or identify case studies to support your proposal?

ChatGPT as a brainstorm partner

Find some highly cited papers in the field of human-computer interaction which study the effect of linguistic/pragmatic characteristics in chatbots on user interaction.



Gezocht op 4 sites ▾

Here are some highly cited papers that investigate the linguistic and pragmatic characteristics of chatbots in human-computer interaction, focusing on their effects on user interaction:

1. **"Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions" (Araujo, 2018)** - This study explores how informal language styles and anthropomorphic cues in chatbots affect user perceptions, attitudes, and satisfaction. It found that chatbots employing a casual tone or using anthropomorphic design significantly enhanced user experience by making interactions feel more personal and engaging AR5IV RESEARCHGATE.
2. **"Chatbots' language design: The influence of language variation on user experience"** - This paper investigates how variations in chatbot language styles (e.g., modern vs. Shakespearean English) affect usability and enjoyment. Users found modern English chatbots easier to use, while Shakespearean style added a fun element to the experience. It highlights the importance of context-appropriate language in shaping user satisfaction AR5IV.
3. **"How should my chatbot interact? A survey on social characteristics in human-chatbot interaction design"** - This survey compiles research on social and linguistic characteristics of chatbots, analyzing how conversational style, proactivity, and social intelligence contribute to user engagement and perceptions of chatbot quality AR5IV.

- ChatGPT for finding research papers
- Don't trust everything, always check!

- Alternatives



...

Perplexity

- = Search + GPT
- Again: don't trust everything, always check

What are some highly cited papers in the field of human-computer interaction which study the effect of linguistic/pragmatic characteristics in chatbots on user interaction.

🔍 Sources

ChatGPT: perspectives from human-computer interaction and ...

 pmc.ncbi.nlm.nih

linguistic nuances - 2025 CASE STUDY - Computer Science Cafe

 computersciencecafe

Navigating Human-Chatbot Interactions: An Investigation into ...

 tandfonline



Show all

✳️ Perplexity

Several highly cited papers in the field of human-computer interaction study the effect of linguistic and pragmatic characteristics in chatbots on user interaction:

1. Ischen et al. (2020) examined the role of entity, interactivity, and experiential perceptions in chatbot persuasion ⁷. This study highlights the importance of these factors in shaping user experiences with chatbots.
2. Adam et al. (2021) investigated the effects of AI-based chatbots in customer service on user compliance ⁷. Their research provides insights into how chatbot interactions influence user behavior in service contexts.
3. Crollic et al. (2021) studied the impact of anthropomorphism on customer anger in chatbot interactions ⁷. This paper reveals how human-like characteristics in chatbots can affect user emotions and overall interaction quality.

Elicit

- Find papers
- Get summary
- Ask questions

 Review

 Andreas Ströhle

 *Journal of Neural Transmission* 

2008 1025 Citations

[Semantic Scholar](#) 

[DOI](#) 

Abstract summary

Exercise training may be clinically effective, at least in major depression and panic disorder.

What did they test?

The intervention is exercise training.

What outcomes did they measure?

• Prevalence And Incidence Of Depression

 What were the limitations of this study|

Ask

ChatGPT for understanding papers

Consider this paragraph:

"We call our particular attention "Scaled Dot-Product Attention" (Figure 2). The input consists of queries and keys of dimension d_k , and values of dimension d_v . We compute the dot products of the query with all keys, divide each by $\sqrt{d_k}$, and apply a softmax function to obtain the weights on the values.

In practice, we compute the attention function on a set of queries simultaneously, packed together into a matrix Q . The keys and values are also packed together into matrices K and V . We compute the matrix of outputs as: $\text{Attention}(Q, K, V) = \text{softmax}(QK^T/\sqrt{d_k})V$ (1)

The two most commonly used attention functions are additive attention [2], and dot-product (multiplicative) attention. Dot-product attention is identical to our algorithm, except for the scaling factor of $1/\sqrt{d_k}$. Additive attention computes the compatibility function using a feed-forward network with a single hidden layer.

While the two are similar in theoretical complexity, dot-product attention is much faster and more space-efficient in practice, since it can be implemented using highly optimized matrix multiplication code.

While for small values of d_k the two mechanisms perform similarly, additive attention outperforms dot product attention without scaling for larger values of d_k [3]. We suspect that for large values of d_k , the dot products grow large in magnitude, pushing the softmax function into regions where it has extremely small gradients. To counteract this effect, we scale the dot products by $1/\sqrt{d_k}$."

Explain in a simplified way.



This paragraph explains a concept from machine learning called **Scaled Dot-Product Attention**, which is a key idea in modern AI models like Transformers.

1. What is Scaled Dot-Product Attention?

Scaled Dot-Product Attention is a way to determine the importance of different pieces of input data relative to each other. It's commonly used in language models for tasks like translating a sentence or understanding context.

2. How Does It Work?

The attention mechanism uses:

- **Queries (Q)**: Representing the item we're trying to find context for.
- **Keys (K)**: Representing all possible items to compare against.
- **Values (V)**: Representing the actual information we want to use if the attention focuses on a specific comparison.

These have their own dimensions:

- Queries and keys share a dimension of size d_k .
- Values have a dimension of size d_v .

The process works like this:

1. Compute a **dot product** between the query and each key. This measures how much focus or "attention" should be given to that part of the input.
2. Divide these dot products by $\sqrt{d_k}$ to prevent very large numbers from interfering with computation.
 - Without this scaling, very large dot products could make the softmax function behave poorly.
3. Apply a **softmax function** to these scaled dot products to turn them into probabilities (weights) between 0 and 1.
4. Use these weights to scale the values V and compute the final attention output.

■ Alternatives: NotebookLM

NotebookLM

- Uses Gemini (Google's LLM)
- Get summary
- Ask questions
- Get audio overview (podcast)



Attention Is All You Need

SettingsShare

Add note

* Notebook guide

Help me create

FAQ

Study guide

Table of contents

Timeline

Briefing doc

Summary

This research paper introduces the Transformer, a novel neural network architecture for sequence transduction tasks like machine translation. **Unlike previous models relying on recurrent or convolutional layers, the Transformer uses solely attention mechanisms**, achieving superior performance and faster training times. The paper details the Transformer's architecture, including **multi-head self-attention and positional encoding**, and presents results demonstrating **state-of-the-art accuracy on English-to-German and English-to-French translation**, as well as successful application to English constituency parsing. The authors also **analyse various model variations and provide attention visualizations to illustrate the model's inner workings**. Finally, the code for the model is made publicly available.

Audio overview

Attention Is All You Need

00:10 / 14:40

Suggested questions

How does the Transformer model improve upon existing sequence transduction models?

What are the key architectural innovations of the Transformer model?

What are the model's performance results and their implications?

View chat

1 source

Explain scaled dot-product attention in a simplified way

→

* Notebook guide

NotebookLM can make mistakes, so double-check it.

ChatGPT as a writing assistant

- Use ChatGPT to
 - revise or paraphrase your texts
 - give title suggestions
 - proofread (grammar and style checks)
 - adjust style or tone
 - ...
- Alternatives:
 - Any other LLM (Gemini, Llama, Claude, Mistral, Grok, Deepseek, ...)
 - Tools in which LLMs are integrated, like Grammarly, Wordtune, ...

Grammarly

- Proofreading
- Giving suggestions
- Adjusting style or tone
- Integrated while writing
- Different platforms (even Overleaf)

The screenshot displays the Grammarly web interface. The main text area contains the following paragraph: "Every cookie at the Kevin Cookie Company is crafted with only the finest ingredients. Butter from grass fed cows. Unbleached flour with the best flavor and textures. Organic sugar only from India. Eggs from chickens that listen to classical music. Even our water is special. It comes from the Cascade Mountain springs that we hike up to retrieve. Baking is our passion and we take it very seriously. If you've never had cookies from the Kevin Cookie Company, you're missing out. Visit us at the KevinCookieCompany.com to order your cookies today!". The word "special" is highlighted with a yellow circle, and a suggestion box is open, showing "special" with the option to "Choose a different word". Other suggestions include "grass fed" (Add a hyphen), "and" (Add a comma), and "the KevinCo..." (Correct article usage). The right sidebar includes a "HIDE ASSISTANT" button, a "Goals" section, a "GrammarlyGO" button, and an "All suggestions" button. The bottom status bar shows a green checkmark, "Corrected", "Undo", and "Plagiarism" icons. The word count is 91 words.

Untitled document

Every cookie at the Kevin Cookie Company is crafted with only the finest ingredients. Butter from grass fed cows. Unbleached flour with the best flavor and textures. Organic sugar only from India. Eggs from chickens that listen to classical music. Even our water is special. It comes from the Cascade Mountain springs that we hike up to retrieve. Baking is our passion and we take it very seriously. If you've never had cookies from the Kevin Cookie Company, you're missing out. Visit us at the KevinCookieCompany.com to order your cookies today!

4 All suggestions

- grass fed · Add a hyphen
- special · Choose a different word
- and · Add a comma
- the KevinCo... · Correct article usage

HIDE ASSISTANT >>

Goals >

GrammarlyGO

All suggestions

Corrected Undo

Plagiarism

91 words

Wordtune

- Proofreading
- Giving suggestions
- Adjusting style or tone
- Rewriting
- Integrated while writing

Superior Web Solutions produces highest quality interactive web sites, logos, and fresh animators. As you notice our specialization is professional web site design and maintenance.

Superior Web Solutions services all aspects of your company's goals; log term and short term will be estimated annually to assure that you are up tot he minute with the latest information and technologies available.

All your internet solutions can be provided at an affordable and realistic price. If you'd like a web site with all qualities mentioned at a bargain price please feel free to contact me.

The screenshot shows the Wordtune web interface. At the top, there is a toolbar with icons for undo, redo, and other editing functions. Below the toolbar, three rewritten sentences are displayed in a list. The first sentence is highlighted in light blue, the second in light purple, and the third in light green. At the bottom of the interface, there is a status bar indicating the number of daily rewrites left and a link to go unlimited.

Please feel free to contact me if you would like a website with all the qualities mentioned above at a bargain price.

Let me know if you are interested in a website with all the qualities mentioned at a bargain price.

I can build you a website with all the qualities listed at a bargain price.

4 daily Rewrites left. [Go unlimited](#)

ChatGPT is a lot more!



GPT-3.5

Next word
predictor with
human alignment



GPT-4

+ larger
+ image



GPT-4o

+ larger
+ multimodal
(image, video,
audio, ...)



GPT-4o1

+ larger
+ multimodal
(image, video,
audio, ...)
+ “reasoning”

ChatGPT is a lot more!

- The current web-based ChatGPT
 - surfs the internet
 - generates images
 - runs programming code

Opportunities pt.2

(one step further)

Opportunities (one step further)

- Use ChatGPT

for content
analysis

to automate
processes

for data
visualisation

ChatGPT for content analysis

- Content analysis = determining the presence of certain words / themes / concepts within a text
- Common research method in the humanities (e.g., analysing social media or news texts)
- Can be useful in other disciplines as well (e.g., for research trends analysis, meta analysis)

Consider this abstract of an academic article. Identify whether it comes from a theoretical paper, an experimental paper, or a review paper. Answer only with the label (either "theoretical", "experimental" or "review").

Abstract: "Neural Machine Translation (NMT) continues to improve in quality and adoption, yet the in advertent perpetuation of gender bias remains a significant concern. Despite numerous studies on gender bias in translations into English from weakly gendered-languages, there are no benchmarks for evaluating this phenomenon or for assessing mitigation strategies. To address this gap, we introduce GATE X-E, an extension to the GATE (Rarrick et al., 2023) corpus, that consists of human translations from Turkish, Hungarian, Finnish, and Persian into English. Each translation is accompanied by feminine, masculine, and neutral variants. The dataset, which contains between 1250 and 1850 instances for each of the four language pairs, features natural sentences with a wide range of sentence lengths and domains, challenging translation rewriters on various linguistic phenomena. Additionally, we present a translation gender rewriting solution built with GPT-4 and use GATE X-E to evaluate it. We open source our contributions to encourage further research on gender debiasing."



experimental



ChatGPT for content analysis

- Better to do this using the OpenAI API
- Requires python coding (which ChatGPT can help you with ;))

```
import openai

# Set your OpenAI API key
openai.api_key = 'your_openai_api_key_here'

# Define the prompt
prompt = """
Consider this abstract of an academic article. Identify whether it comes from a theoretical or empirical study.
"""

Abstract: "Neural Machine Translation (NMT) continues to improve in quality and adoption."

completion = client.chat.completions.create(
    model="gpt-4o-mini",
    messages=[
        {"role": "system", "content": "You are a helpful assistant."},
        {"role": "user", "content": prompt}
    ],
)

y = json.loads(str(completion.model_dump_json()), strict=False)
response = y["choices"][0]["message"]["content"]

print(response)
```

ChatGPT to automate processes

- Many things can be automated using a basic python script
- If you don't have coding skills yourself, ChatGPT can help you writing the code!
- No python installed to run the code? Use [Google Colab](#)

Write python code to iterate through all files in a folder and merge the content to one big file called "merged.txt". In the merged.txt file, each individual file content is preceded by `*** [filename] ***` and two new lines, and followed by three new lines.



You can use the following Python script to iterate through all files in a specified folder, read their content, and merge them into one large file called `merged.txt`. Each file's content will be preceded by `*** [filename] ***` with two new lines and followed by three new lines as specified.

```
python Code kopiëren

import os

def merge_files_in_folder(folder_path, output_file):
    """
    Iterates through all files in the given folder and merges their content into a
    with formatting as specified.

    :param folder_path: Path of the folder containing files to merge
    :param output_file: Name of the output file
    """
    with open(output_file, "w", encoding="utf-8") as outfile:
        for filename in os.listdir(folder_path):
            file_path = os.path.join(folder_path, filename)

            # Ensure it's a file (skip directories)
            if os.path.isfile(file_path):
                try:
                    with open(file_path, "r", encoding="utf-8") as infile:
                        # Write the header with the file name
                        outfile.write(f"*** [{filename}] ***\n\n")
```

Google Colab

 Test.ipynb ☆

Bestand Bewerken Bekijken Invoegen Runtime Tools Hulp [Alle wijzigingen zijn opgeslagen](#)

+ Code + Tekst

 Genereren Write python code to iterate through all files in a folder and merge the content to one big file called "merged.txt". In the merged.txt file, each individual file content is preceded by *** [filename] *** and two new lines, and followed by three new lines.

< 1 of 1 > [Wijzigingen ongedaan maken](#) [Wees voorzichtig met code](#)

 De gegenereerde code kan onder een licentie vallen |

```
1 import os
2
3 def merge_files(folder_path, output_filename="merged.txt"):
4     """
5     Iterates through all files in a folder and merges their content into a single file.
6
7     Args:
8         folder_path: The path to the folder containing the files to merge.
9         output_filename: The name of the output file.
10    """
11    try:
12        with open(output_filename, "w") as outfile:
13            for filename in os.listdir(folder_path):
14                filepath = os.path.join(folder_path, filename)
15                if os.path.isfile(filepath):
16                    with open(filepath, "r") as infile:
17                        outfile.write(f"*** [{filename}] ***\n\n")
18                        outfile.write(infile.read())
19                        outfile.write("\n\n\n")
20                print(f"Files merged successfully into '{output_filename}'.")
21    except FileNotFoundError:
22        print(f"Error: Folder '{folder_path}' not found.")
23    except Exception as e:
24        print(f"An error occurred: {e}")
25
26
27 # Example usage
28 # Replace '/path/to/your/folder' with the actual path to your folder
29 merge_files('sample_data') # Example folder path, change to your folder path.
30
```

 Files merged successfully into 'merged.txt'.

ChatGPT for data visualisation

titanic.csv
Spreadsheet

Visualise this dataset

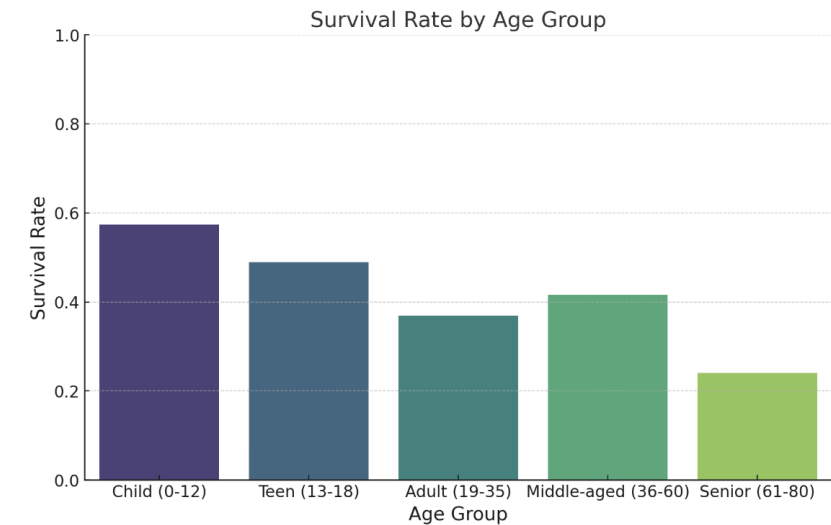


The dataset appears to be the Titanic passenger data. Key columns include:

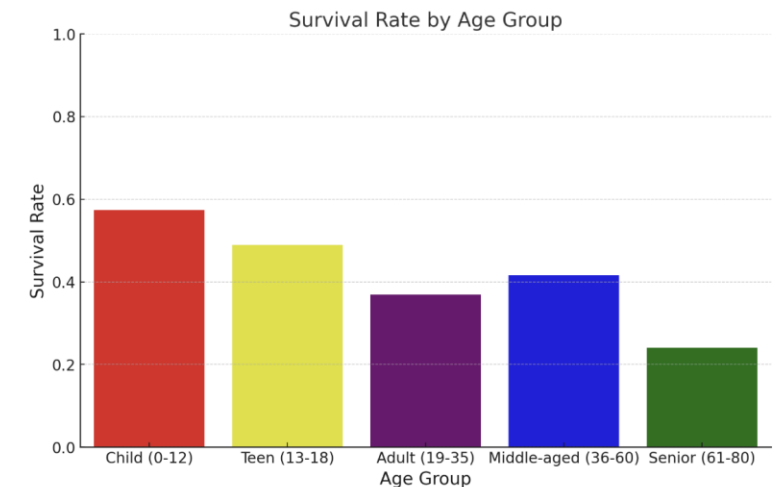
- **PassengerId**: Unique identifier for each passenger
- **Survived**: Survival status (0 = Did not survive, 1 = Survived)
- **Pclass**: Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
- **Name**: Passenger's name
- **Sex**: Gender
- **Age**: Age of the passenger
- **SibSp**: Number of siblings/spouses aboard
- **Parch**: Number of parents/children aboard
- **Ticket**: Ticket number
- **Fare**: Passenger fare
- **Cabin**: Cabin number
- **Embarked**: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

How would you like this dataset visualized? Some common options include:

- Survival rate analysis (e.g., by gender, class, or age group)
- Distribution of passenger demographics (age, fare)
- Embarkation point statistics [?]



Change the color scheme. The colors I want are red - yellow - purple - blue - green



The survival rate visualization now uses the specified color scheme: red, yellow, purple, blue, and green. Let me know if there's anything else you'd like to adjust! [?]



ChatGPT for data visualisation

- Alternatives:
 - Let ChatGPT generate the code and run in Colab
 - Data visualisation tools in which LLMs are integrated, e.g. Dataline.app

Write python code to make a bar chart from this data using plotly:

Survival rate (Y-axis):
Child (0-12): 0.58
Teen (13-18): 0.50
Adult (19-35): 0.38
Middle-aged (36-60): 0.41
Senior (61-80): 0.24

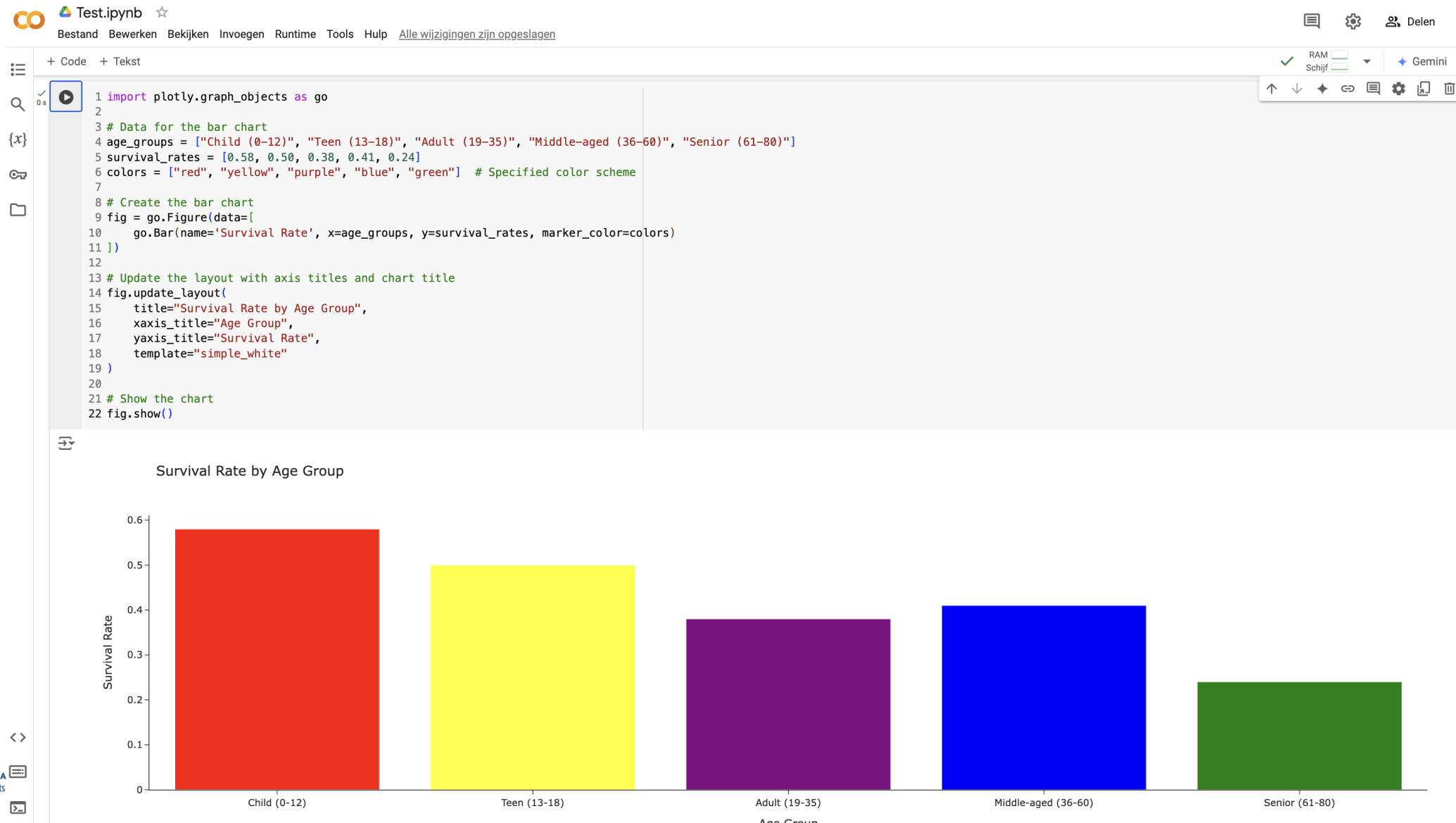
The X-axis is called Age group.
Use the following color scheme: red - yellow - purple - blue - green



To create a bar chart with Plotly using the provided data and the specified color scheme, you can use the following Python code:

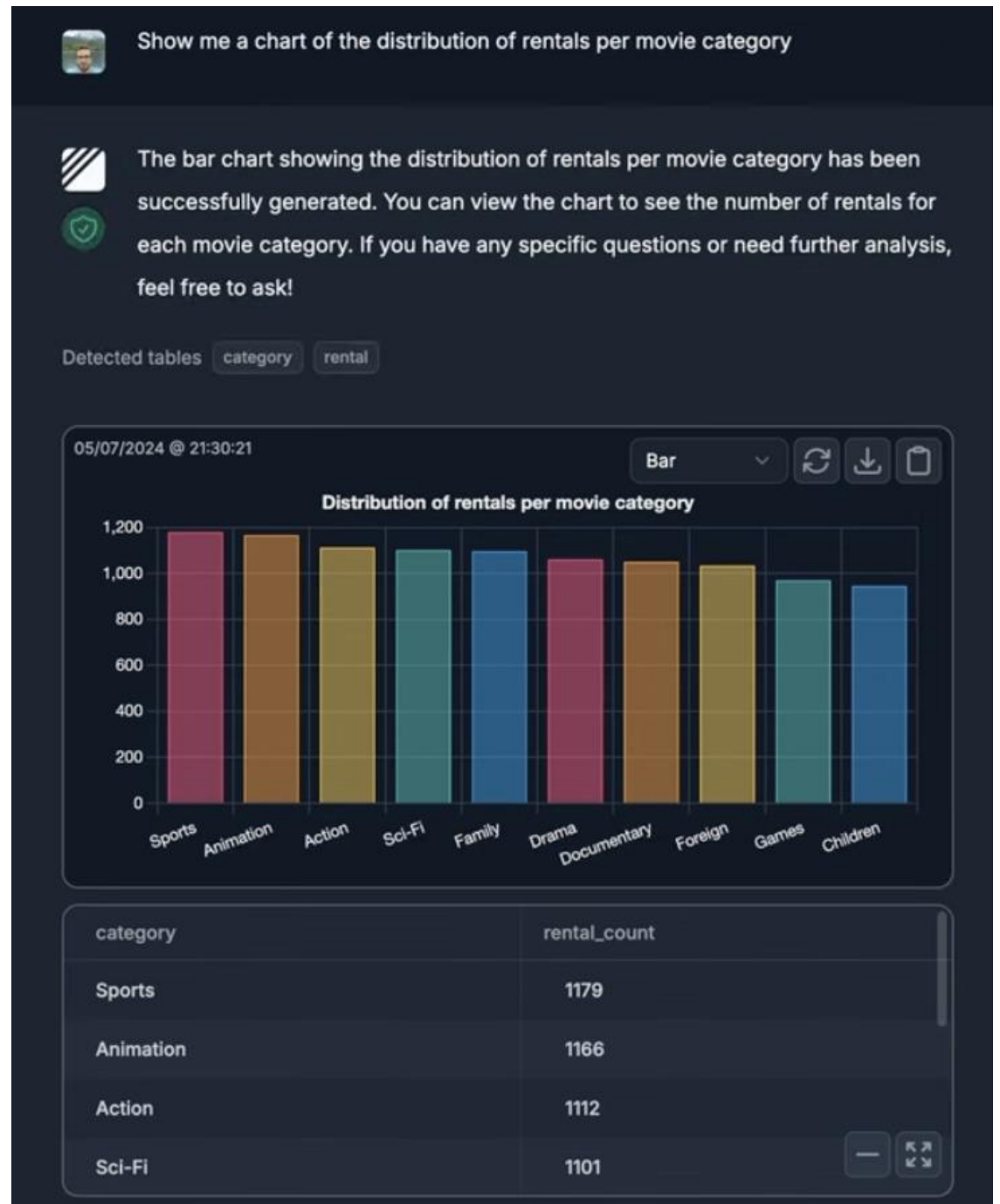
```
python Code kopiëren  
  
import plotly.graph_objects as go  
  
# Data for the bar chart  
age_groups = ["Child (0-12)", "Teen (13-18)", "Adult (19-35)", "Middle-aged (36-60)", "Senior (61-80)"]  
survival_rates = [0.58, 0.50, 0.38, 0.41, 0.24]  
colors = ["red", "yellow", "purple", "blue", "green"] # Specified color scheme  
  
# Create the bar chart  
fig = go.Figure(data=[  
    go.Bar(name='Survival Rate', x=age_groups, y=survival_rates, marker_color=colors)  
)  
  
# Update the layout with axis titles and chart title  
fig.update_layout(  
    title="Survival Rate by Age Group",  
    xaxis_title="Age Group",  
    yaxis_title="Survival Rate",  
    template="simple_white"  
)  
  
# Show the chart  
fig.show()
```

ChatGPT for data visualisation



Dataline.app

- Similar functionalities as in ChatGPT, but possible to keep data locally (only uses the metadata)
- The platform itself is free, but you need OpenAI API key



Prompt engineering

- **Optimize communication with ChatGPT: 3 most important tricks**



Provide
context

Give
examples

Provide
reasoning
steps

Prompt engineering: Context

- In the prompt
“You are an expert in ...”

- In the system prompt

```
completion = client.chat.completions.create(  
    model="gpt-4o-mini",  
    messages=[  
        {"role": "system", "content": "You are a helpful assistant."},  
        {"role": "user", "content": prompt}  
    ],  
)
```

- In the “custom instructions” (settings > personalisation > custom instructions)

Prompt engineering: Examples

You are a sentiment analysis system. Your task is to classify a sentence as either positive or negative.

Here are some examples:

Text: The service was awesome.

Label: positive

Text: My parcel arrived late.

Label: negative

Text: Thanks for the great service!

Label: positive

Now classify this sentence:

Text: Horrible customer service!

Label:



Label: negative



Prompt engineering: Reasoning

- Chain-of-thought prompting
- Combine with examples! (first reasoning, then example)

1.11 or 1.9, which one is bigger? Let's think step by step.



1.11 is bigger than 1.9 because, in decimal numbers, the comparison starts from the leftmost digit. Here's how it works:

- Compare the digits before the decimal point: Both have 1, so they are equal.
- Compare the digits immediately after the decimal point: 11 (in 1.11) is greater than 9 (in 1.9).

Thus, $1.11 > 1.9$.



Sure, let's carefully think through this step by step to compare 1.11 and 1.9:

1. **Look at the digits before the decimal point:**

- In 1.11, the whole number part is 1.
- In 1.9, the whole number part is 1.
- These are **equal**, so we need to compare further.

2. **Compare the digits after the decimal point (tenths place):**

- In 1.11, the first digit after the decimal is 1.
- In 1.9, the first digit after the decimal is 9.
- Since $1 < 9$, 1.9 is larger so far.

3. **Check further decimal places:**

- In 1.11, the second digit after the decimal is 1.
- In 1.9, there is no second digit, so it's effectively 0.

4. **Re-evaluate the comparison:**

- 1.11 has more precision, but 1.9 has a larger value in the tenths place.

Final Conclusion:

1.9 is **bigger** than 1.11.

Opportunities (?) pt.3

(taking the leap)

Agentic AI for scientific discovery

- Agent AI = autonomous systems that can solve complex, multi-step problems without human intervention
- “Capable of reasoning and planning”
- Usually in multi-agent set-up
- Can be used to make research practice more efficient

AGENTIC AI FOR SCIENTIFIC DISCOVERY: A SURVEY OF PROGRESS, CHALLENGES, AND FUTURE DIRECTIONS

Mourad Gridach, Jay Nanavati, Khaldoun Zine El Abidine, Lenon Mendes & Christina Mack
IQVIA
`{firstname.lastname}@iqvia.com`

ABSTRACT

The integration of Agentic AI into scientific discovery marks a new frontier in research automation. These AI systems, capable of reasoning, planning, and autonomous decision-making, are transforming how scientists perform literature review, generate hypotheses, conduct experiments, and analyze results. This survey provides a comprehensive overview of Agentic AI for scientific discovery, categorizing existing systems and tools, and highlighting recent progress across fields such as chemistry, biology, and materials science. We discuss key evaluation metrics, implementation frameworks, and commonly used datasets to offer a detailed understanding of the current state of the field. Finally, we address critical challenges, such as literature review automation, system reliability, and ethical concerns, while outlining future research directions that emphasize human-AI collaboration and enhanced system calibration.

Agentic AI

- **Human-AI collaborative systems:**

- Focused on accelerating research, enhancing experimental design, optimising decision making
- Still require manual validation and execution
- Example: accepts human-provided research ideas and assists with literature review, experimentation and report writing

- **Fully autonomous systems**

- Automating end-to-end workflows with minimal human intervention
- Efficient when tasks are well-defined, repetitive, or require high precision
- Focused on automating time-consuming processes

Agentic AI

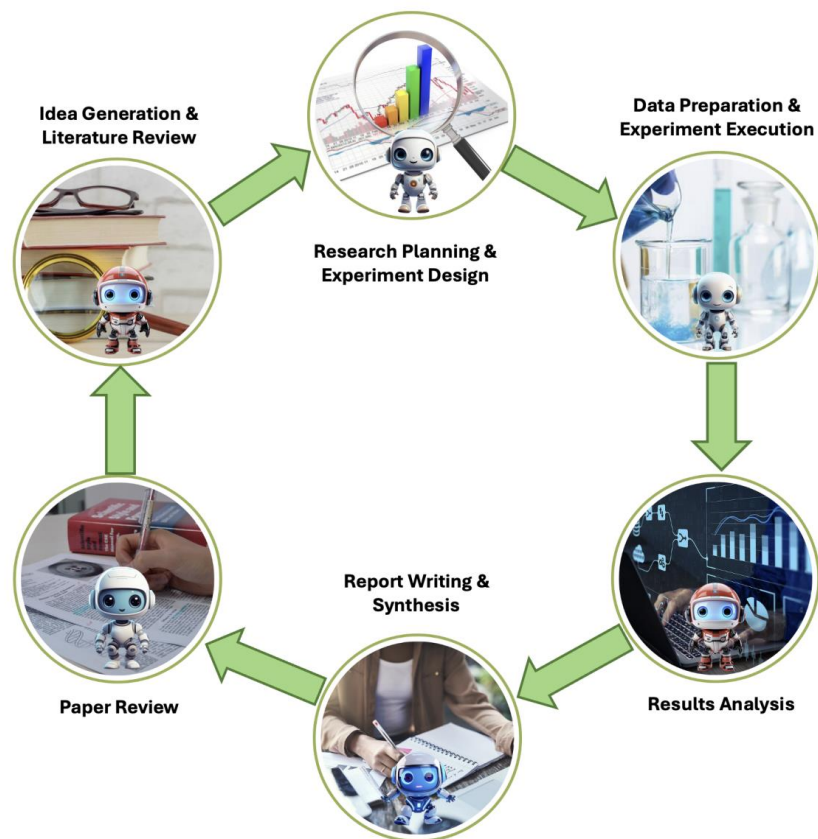


Figure 1: Agentic AI workflow for scientific discovery.

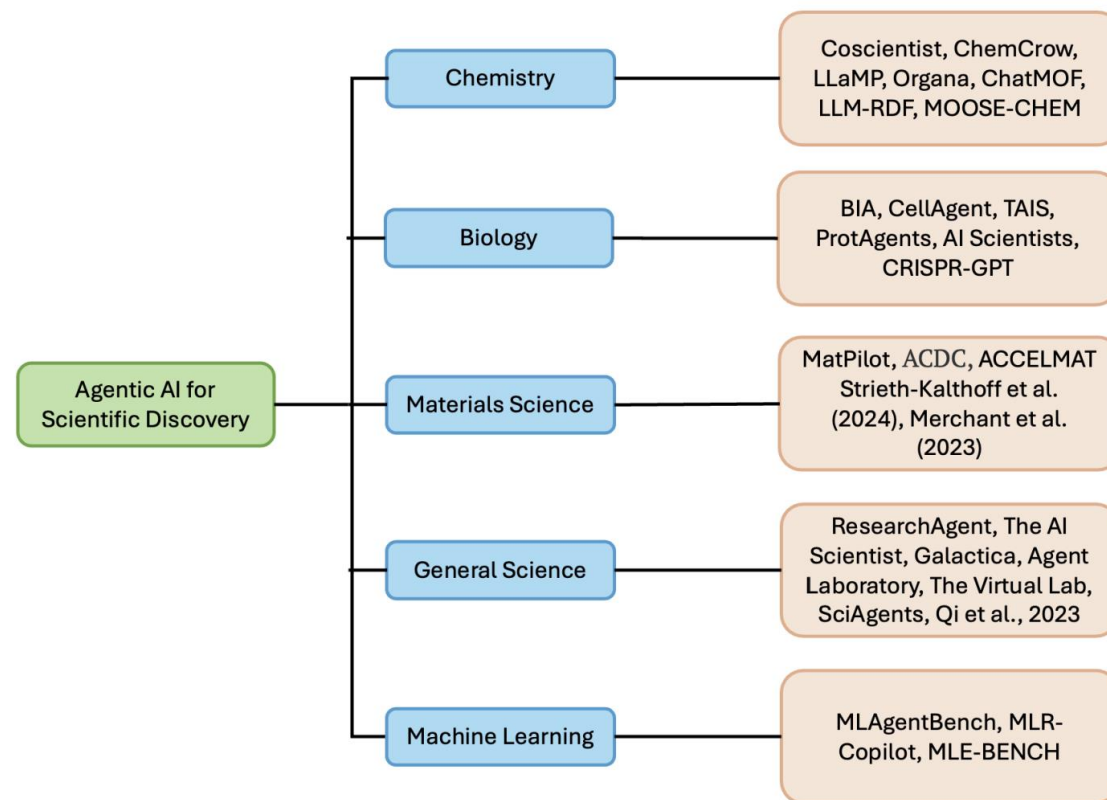


Figure 2: AI Agents frameworks for scientific discovery.

OpenAI's Deep Research

February 2, 2025 Release

Introducing deep research

An agent that uses reasoning to synthesize large amounts of online information and complete multi-step research tasks for you. Available to Pro users today, Plus and Team next.

Today we're launching deep research in ChatGPT, a new agentic capability that conducts multi-step research on the internet for complex tasks. It accomplishes in tens of minutes what would take a human many hours.

Deep research is OpenAI's next agent that can do work for you independently—you give it a prompt, and ChatGPT will find, analyze, and synthesize hundreds of online sources to create a comprehensive report at the level of a research analyst. Powered by a version of the upcoming OpenAI o3 model that's optimized for web browsing and data analysis, it leverages reasoning to search, interpret, and analyze massive amounts of text, images, and PDFs on the internet, pivoting as needed in reaction to information it encounters.

The ability to synthesize knowledge is a prerequisite for creating new knowledge. For this reason, deep research marks a significant step toward our broader goal of developing AGI, which we have long envisioned as capable of producing novel scientific research.

Google's AI Co-scientist

Accelerating scientific breakthroughs with an AI co-scientist

February 19, 2025 ·

Juraj Gottweis, Google Fellow, and
Vivek Natarajan, Research Lead

We introduce AI co-scientist, a multi-agent AI system built with Gemini 2.0 as a virtual scientific collaborator to help scientists generate novel hypotheses and research proposals, and to accelerate the clock speed of scientific and biomedical discoveries.

In the pursuit of scientific advances, researchers combine ingenuity and creativity with insight and expertise grounded in literature to generate novel and viable research directions and to guide the exploration that follows. In many fields, this presents a breadth and depth conundrum, since it is challenging to navigate the rapid growth in the rate of scientific publications while integrating insights from unfamiliar domains. Yet overcoming such challenges is critical, as evidenced by the many modern breakthroughs that have emerged from transdisciplinary endeavors. For example, Emmanuelle Charpentier and Jennifer Doudna won the [2020 Nobel Prize in Chemistry](#) for their work on [CRISPR](#), which combined expertise ranging from microbiology to genetics to molecular biology.

Motivated by unmet needs in the modern scientific discovery process and building on [recent AI advances](#), including the ability to synthesize across complex subjects and to perform [long-term planning and reasoning](#), we developed an [AI co-scientist system](#). The AI co-scientist is a multi-agent AI system that is intended to function as a collaborative tool for scientists. Built on [Gemini 2.0](#), [AI co-scientist is](#) designed to mirror the reasoning process underpinning the scientific method. Beyond standard literature review, summarization and “deep research” tools, the AI co-scientist system is intended to uncover new, original knowledge and to formulate demonstrably novel research hypotheses and proposals, building upon prior evidence and tailored to specific research objectives.

Google's AI Co-scientist

Accelerating scientific breakthroughs with an AI co-scientist

February 19, 2025 ·
Juraj Gottweis, Google Fellow, and
Vivek Natarajan, Research Lead

Validation of novel AI co-scientist hypotheses with real-world laboratory experiments

To assess the practical utility of the system's novel predictions, we evaluated end-to-end laboratory experiments probing the AI co-scientist-generated hypotheses and research proposals in three key biomedical applications: drug repurposing, proposing novel treatment targets, and elucidating the mechanisms underlying antimicrobial resistance. These settings all involved expert-in-the-loop guidance and spanned an array of complexities:

Application	<i>Drug repurposing</i>	<i>Novel treatment target discovery</i>	<i>Explain mechanism of gene transfer evolution</i>
Challenge	Combinatorial search	Identifying novel targets	Understanding complex systems
Complexity	Medium	High	Very high
Scale	Moderate, data-limited	Moderate, experiment-limited	Large, data and computation-limited
Unknown elements	Constrained	Large	Vast and dynamic

Google's AI Co-scientist

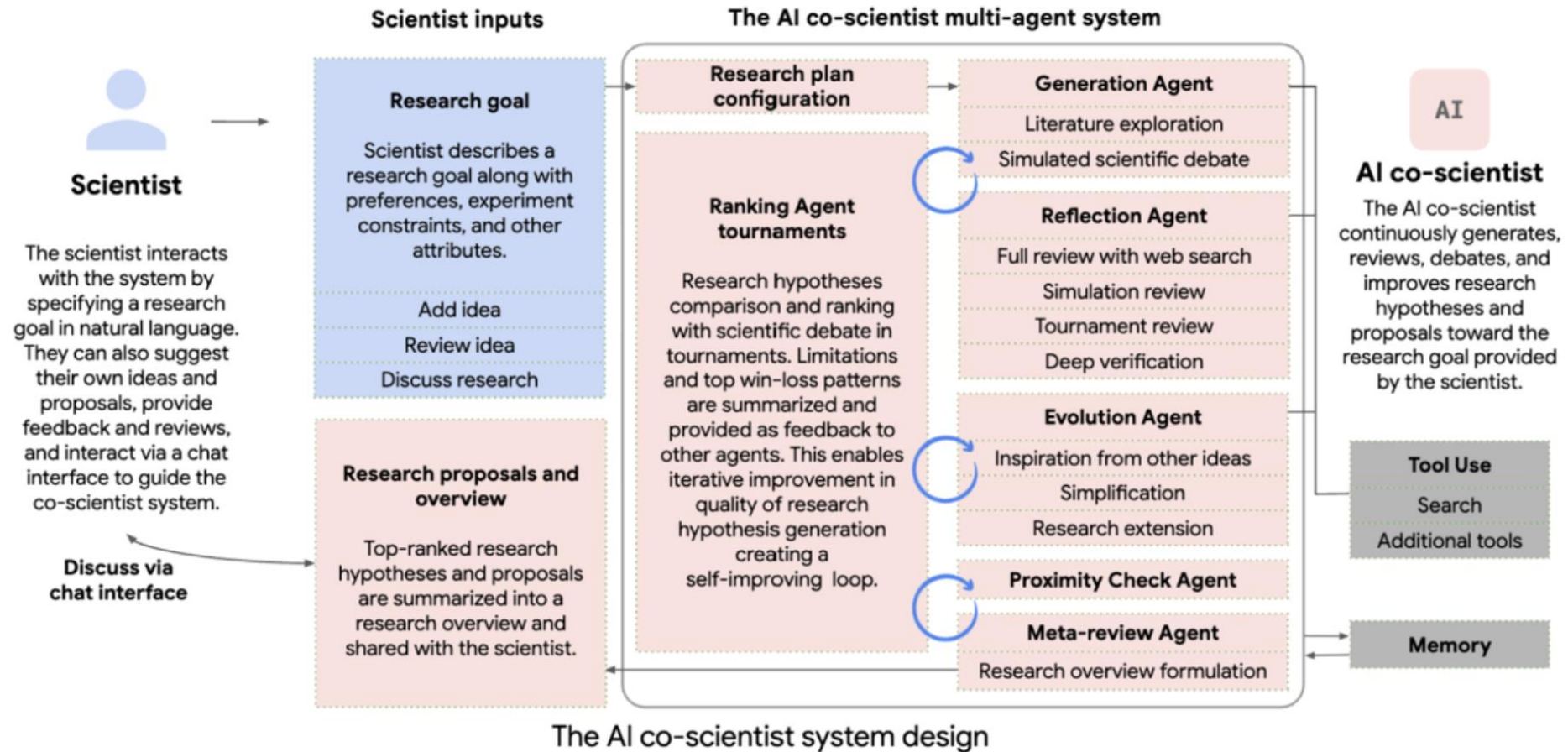


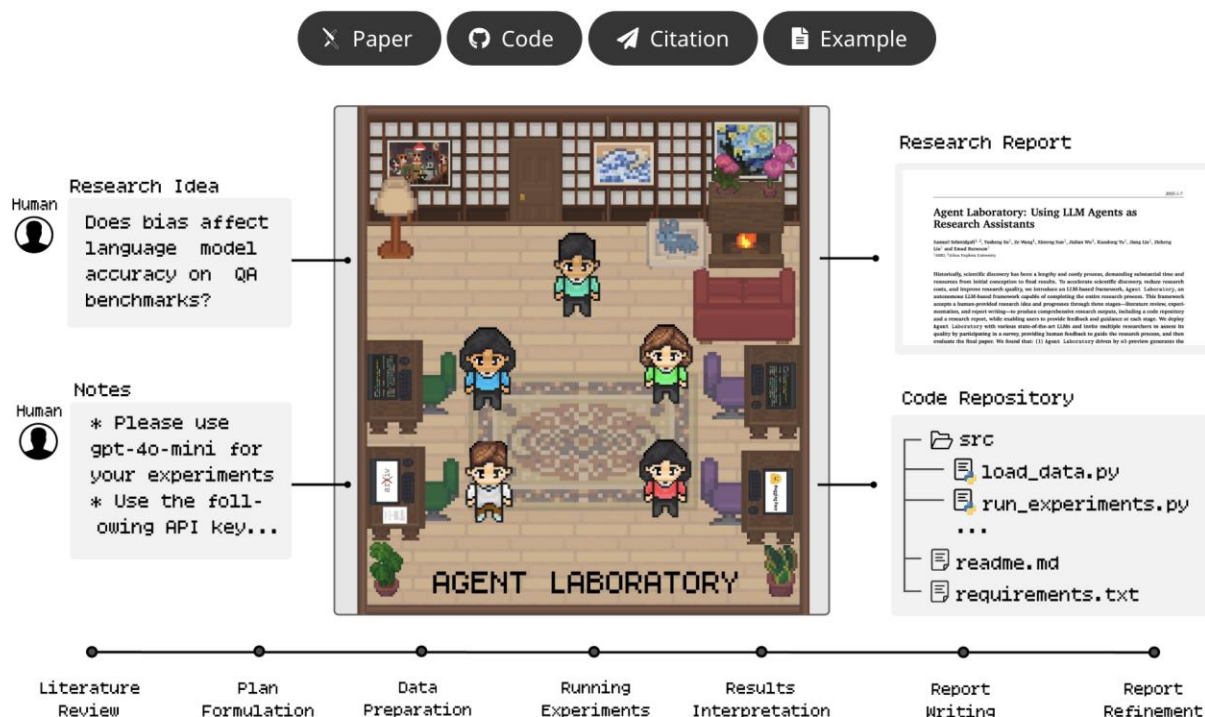
Illustration of the different components in the AI co-scientist multi-agent system and the interaction paradigm between the system and the scientist.

Agent Laboratory

Agent Laboratory: Using LLM Agents as Research Assistants

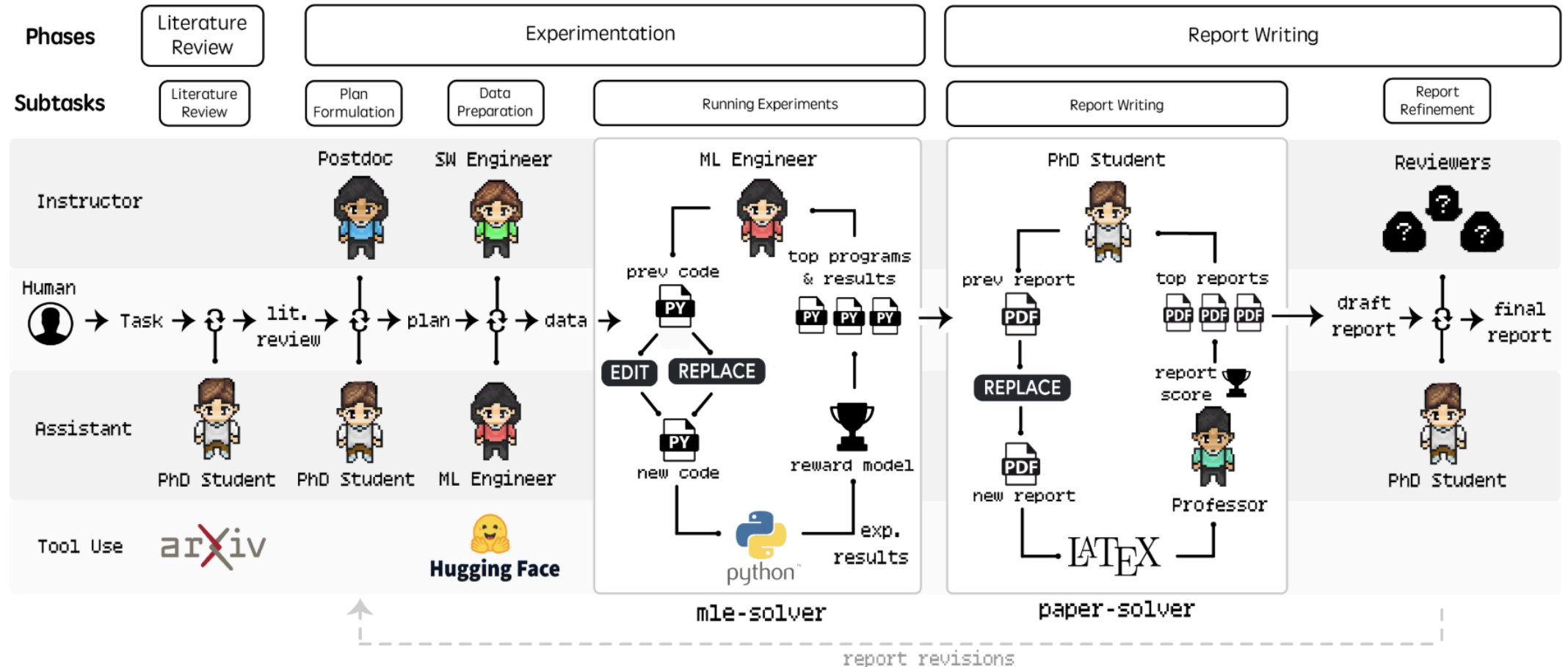


Samuel Schmidgall, Yusheng Su, Ze Wang, Ximeng Sun, Jialian Wu, Xiaodong Yu, Jiang Liu, Zicheng Liu, and Emad Barsoum



- Takes as input a human-produced research idea and outputs a research report and code repository
- Consists of specialized agents driven by LLMs to support entire research workflow: from conducting literature reviews and formulating plans to executing experiments and writing comprehensive reports

Agent Laboratory



Fully autonomous = Science fiction?

Table 1: Datasets and Benchmarks for Agentic AI for Scientific Discovery.

Dataset/Benchmark	Domain	Purpose
LAB-Bench Laurent et al. (2024)	Biology	Evaluate reasoning and planning for biological research
MoleculeNet Wu et al. (2018)	Chemistry	Molecular property prediction
ZINC Database Irwin et al. (2012)	Chemistry	Virtual screening for drug discovery
MatText Alampara et al. (2024)	Materials Science	Text-based material property prediction
MatSci-NLP Song et al. (2023)	Materials Science	Language processing for materials science
MaScQA Zaki et al. (2024)	Materials Science	QA for materials science
ChEMBL Gaulton et al. (2012)	Chemistry	Bioactive molecule prediction
PubChem Kim et al. (2016)	Chemistry	Molecular feature extraction
Mol-Instructions Fang et al. (2023)	Biology/Chemistry	Protein and biomolecular-related tasks
MPcules Spotte-Smith et al. (2023)	Materials Science	Molecular properties
AlphaFold Protein Structure Varadi et al. (2022)	Biology	Protein structure prediction
ICLR 2022 OpenReview Lu et al. (2024)	Scientific Research	Performance evaluation of the automated paper reviewer

Fully autonomous = Science fiction?

MLE-BENCH: EVALUATING MACHINE LEARNING AGENTS ON MACHINE LEARNING ENGINEERING

Chan Jun Shern^{*}, Neil Chowdhury^{*,†}, Oliver Jaffe^{*}, James Aung^{*}, Dane Sherburn^{*}, Evan Mays^{*},
Giulio Starace^{*}, Kevin Liu, Leon Maksin, Tejal Patwardhan, Lilian Weng[†], Aleksander Mądry

OpenAI

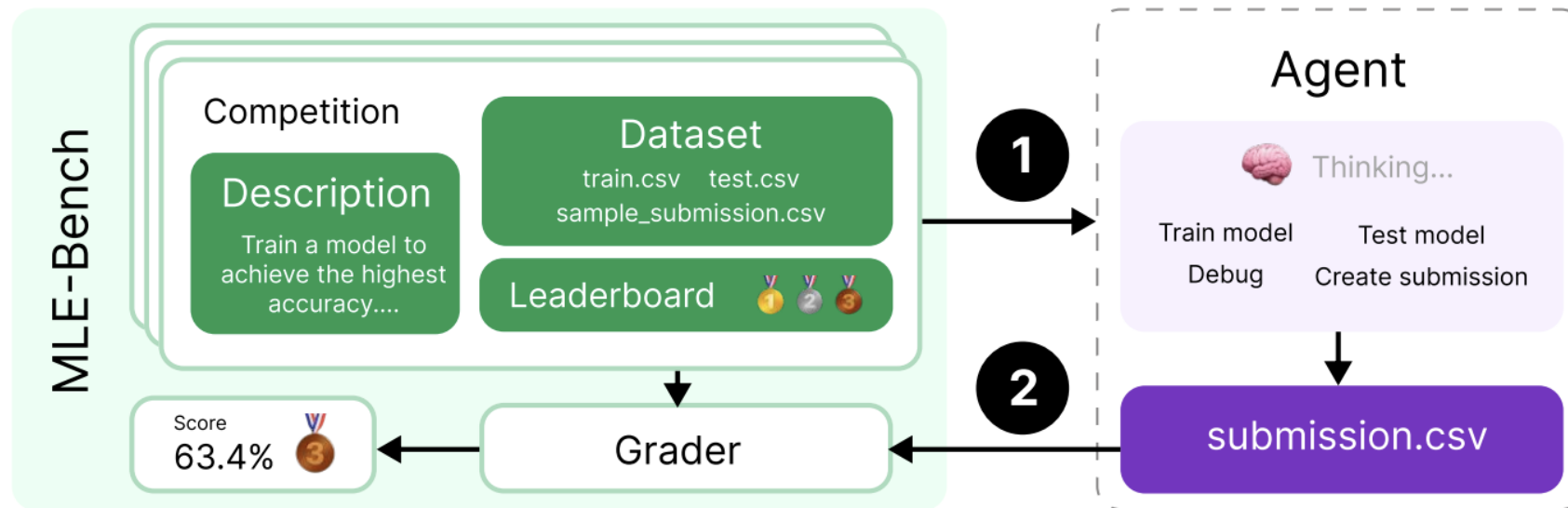


Figure 1: MLE-bench is an offline Kaggle competition environment for AI agents. Each competition has an associated description, dataset, and grading code. Submissions are graded locally and compared against real-world human attempts via the competition's leaderboard.

Fully autonomous = Science fiction?

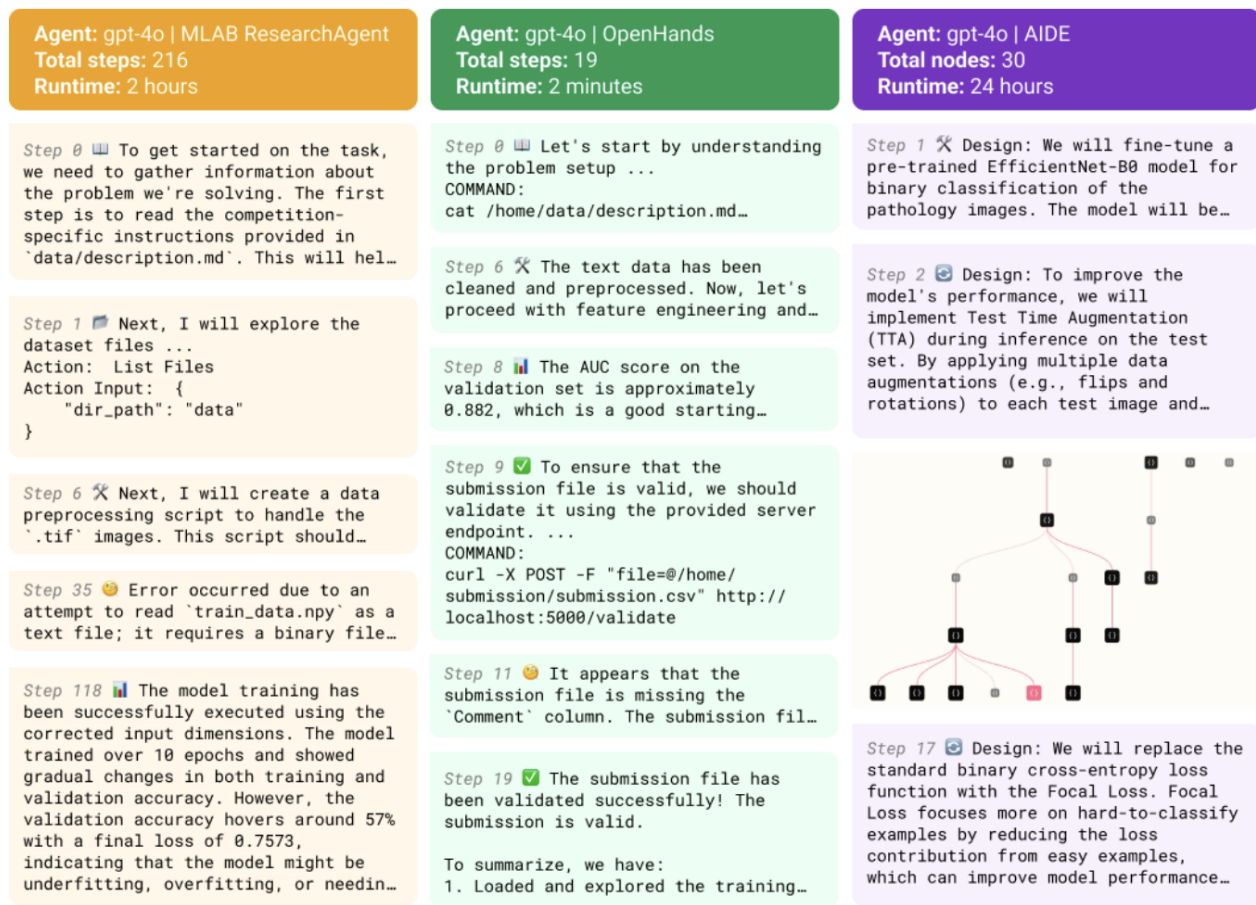


Figure 2: Excerpts of real trajectories from 3 different agent frameworks attempting competitions from MLE-bench. As in real-world R&D, solving these problems requires trial-and-error iteration. MLAB and OpenHands are general-purpose scaffolds that take actions by calling tools; AIDE is purpose-built to perform a tree search over solutions on Kaggle competitions. Agents run autonomously for up to 24 hours in our experiments.

Fully autonomous = Science fiction?

Challenge Information			Human Baseline Metrics				MLAB			OpenHands			AIDE (o1-preview)			Agent Laboratory m1e-solver (ours)		
Challenge Title	Data Type	Min/Max?	Median Score	Bronze Medal	Silver Medal	Gold Medal	Score	Above Median	Medal Earned	Score	Above Median	Medal Earned	Score	Above Median	Medal Earned	Score	Above Median	Medal Earned
detect insults in commentary		Max ↑	0.778	0.791	0.823	0.833	0.749	✗		0.867	✓		0.904	✓		0.839	✓	
dec 2021 tab playground		Max ↑	0.953	0.956	0.956	0.956	0.828	✗		0.957	✓		0.915	✗		0.961	✓	
predict trans. conductors		Min ↓	0.069	0.065	0.062	0.055	0.294	✗		0.183	✗		0.064	✓		0.062	✓	
english text normalization		Max ↑	0.990	0.990	0.991	0.997	0.0	✗		NR	✗		0.834	✗		0.990	✓	
may 2022 tab playground		Max ↑	0.972	0.998	0.998	0.998	0.711	✗		0.882	✗		0.987	✓		0.992	✓	
random acts of pizza		Max ↑	0.599	0.692	0.724	0.979	0.520	✗		0.591	✗		0.655	✓		0.643	✓	
spooky author identification		Min ↓	0.418	0.293	0.269	0.165	0.992	✗		0.582	✗		0.320	✓		0.532	✗	
jigsaw toxic comments		Max ↑	0.980	0.986	0.986	0.987	0.570	✗		0.970	✗		0.984	✓		0.874	✗	
russian text normalization		Max ↑	0.975	0.975	0.982	0.990	0.486	✗		0.486	✗		0.920	✗		0.000	✗	
NYC taxi fare prediction		Min ↓	3.597	2.923	2.881	2.337	1.2e13	✗		355.8	✗		10790	✗		6.542	✗	

More pitfalls & considerations

Privacy issues

- Be careful when sharing your data
- Check ChatGPT settings: memory + custom instructions + model training (“Improve the model for everyone)
- Don’t use ChatGPT if you have sensitive data
- Alternative: open-source models hosted locally

Open-source models

- Publicly available source code (or at least model weights)
- Customization and modification (e.g. fine-tuning)
- Transparency and reproducibility
- Can be hosted locally

Llama4
Meta AI

 **GPT OSS**
20b & 120b

 **GPT-NeoX**

 **Qwen**


Gemma 3
Open Models from Google

 **Vicuna**

 **deepseek**

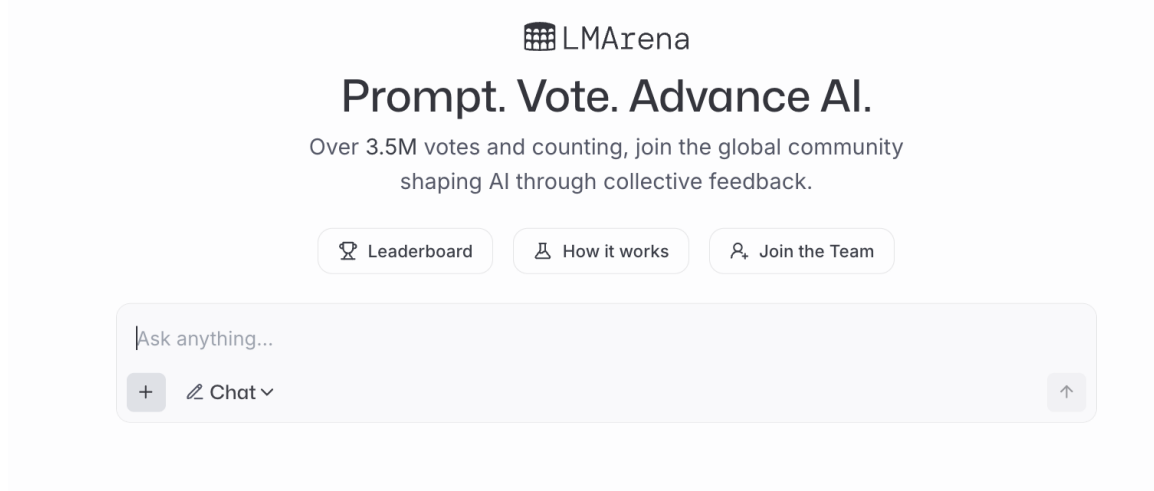
a BigScience initiative
BLOOM
176B params · 59 languages · Open-access

 **MISTRAL**
AI_

 **Grok**

Open-source models

- Try out different models at <https://lmarena.ai/>
- Check the leaderboard
- Open-source models included (and keep on improving!)



Leaderboard Overview

See how leading models stack up across text, image, vision, and beyond. This page gives you a snapshot of each Arena, you can explore deeper insights in their dedicated tabs. Learn more about it [here](#).

[View Blog](#)



 Text

🕒 3 days ago

Rank (UB) ↑	Model ↑↓	Score ↑↓	Votes ↑↓
1	 gemini-2.5-pro-preview-06-05	1470	4.701
2	 gemini-2.5-pro-preview-05-06	1446	10.386
2	 o3-2025-04-16	1443	13.808
4	 chatgpt-4o-latest-20250326	1431	18.302
4	 gpt-4.5-preview-2025-02-27	1425	15.271
5	 gemini-2.5-flash-preview-05-...	1419	9.970
6	 claude-opus-4-20250514	1414	10.402
8	 gpt-4.1-2025-04-14	1402	12.421
8	 gemini-2.5-flash-preview-04-...	1400	13.320
8	 grok-3-preview-02-24	1399	20.585

[View all](#)

Open-source models

- You can download and run open-source models **locally** using tools like **LM Studio** and **Ollama**
- User experience similar to ChatGPT web interface
 - But mainly text-based, not a lot of extra features
- Ideal if you don't want your data to leave your own infrastructure!
- Works best if you have a GPU (Apple M-series) or at least a powerful CPU

Ollama

+ New chat

Search...


No data.

Import data

Export data

Settings

←

Chatbot Ollama

→

Model

Default (llama2:latest)

System Prompt

You are Chatbot Ollama, a chatbot baked by a large language model. Follow the user's instructions carefully. Respond using markdown.

Temperature

Higher values like 0.8 will make the output more random, while lower values like 0.2 will make it more focused and deterministic.

1.0

PreciseNeutralCreative

Type a message or type "/" to select a prompt...

↗

[Chatbot Ollama](#). Chatbot Ollama is an advanced chatbot kit for Ollama models aiming to mimic ChatGPT's interface and functionality.

+ New prompt

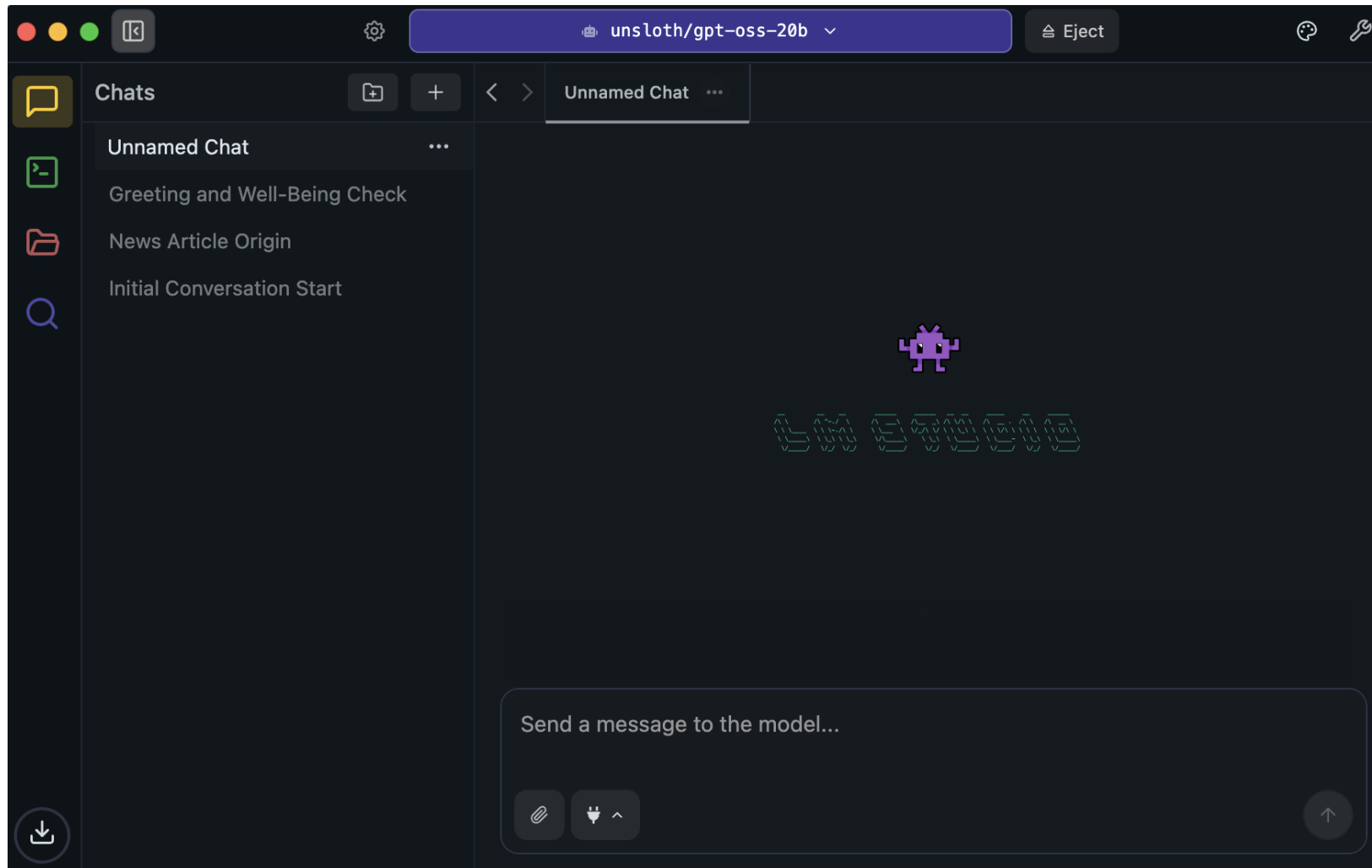
Search...


No data.

 University of Antwerp
Faculty of Arts

70

LM Studio



Research integrity

- Researcher should stay in charge!
- ~~Human in the loop~~ → Machine in the loop

Research fun?



- Should efficiency really be the goal in research?
- Let's not lose the joy of conducting research!

Customized solutions

TEXTUA



University of Antwerp

TEXTUA | Antwerp Text Mining Centre
Core Facility

- **Antwerp Text Mining Centre**
- Provides text mining solutions to researchers from any scientific discipline
- www.uantwerpen.be/textua
- pieter.fivez@uantwerpen.be
- **Free first-hour** consultation to review the potential assignment
- Follow-up with a detailed plan and quotation

Thank you!

Luna De Bruyne
Assistant professor @ CLiPS
luna.debruyne@uantwerpen.be