



中國人民大學  
RENMIN UNIVERSITY OF CHINA

# How to use ChatGPT to facilitate academic research

Yuhao Mu  
(yuhaomuruc@126.com)

Workshop at EwDS research cluster  
School of Finance, Renmin University of China, Beijing, China

- 1. In my presentation, ChatGPT is used only for assisting and supporting academic research, based on the scholars' original research and original writing.**
- 2. In particular, ChatGPT can enhance the efficiency of academic research by reducing repetitive work.**
- 3. Technically speaking, ChatGPT actually cannot generate original innovative research ideas. Ultimately, we still need excellent researchers to push the front of human knowledge.**

**1**

Introduction to ChatGPT

**2**

Applications in academic research

**3**

Other Applications

*01*

**Introduction to ChatGPT**

# 01

## Products

01

In November 2022 - *ChatGPT*, an AI chatbot program developed by *OpenAI*.

02

In February 2023 - the new subscription plan, *ChatGPT Plus*.

03

In March 2023 - *ChatGPT* model became available on Application Programming Interface (API), which gives developers access to speech-to-text and language capabilities.

04

On March 14 2023 - *ChatGPT4.0* was able to process up to 25,000 words, to handle images and more detailed instructions than *ChatGPT3.5*.

05

*ChatGPT4.0* is available through *ChatGPT Plus* and API.



# 01

## Development

Transformer in June 2017

GPT-1 (117M, 2018.6) , GPT-2 (1.5 billion, 2019.2) and GPT-3 (175 billion, 2020.5)

From GPT-3 to ChatGPT

Reinforcement learning from human feedback (**RLHF**) is introduced in *InstructGPT* and it makes the model to generate neutral, objective, safe, and informative answers.

The ability to perform complex reasoning is likely to be from training on code (**CodeX**)

**ChatGPT** utilizes conversations provided by human AI trainers for supervised fine-tuning and the dialog ability emerges.



# 01

## Usage



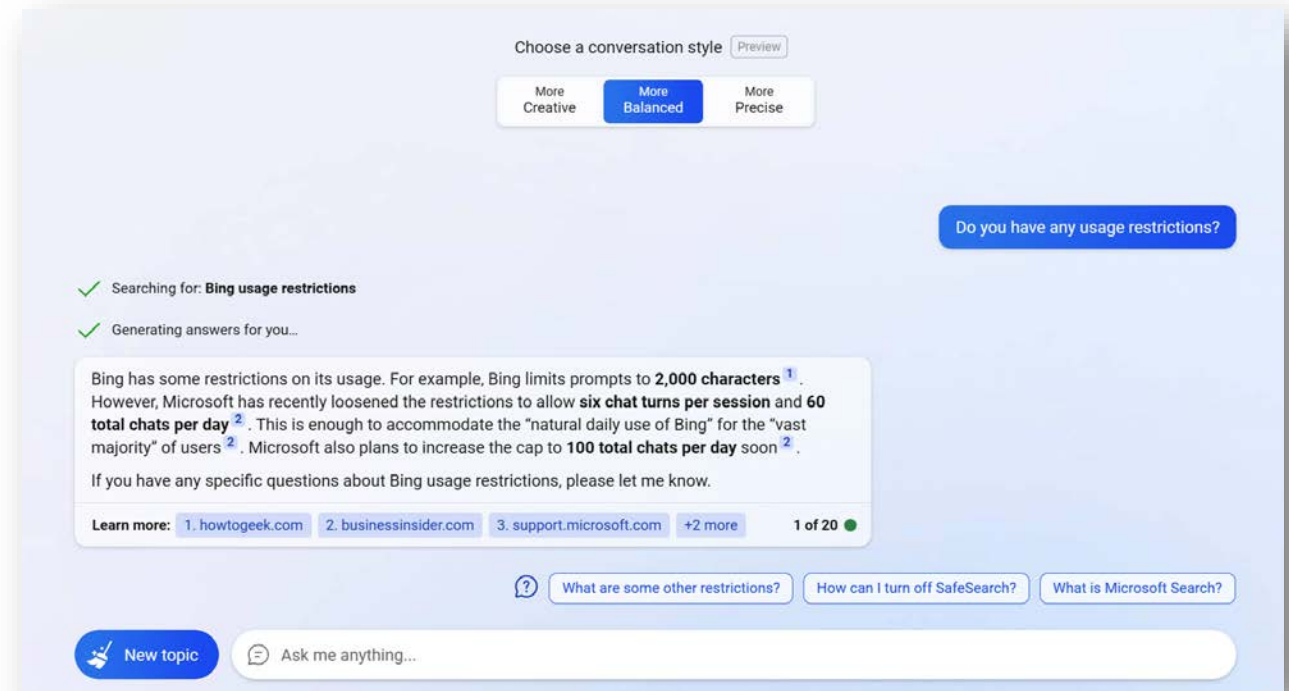
The website of ChatGPT:  
<https://chat.openai.com>



New bing: [www.bing.com/new](http://www.bing.com/new)



Application Programming Interface (API  
the most flexible way):  
[https://platform.openai.com/account/api-  
keys](https://platform.openai.com/account/api-keys)





# 01

## API and tokens: Use ChatGPT as a model

1

Using an **API** involves calling the **ChatGPT model** directly and inputting the desired content to generate explanatory or responsive output.

2

The model understands and processes text by breaking it down into **tokens**, and **OpenAI** charges the usage of **API** by tokens.

3

Tokens can be **words** or just **chunks of characters**. For example, the word “**hamburger**” gets broken up into the tokens “*ham*”, “*bur*” and “*ger*”, while a short and common word like “*pear*” is a single token. **1 token** is approximately **4** characters or **0.75** words for English text.

4

**New users** get **\$5** credit in their accounts, which expires in a few months.

### GPT-4

With broad general knowledge and domain expertise, GPT-4 can follow complex instructions in natural language and solve difficult problems with accuracy.

[Learn more](#)

Model	Prompt	Completion
8K context	\$0.03 / 1K tokens	\$0.06 / 1K tokens
32K context	\$0.06 / 1K tokens	\$0.12 / 1K tokens

### Chat

ChatGPT models are optimized for dialogue. The performance of gpt-3.5-turbo is on par with Instruct Davinci.

[Learn more about ChatGPT](#) ↗

Model	Usage
gpt-3.5-turbo	\$0.002 / 1K tokens



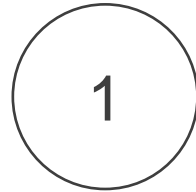
02

**Applications in academic research**



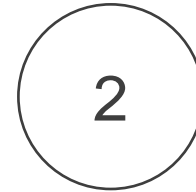
## How **ChatGPT** can help you in academic research?

### Reading



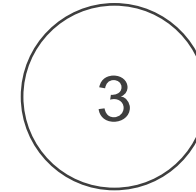
- **Skim Reading:** Summarize and extract main ideas.
- **Detailed Reading:** Understand details and the code.

### Coding



- **Data:** Data processing, such as missing value imputation and descriptive statistics.
- **Model:** Modeling data using a programming language and debug.
- **Plot:** Displaying the results.

### Proofreading



- **Proofreading:** Spelling, grammar, clarity.

# 02

## Skim Reading: ChatPaper

1

Skim reading of papers can't be achieved through the website of ChatGPT.

2

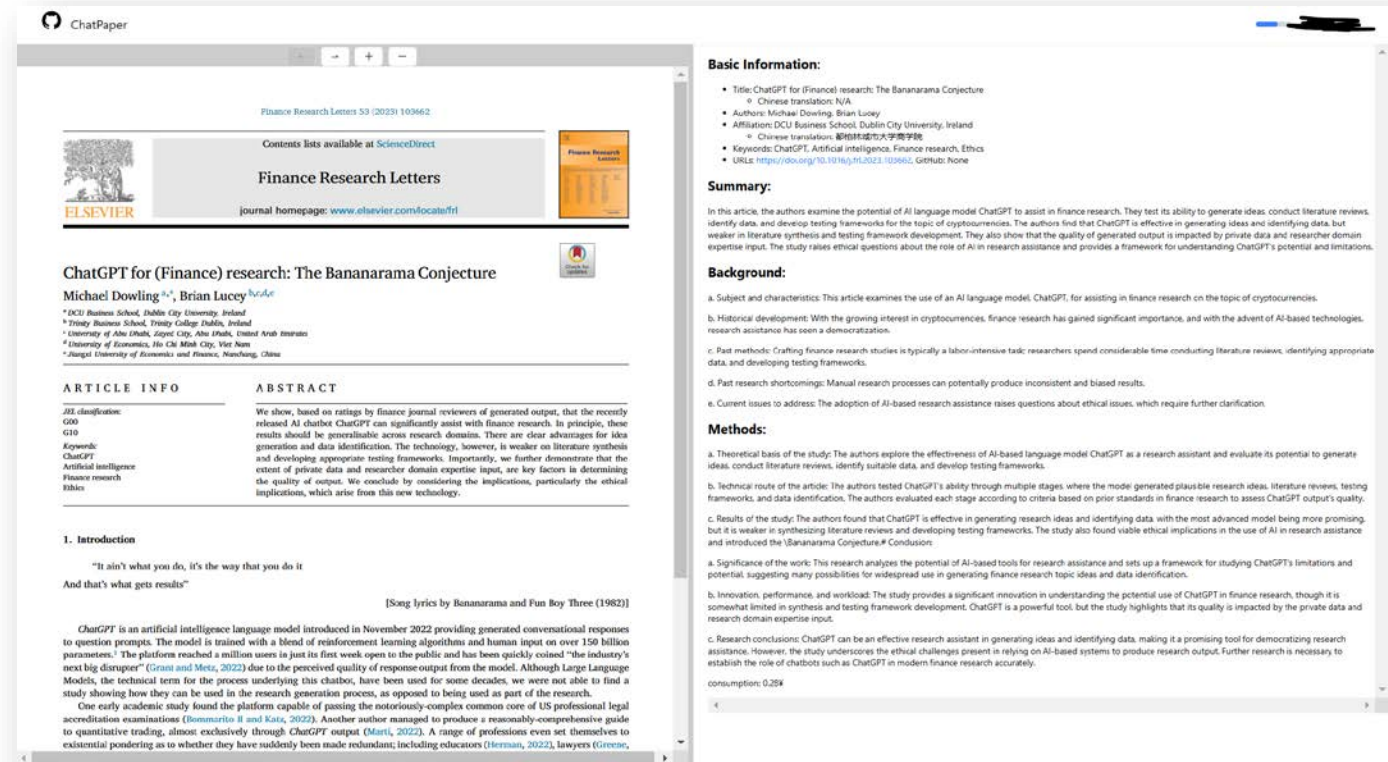
It relies on various websites developed based on API for skim reading of papers.

3

Convert PDF into text, and use API to call models to analyze the text.

4

The website of ChatPaper: <https://chatpaper.org/>

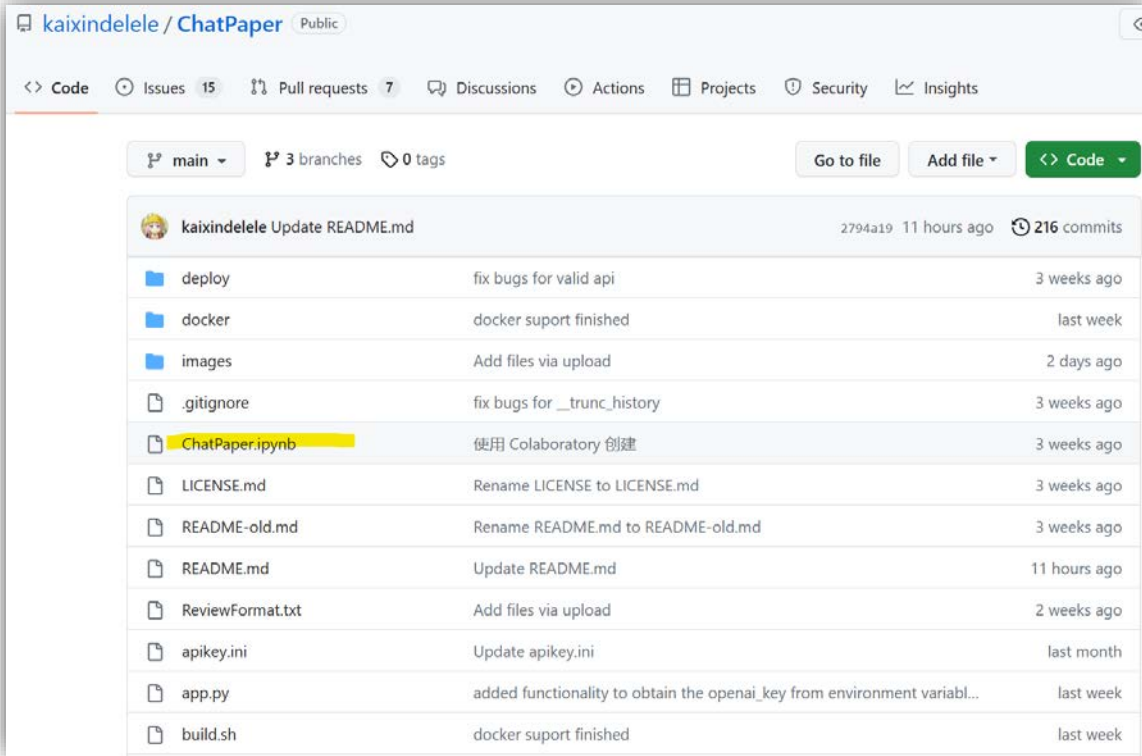




# 02

## ChatPaper

➤ **Github of ChatPaper:**  
<https://github.com/kaixindelele/ChatPaper>



➤ **Download ChatPaper.ipynb, use your own API key and change the format of the summary:**

```
if __name__ == '__main__':
    parser = argparse.ArgumentParser()
    parser.add_argument("--pdf_path", type=str, default="", help="if none, the bot will download from arxiv with query")
    parser.add_argument("--query", type=str, default='all: return', help="the query string, ti: xx, au: xx, all: xx.")
    parser.add_argument("--key_word", type=str, default='Portfolio Management', help="the key word of user research fields")
    parser.add_argument("--filter_keys", type=str, default='', help="the filter key words")
    parser.add_argument("--max_results", type=int, default=2, help="the maximum number of results")
    parser.add_argument("--sort", type=str, default="Relevance", help="another is LastUpdatedDate, and Relevance")
    parser.add_argument("--save_image", default=False, help="save image? It takes a minute or two to save a picture! But pretty")
    parser.add_argument("--file_format", type=str, default='md', help="")
    parser.add_argument("--language", type=str, default='en', help="The other output language is English, is en")
    parser.add_argument("--api_key", type=str, default='sk-v...', help="your openai api key!")
    parser.add_argument('-f')
    args = parser.parse_args()
    import time
    start_time = time.time()
    main(args=args)
    print("summary time:", time.time() - start_time)
```

Key word: Portfolio Management  
Query: all: return  
Sort: SortCriterion.Relevance  
all search:  
0 The Market-Based Probability of Stock Returns 2023-02-06 11:16:18+00:00  
abs\_text: We show how time-series of random market trade values and volumes completely describe stochasticity of stock returns. We derive equation that links up returns with current and past trade values and show how statistical moments of the trade values and volumes determine statistical moments of stock returns. We estimate statistical moments of the trade values and volumes by the conventional frequency-based probability. However we believe that frequencies of stock returns don't define its probabilities as market and financial concepts. We present the market-based treatment of the probability of stock returns that defines average returns during "trading day" that completely match conven

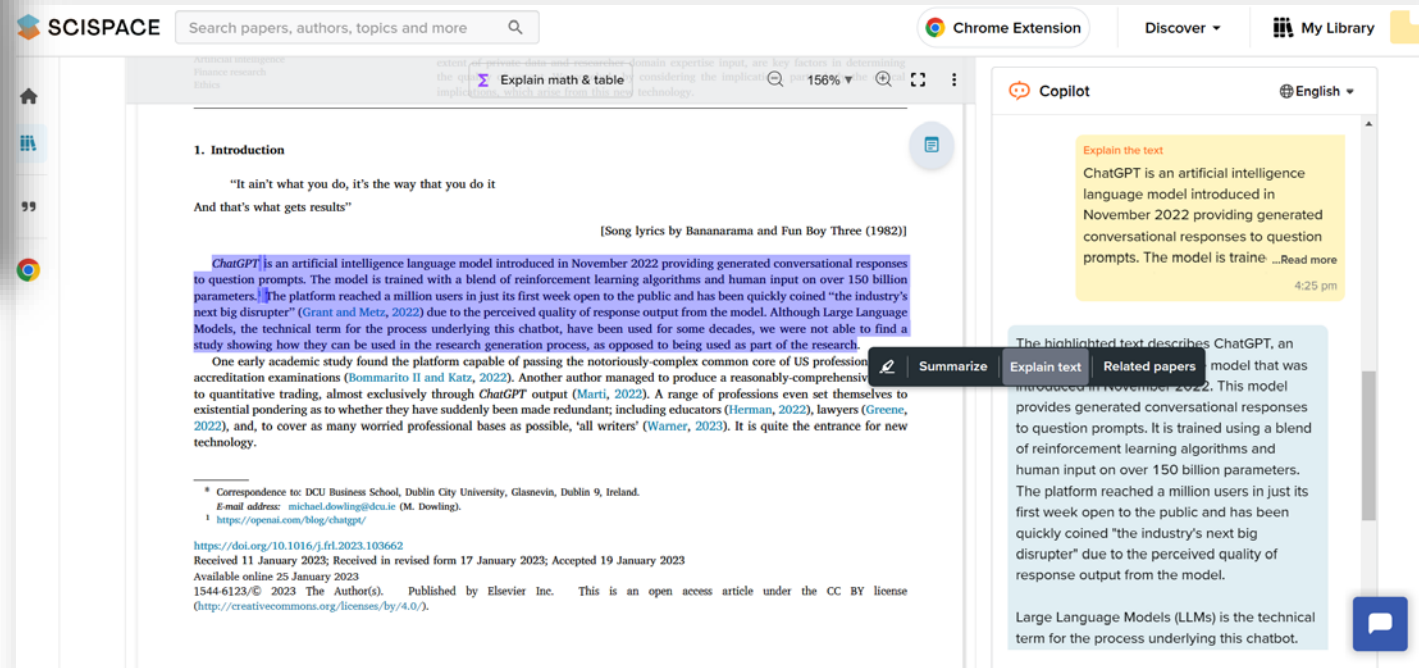
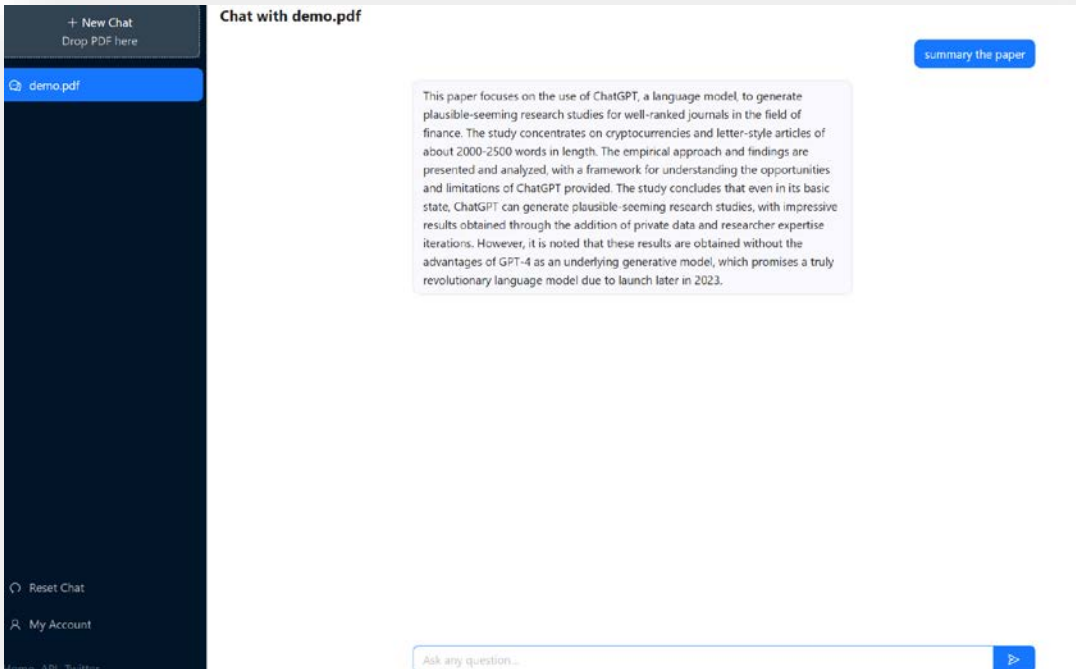


# 02

## Detailed reading

The following websites can aid in the comprehension of academic content:

- ❑ Scispace: <https://typeset.io/library>
- ❑ ChatPDF: <https://www.chatpdf.com>



- ❑ PandaGPT: <https://pandagpt.io>
- ❑ ChatDOC: <https://chatdoc.com>

# 02

## Data

- What kind of data
- Which programming language
- What target to achieve

MU

Please write a piece of code for me, concerning a Python dataframe that contains 10 stocks of 5 days and the columns include the stock code ,tradedate ,10 features and 1 column of returns. There are missing values in two features of the first day of the first stock. Please perform descriptive statistics on this data and fill in the missing values.



Sure! Here's some code that should do what you're asking for:

```
python Copy code

import pandas as pd
import numpy as np

# create sample data
stocks = ['AAPL', 'GOOG', 'MSFT', 'AMZN', 'FB', 'NFLX', 'TSLA', 'NVDA', 'AMD', 'PYPL']
dates = pd.date_range(start='2022-01-01', end='2022-01-05', freq='D')
data = {'Stock Code': [], 'Trade Date': [], 'Feature 1': [], 'Feature 2': [], 'Feature 3': []}

for stock in stocks:
    for date in dates:
        data['Stock Code'].append(stock)
        data['Trade Date'].append(date)
        data['Feature 1'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 2'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 3'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 4'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 5'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 6'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 7'].append(np.random.normal(loc=0.0, scale=1.0))
```

Regenerate response

Send a message...

MU

How to use Lasso for predicting returns based on the above data (Python)



Sure! Here's an example of how to use Lasso regression for predicting returns based on the data you provided using Python's scikit-learn library:

```
python Copy code

import pandas as pd
from sklearn.linear_model import Lasso
from sklearn.metrics import r2_score

# create sample data
stocks = ['AAPL', 'GOOG', 'MSFT', 'AMZN', 'FB', 'NFLX', 'TSLA', 'NVDA', 'AMD', 'PYPL']
dates = pd.date_range(start='2022-01-01', end='2022-01-05', freq='D')
data = {'Stock Code': [], 'Trade Date': [], 'Feature 1': [], 'Feature 2': [], 'Feature 3': []}

for stock in stocks:
    for date in dates:
        data['Stock Code'].append(stock)
        data['Trade Date'].append(date)
        data['Feature 1'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 2'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 3'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 4'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 5'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 6'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 7'].append(np.random.normal(loc=0.0, scale=1.0))
```

Stop generating

Send a message...



# 02

## Model and Plot

MU How to use Random forest for predicting returns based on the above data (Python)

Sure! Here's an example of how to use Random Forest regression for predicting returns based on the data you provided using Python's scikit-learn library:

```
python Copy code

import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score

# create sample data
stocks = ['AAPL', 'GOOG', 'MSFT', 'AMZN', 'FB', 'NFLX', 'TSLA', 'NVDA', 'AMD', 'PYPL']
dates = pd.date_range(start='2022-01-01', end='2022-01-05', freq='D')
data = {'Stock Code': [], 'Trade Date': [], 'Feature 1': [], 'Feature 2': [], 'Feature 3': []}

for stock in stocks:
    for date in dates:
        data['Stock Code'].append(stock)
        data['Trade Date'].append(date)
        data['Feature 1'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 2'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 3'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 4'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 5'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 6'].append(np.random.normal(loc=0.0, scale=1.0))
        data['Feature 7'].append(np.random.normal(loc=0.0, scale=1.0))
```

Send a message...

MU How to visualize variable importance using random forests in a graph.

To visualize variable importance using Random Forest in a graph, you can use the `feature_importances_` attribute of the `RandomForestRegressor` or `RandomForestClassifier` class in scikit-learn library. Here's an example code snippet that demonstrates how to do this:

```
python Copy code

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor

# create sample data
np.random.seed(42)
stocks = ['AAPL', 'GOOG', 'MSFT', 'AMZN', 'FB', 'NFLX', 'TSLA', 'NVDA', 'AMD', 'PYPL']
dates = pd.date_range(start='2022-01-01', end='2022-01-05', freq='D')
data = {'Stock Code': [], 'Trade Date': [], 'Feature 1': [], 'Feature 2': [], 'Feature 3': []}

for stock in stocks:
    for date in dates:
        data['Stock Code'].append(stock)
        data['Trade Date'].append(date)
```

Send a message...



Polish

“Please proofread and polish the passage from an academic angle and highlight the modification:”

Polish

“Below is a paragraph from an academic paper. Polish the writing to meet the academic style, and improve the spelling, grammar, clarity, concision, and overall readability. When necessary, rewrite the whole sentence. Furthermore, list all modifications and explain the reasons to do so.”

Grammar

“Can you help me ensure that the grammar and the spelling are correct? Do not try to polish the text, if no mistake is found, tell me that this paragraph is good. If you find grammar or spelling mistakes, please list mistakes you find in a two-column markdown table, put the original text in the first column, put the corrected text in the second column, and highlight the keywords you fixed.”





# 02

## Proofreading

MU

Below is a paragraph from an academic paper. Polish the writing to meet the academic style, improve the spelling, grammar, clarity, concision and overall readability. When necessary, rewrite the whole sentence. Furthermore, list all modification and explain the reasons to do so. "When conducting research projects, it is crucial to clearly describe the problem to be addressed. A good problem description should include the characteristics of the data to be analyzed, the programming language to be used, and the goals to be achieved. This paper discusses the importance of a well-defined problem description and provides guidance on how to construct an effective problem statement for research projects. By following these guidelines, researchers can increase the chances of success in their projects and better communicate their research objectives to their peers and stakeholders."



When embarking on research projects, it is crucial to provide a clear and concise problem statement that outlines the data characteristics, programming language, and research objectives. This paper highlights the significance of a well-defined problem statement and offers guidance on how to construct an effective one. By following these guidelines, researchers can improve the likelihood of project success and enhance communication with peers and stakeholders.



Modifications:

- Replaced "conducting" with "embarking on" to improve clarity and concision.
- Rewrote the sentence to emphasize the importance of a clear and concise problem statement.
- Changed "good" to "effective" to convey a more academic tone.

Regenerate response

Send a message...



MU

Can you help me ensure that the grammar and the spelling is correct? Do not try to polish the text, if no mistake is found, tell me that this paragraph is good. If you find grammar or spelling mistakes, please list mistakes you find in a two-column markdown table, put the original text the first column, put the corrected text in the second column and highlight the key words you fixed. "We show, based on ratings by finance journal reviewers of generated output, that the recently released AI chatbot ChatGPT can significantly assist with finance research. In principle, these results should be generalisable across research domains. There are clear advantages for idea generation and data identification. The technology, however, is weaker on literature synthesis and developing appropriate testing frameworks. Importantly, we further demonstrate that the extent of private data and researcher domain expertise input, are key factors in determining the quality of output. We conclude by considering the implications, particularly the ethical implications, which arise from this new technology."



Original Text	Corrected Text
We show, based on ratings by finance journal reviewers of generated output, that the recently released AI chatbot ChatGPT can significantly assist with finance research.	We show, based on ratings by finance journal reviewers of the generated output, that the recently released AI chatbot ChatGPT can significantly assist with finance research.
In principle, these results should be generalisable across research domains.	In principle, these results should be generalizable across research domains.
There are clear advantages for idea generation and data identification.	There are clear advantages for generating ideas and identifying data.
The technology, however, is weaker on literature synthesis and developing	technology is weaker in literature synthesis and developing



Regenerate response

Send a message...





02

A comprehensive platform:  
Chatgpt\_academic

A research assistant platform based on API development.

English version: [https://github.com/binary-husky/chatgpt\\_academic/files/11145631/generated\\_english\\_version.zip](https://github.com/binary-husky/chatgpt_academic/files/11145631/generated_english_version.zip)

### ChatGPT Academic Optimization

Chatbot

Based on your analysis above, please summarize the overall function and structure of the program. Then use a markdown table to organize the functions of each file (including main.py).

Based on the analysis above, the `main.py` file contains the main code for running a Gradio interface for using the ChatGPT model. It sets up a user interface that allows users to make predictions using the ChatGPT model, and provides various helper functions to support the user's interactions. Here's a table summarizing the functions of each file:

File	Description
<code>check_proxy.py</code>	Contains a function to check whether a proxy is working
<code>config_private.py</code>	Contains configuration values like proxies, web port, LLM model, concurrent count, and authentication
<code>functional.py</code>	Contains a dictionary of helper functions for the Gradio UI
<code>functional_crazy.py</code>	Contains another dictionary of helper functions for the Gradio UI
<code>gradio.py</code>	Contains various UI elements like text boxes, sliders, buttons, and file upload fields
<code>logging.py</code>	Contains a logging function to save user inputs and predictions to a log file
<code>main.py</code>	The main program file for running the Gradio interface and making predictions using the ChatGPT model
<code>predict.py</code>	Contains a function for making predictions using the ChatGPT model
<code>theme.py</code>	Contains a function for creating a custom UI theme
<code>toolbox.py</code>	Contains various helper functions like <code>get_conf</code> to get configuration values from a <code>config_private.py</code> file

Are you done?

The above materials have been written in D:\Chatgpt\generated\_english\_version\gpt\_log\chatGPT\_analysis\_report2023-04-06-19-52-31.md

How to get the summary report remotely?

The summary report has been added to the "File Upload Area" on the right side (may be folded), please check.

private\_upload/2023-04-06-19-50-58

Submit

Reset      Stop

normal

Basic Functions Area

English Academic Proofreading

Chinese Academic Proofreading

Find Grammar Errors

Chinese to English Translation

Academic Translation between Chinese and English

English to Chinese Translation

Find Pictures

ExplainCode

Function Plugin Area

Note: The function plugins marked in 'Red' below need to read the path from the input area as a parameter.

Parse the entire Python project

Parse the entire C++ project header file

Read Tex article and write summary

Batch generate function comment

[multi-thread demo] switch this project to english      [function plugin demo] today in history

More Function Plugins

Click to Expand "File Upload Area". Upload local files for the Red Function Plugin to call.

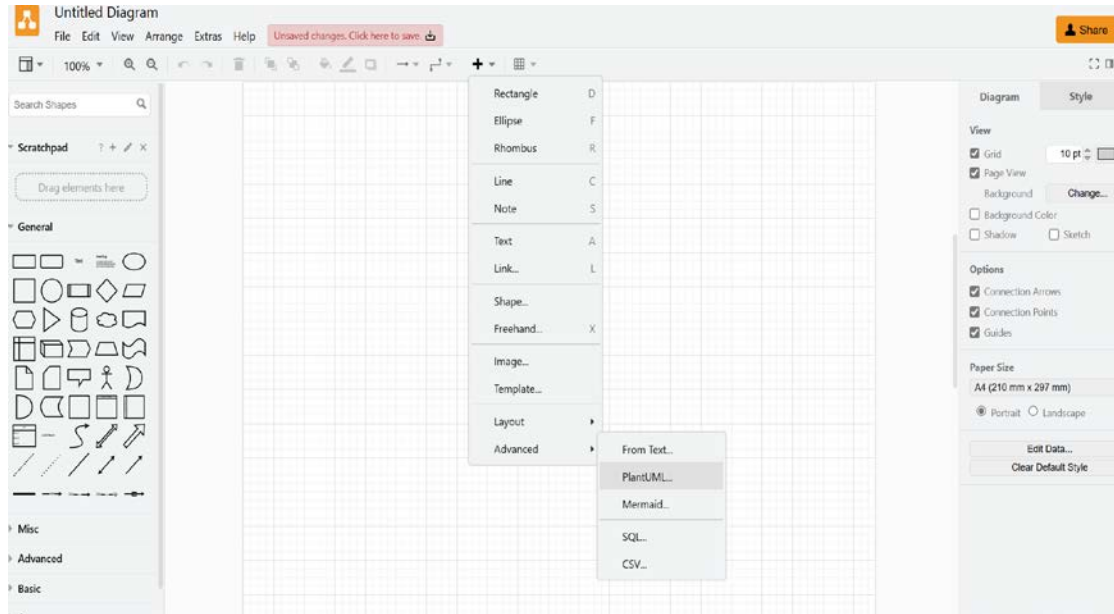
Any file, but compressed file (zip, tar) recommended

chatGPT\_analysis\_report2023-04-...      3.0 KB      [Download](#)

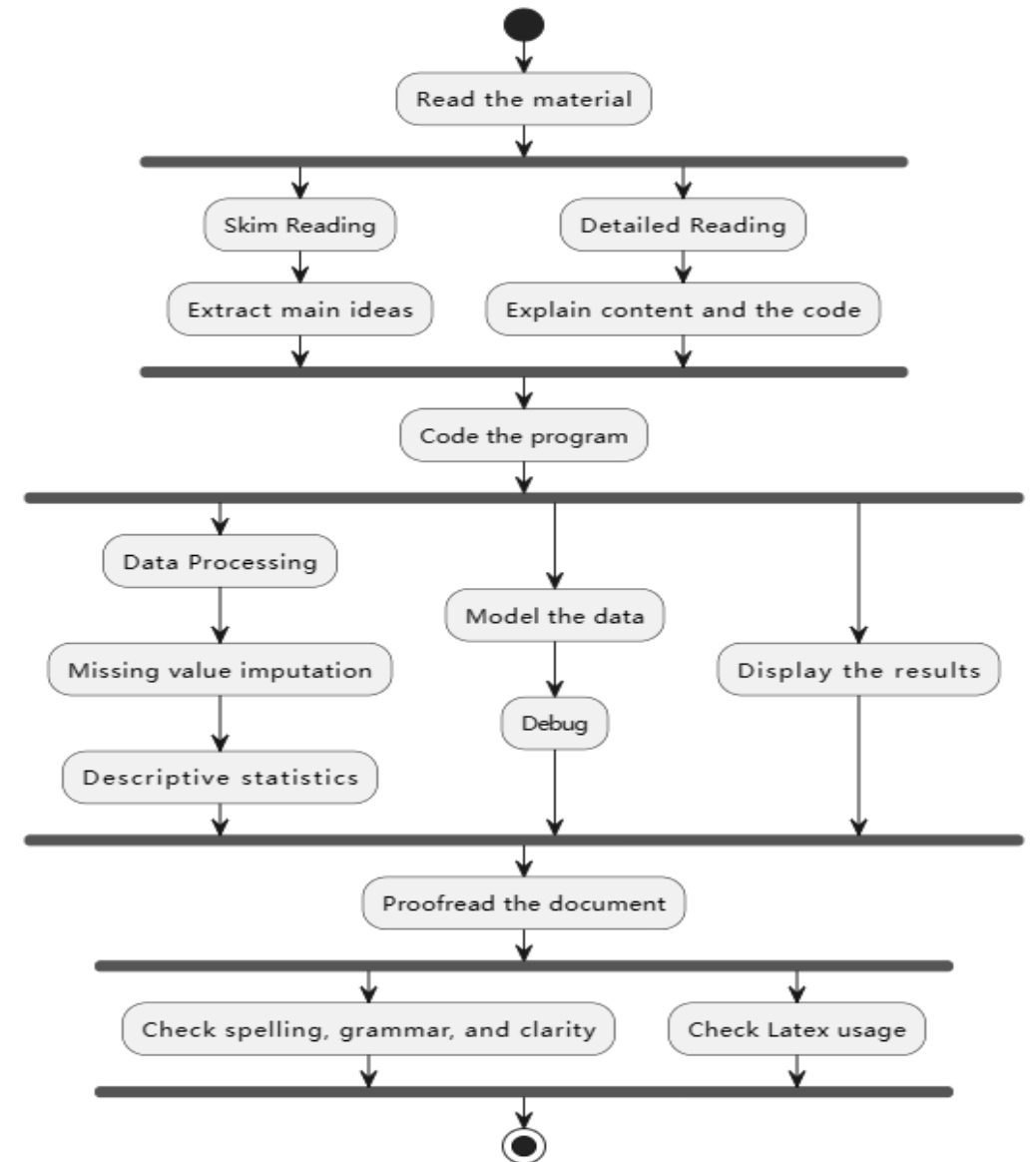
# 02

## Research Process Diagram

- “As a **PlantUML** activity diagram generator, you need to create a flowchart based on my prompts, and the flowchart should have branches and parallelism, and output the corresponding **PlantUML** code.”



- Use the **PlantUML** code for diagrams:  
<https://app.diagrams.net/>



03

**Other Applications**



## ChatReviewer:

<https://huggingface.co/spaces/ShiwenNi/ChatReviewer>

请输入你的API-key(sk开头的字符串)

.....

请输入特定的评审要求和格式(否则为默认格式)

respond to the concerns individually).

These may include, but are not limited to: inadequate implementation details for reproducing the study, limited evaluation and ablation studies for the proposed method, correctness of the theoretical analysis or experimental results, lack of comparisons or discussions with widely-known baselines in the field, lack of clarity in exposition, or any other factors that may impede the reader's understanding or benefit from the paper. Please kindly refrain from providing a general assessment of the paper's novelty without providing detailed explanations. (Maximum length: 2,000 characters)

(1) xxx

(2) xxx

(3) xxx

\* Questions To Authors And Suggestions For Rebuttal

Please provide a numbered list of specific and clear questions that pertain to the details of the proposed method, evaluation setting, or additional results that would aid in supporting the authors' claims.

The questions should be formulated in a manner that, after the authors have answered them during the rebuttal, it would enable a more thorough assessment of the paper's quality.

(Maximum length: 2,000 characters)

\*Overall score (1-10)

The paper is scored on a scale of 1-10, with 10 being the full mark, and 6 stands for borderline accept. Then give the reason for your rating.

xxx

评审结果

- (1) The paper does not compare ChatGPT-generated output with output generated by human experts or other AI tools, which could have provided a more comprehensive understanding of ChatGPT's potential and limitations.
- (2) The sample size of reviewers is relatively small (32), which could limit the generalizability of the findings.
- (3) The paper doesn't provide any in-depth analysis of the ethical implications of using ChatGPT for research study writing.

Questions to Authors and Suggestions for Rebuttal:

- (1) Have you considered comparing ChatGPT-generated output with output generated by other AI tools or human experts to evaluate ChatGPT's relative performance?
- (2) How do you plan to address the limitation of the small review sample size in future research?
- (3) Have you considered the potential ethical implications of using ChatGPT in research writing, such as issues related to data privacy, intellectual property, and biases?

Overall score: 7

The paper provides a valuable contribution to the limited literature on the potential of ChatGPT for writing a research study. The methodology is well-structured and the findings are insightful, particularly in showing how incorporating researcher domain expertise alongside private data can significantly improve the output quality. However, the paper could be strengthened by addressing the weaknesses outlined above.

资源统计

使用token数: 4019












花费时间: 20.23秒



# 03

## Plugins

❑ **Plugins** are tools designed specifically for language models with safety as a core principle, and help **ChatGPT** access up-to-date information, run computations, or use third-party services. Only **Plugin** developers who have been invited off the waitlist can use.

 <b>Expedia</b> Bring your trip plans to life—get there, stay there, find things to see and do.	 <b>FiscalNote</b> Provides and enables access to select market-leading, real-time data sets for legal, political, and regulatory data and information.	 <b>Instacart</b> Order from your favorite local grocery stores.	 <b>KAYAK</b> Search for flights, stays and rental cars. Get recommendations for all the places you can go within your budget.
 <b>Klarna Shopping</b> Search and compare prices from thousands of online shops.	 <b>Milo Family AI</b> Giving parents superpowers to turn the manic to magic, 20 minutes each day. Ask: Hey Milo, what's magic today?	 <b>OpenTable</b> Provides restaurant recommendations, with a direct link to book.	 <b>Shop</b> Search for millions of products from the world's greatest brands.
 <b>Speak</b> Learn how to say anything in another language with Speak, your AI-powered language tutor.	 <b>Wolfram</b> Access computation, math, curated knowledge & real-time data through Wolfram Alpha and Wolfram Language.	 <b>Zapier</b> Interact with over 5,000+ apps like Google Sheets, Trello, Gmail, HubSpot, Salesforce, and more.	



# 03

## Further prompts

Refer to:

<https://github.com/f/awesome-ChatGPT-prompts>

<https://github.com/torantulino/auto-gpt>



中國人民大學

RENMIN UNIVERSITY OF CHINA

Thank you !

RENMIN UNIVERSITY OF CHINA