THE UNIVERSITY OF CHICAGO

How can we use ChatGPT better: A research of API-Enhanced ChatGPT in stock prediction

By

Qichang Zheng

June 2024

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Joseph Hardwick

Preceptor: Joseph Hardwick

Abstract

In this study, we propose an API-enhanced ChatGPT structure that incorporates stock price and news data to improve stock price movement predictions. By integrating external data sources and prompt engineering techniques, our approach demonstrates a significant improvement in predictive performance compared to using only stock price data. The inclusion of news data alongside stock prices results in an approximately 10% increase in accuracy and F1 scores, as well as a 20% improvement in risk-adjusted returns, as measured by Sharpe ratios and information ratios. Our findings highlight the potential of leveraging conversational AI and large language models for stock market analysis, while also identifying areas for further research and optimization, such as addressing stock-specific challenges and developing cost-effective strategies for implementation. This study contributes to the limited body of literature on the application of large language models in finance and paves the way for future research in enhancing the capabilities of AI-driven investment decision-making tools.

Keywords: Large Language Models; LLMs; ChatGPT; Stock Prediction; FastGPT

1 Introduction

The rapid advancement of conversational AI technologies, such as ChatGPT, has revolutionized the way we access and analyze information across various domains, including finance. These large language models (LLMs) offer unprecedented opportunities for retrieving and processing financial data, potentially transforming the landscape of investment decision-making. However, the effectiveness of using ChatGPT directly for financial queries versus an enhanced approach that combines ChatGPT with external data sources through API-based information collection remains largely unexplored.

This study aims to bridge this gap by evaluating the performance of an API-enhanced ChatGPT structure in the context of U.S. stock market information retrieval and stock price movement prediction. By incorporating stock price and news data alongside prompt engineering techniques, we seek to improve the accuracy, risk-adjusted returns, and overall predictive capabilities of ChatGPT in the financial domain. Our research focuses on metrics that are critical for users such as investors, analysts, and enthusiasts, including accuracy rates, F1 scores, Sharpe ratios, and information ratios.

The literature review synthesizes findings from pivotal papers spanning the integration of big data and artificial intelligence (AI) in accelerating business growth, the effectiveness of AI in stock market prediction, and the transformative potential of AI in the financial industry. Recent studies have highlighted the importance of text analysis in stock price prediction, with LLMs demonstrating promising results in linking textual information with traditional quantitative methods (Schumaker & Chen, 2009; Yoo et al., 2023; Duan et al., 2023; Meng et al., 2023; Jiang et al., 2023). These studies collectively provide a foundational understanding necessary for exploring the efficiency of ChatGPT in stock-related topics, leveraging both direct queries and API-enhanced data collection methods.

Our research builds upon this existing body of literature by specifically investigating the impact of incorporating external data sources and prompt engineering techniques on the performance of ChatGPT in stock market analysis. We hypothesize that the API-enhanced ChatGPT structure will demonstrate improved accuracy, risk-adjusted returns, and predictive capabilities compared to using ChatGPT directly with only stock price data.

Through a comprehensive evaluation of our proposed approach using real-world stock price and news data, we aim to contribute to the limited literature on the application of LLMs in finance and provide valuable insights for researchers, practitioners, and investors. Our findings have the potential to inform the development of more effective and efficient AI-driven tools for stock market analysis and investment decision-making, ultimately revolutionizing the way financial information is accessed and processed.

2 Literature Review

Conversational AI in Financial Services

The financial sector has increasingly adopted conversational AI technologies, including chatbots and conversational robo-advisors, to enhance customer interaction and service delivery. The paper ”Artificial Intelligence for Conversational Robo-Advisor” (Day et al., 2018) demonstrates the application of deep learning in automated financial management, highlighting the potential of AI in optimizing investment portfolios through conversational interfaces. Similarly, ”AI-based Chatbot Service for Financial Industry” discusses the deployment of chatbots for customer support and sales in the financial industry, emphasizing the role of AI in automating communication and enhancing customer service efficiency.

AI and Big Data in Stock Market Analysis

The predictive power of AI in stock market analysis is a subject of growing interest. ”Effectiveness of Artificial Intelligence in Stock Market Prediction Based on Machine Learning” (Mokhtari et al., 2021) explores the use of machine learning algorithms in predicting stock market trends, indicating the median performance of AI technologies in this domain. This finding suggests that while AI holds promise for stock market analysis, its capabilities are yet to be fully realized.

Integrating AI with Big Data for Business Growth

The intersection of AI and big data is revolutionizing business analytics and operational efficiency. "Accelerating Business Growth with Big Data and Artificial Intelligence" (Ghimire et al., 2020) underscores the transformative impact of AI and big data across industries, including finance, where these technologies are used to solve complex problems more efficiently than traditional computational systems. "A Comprehensive Study on Integration of Big Data and AI in Financial Industry and its Effect on Present and Future Opportunities" forecasts the expansion of AI in finance, anticipating a significant transformation in customer transactions and operational efficiency.

Challenges and Future Directions of Conversational Agents

Despite the potential benefits, the deployment of conversational AI faces challenges. "An Overview of Conversational Agent: Applications, Challenges, and Future Directions" (Alnefaie et al., 2021) provides a comprehensive analysis of the evolution and application of conversational agents, highlighting the challenges and recommending areas for future research. This overview is crucial for understanding the limitations and potential improvements in employing conversational AI for financial analysis.

Synthesis and Gap Identification

These papers collectively highlight the rapid advancement of AI and big data technologies in transforming the financial industry, improving operational efficiency, and enhancing customer interaction through conversational agents. However, there is a noted gap in specific research addressing the efficiency of using ChatGPT, particularly in the context of stock market analysis. This gap presents an opportunity to explore how ChatGPT, both directly and with API-enhanced methods, can optimize the retrieval and analysis of stock-related information in terms of speed, accuracy, and cost. The current study aims to fill this gap by comparing these two methods, contributing to the broader discourse on the application of conversational AI in financial services.

Importance of stock prediction and LLMs

Recent studies have highlighted the importance of text analysis in stock price prediction. Schumaker and Chen, 2009 demonstrated that incorporating textual information from 8-K financial reports significantly improves stock movement prediction compared to using financial information alone. Swamy et al., 2023 explored the potential of large language models (LLMs) in predicting market moves by linking LLM responses with features in traditional quantitative methods. Tong et al., 2024 proposed Ploutos, a novel financial LLM framework that fuses textual and numerical information for interpretable stock movement prediction. A comprehensive overview by Zhao et al., 2024 showcased the emerging integration of LLMs into various financial tasks, including market trend forecasting and investor sentiment analysis. Lastly, Wu, 2024 investigated the effectiveness of LLMs in predicting stock returns based on news scores and conducted related economical analysis in the Chinese A-share market.

3 Data

The research will utilize real-time and historical data related to U.S. stocks, specifically targeting share prices and news articles by conducting experiments. The sources will include official stock market feeds, financial news websites, and databases that offer APIs for data collection. These sources will be used to validate the accuracy of the information provided by ChatGPT and to feed data into ChatGPT in the API-enhanced method.

3.1 Adjusted Close Price

Here the adjusted stock close price is used to calculate the stock return. When analyzing stock returns in academic papers, the use of adjusted close prices becomes critical for a comprehensive and accurate evaluation. This approach considers several factors that affect stock prices post-closing, providing a more precise reflection of a stock's value and its returns over time. Below, I introduce and explain the significance of using adjusted close prices for calculating stock returns in academic contexts.

3.1.1 Introduction to Adjusted Close Price

The adjusted close price is a stock's closing price that has been modified to include any actions that occurred at any time after the market close. This adjustment includes dividends, stock splits, and rights offerings. The primary purpose of adjusting the close price is to ensure that the stock's performance reflects its true value, free from distortions caused by corporate actions.

3.1.2 Dividend Adjustments

When a company pays dividends, it distributes a portion of its earnings to shareholders. This distribution typically results in a decrease in the stock's price, equivalent to the dividend amount, on the ex-dividend date. By adjusting for dividends, the adjusted close price offers a clearer picture of the stock's performance by incorporating the total shareholder return, which includes both price appreciation and dividends received.

3.1.3 Stock Splits and Reverse Splits

Stock splits and reverse splits change the number of shares outstanding without altering the company's market capitalization. A stock split reduces the price per share and increases the number of shares, while a reverse split does the opposite. Adjusting the close price for these actions ensures that the analysis reflects the stock's value and performance without the distortion caused by changes in share quantity.

3.1.4 Rights Offerings

Rights offerings, where existing shareholders are given the right to purchase additional shares at a discount, can also impact the stock's price. Adjusting for such corporate actions ensures that the stock's performance analysis is not skewed by temporary price changes resulting from these events.

3.2 Summary

Using adjusted close prices in calculating stock returns is paramount in academic research for achieving accurate and meaningful insights. This methodology ensures that stock performance metrics truly reflect the underlying value changes in a company's stock, accounting for dividends, stock splits, and other corporate actions. By providing a more precise measure of stock returns, researchers and analysts can better understand market dynamics, make more informed decisions, and contribute to the robust body of financial literature.

3.3 Data Collection

3.3.1 Stock Price Data

In the data collection process, we will utilize the yfinance API to gather stock price data for the selected companies. Yfinance is a popular open-source Python library that provides a convenient way to access historical and real-time financial data from Yahoo Finance, a widely used platform for stock market information.

Key features of the yfinance API include:

Comprehensive Data Coverage

Yfinance offers access to a wide range of financial data for stocks, indices, currencies, and commodities from various markets worldwide. It covers a vast majority of publicly traded companies, making it an ideal choice for our study, which focuses on U.S. stocks. Historical Data Retrieval: The library enables users to retrieve historical stock price data, including open, high, low, close prices, and trading volume, for specified time periods. This feature is crucial for our research, as we will analyze the performance of the API-enhanced ChatGPT structure in predicting stock price movements based on historical data.

Real-time Data Access
Yfinance provides access to real-time stock price data, allowing users to fetch the most recent trading information. Although our study primarily focuses on historical data, the ability to incorporate real-time data can be valuable for future extensions or applications of our proposed structure.

Dividend and Stock Split Adjustments
The API automatically adjusts historical price data for dividends and stock splits, ensuring that the retrieved data accurately reflects the stock's performance over time. This feature saves considerable data preprocessing efforts and ensures the integrity of our analysis.

Easy Integration
Yfinance seamlessly integrates with Python, the programming language used in our study. It provides a user-friendly interface for querying and retrieving financial data, making it straightforward to incorporate into our data collection pipeline.

3.3.2 Stock News Data

In addition to stock price data, our study also incorporates news data to enhance the information available to the API-enhanced ChatGPT structure. For this purpose, we utilize the Finnhub API, which provides access to a wide range of financial news articles from various sources worldwide.

Finnhub is a leading provider of financial data and offers a comprehensive API for retrieving real-time and historical market data, including stock prices, company fundamentals, and news. The Finnhub API covers a vast array of companies and financial instruments, making it an ideal choice for our research on U.S. stocks.

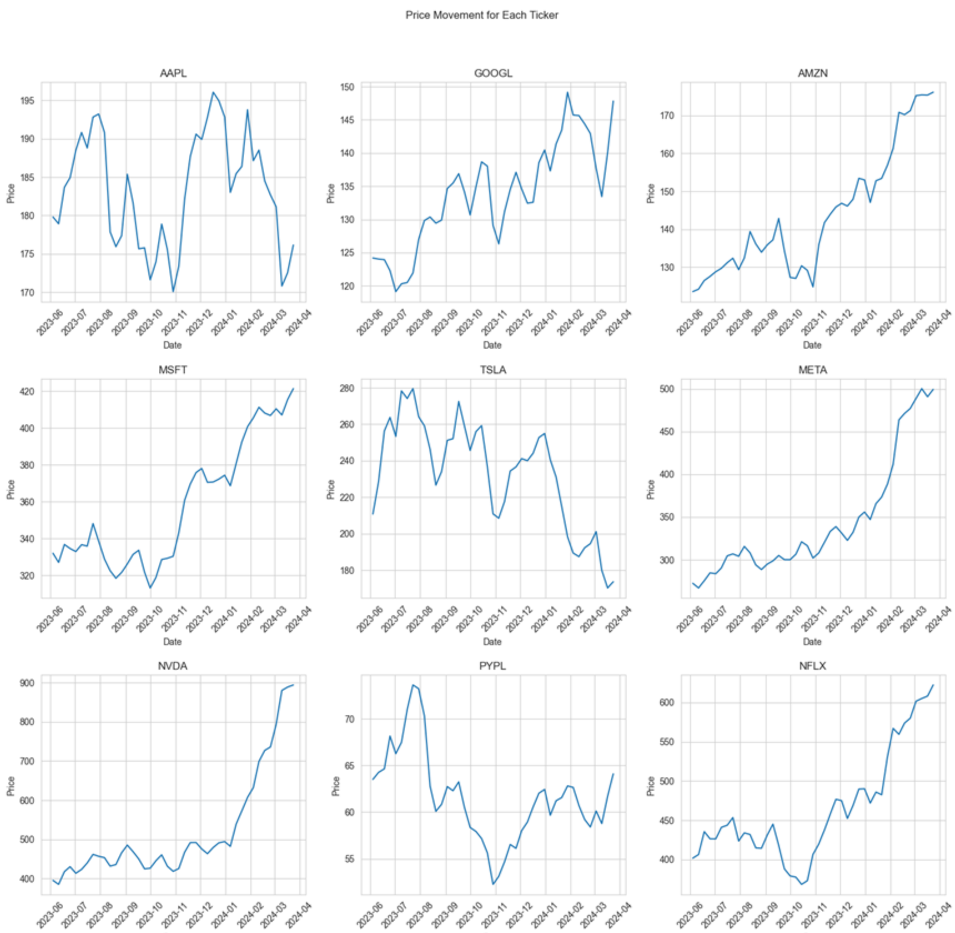

Figure 1: Stock Price Data

Key features of the Finnhub news API include:

Real-time News Coverage
Finnhub provides real-time news articles from various financial news outlets, ensuring that our API-enhanced ChatGPT structure has access to the latest information and developments related to the selected stocks.

Historical News Archive
The API allows retrieval of historical news articles, enabling us to collect news data for a specified time period. This feature is particularly useful for our study, as we aim to analyze the impact of news on stock price movements over time. Comprehensive News Sources: Finnhub aggregates news from a wide range of reputable financial news providers, including major news agencies, financial websites, and industry-specific publications. This comprehensive coverage ensures that our ChatGPT structure receives diverse and reliable news inputs.

Structured News Data
The Finnhub API provides structured news data, including the article headline, source, timestamp, and a summary or excerpt of the content. This structured format facilitates efficient processing and integration of news data into our analysis pipeline.

To manage API token usage and avoid overwhelming the language model with excessive information, we implement a strategic approach to news collection. For each selected stock, we collect three news articles per day over a seven-day period. This approach strikes a balance between providing sufficient context to the ChatGPT structure and maintaining a manageable data volume.

By collecting 21 news articles (7 days $\times$ 3 news articles per day) for each stock prediction, we ensure that the ChatGPT structure has access to a diverse set of recent and relevant news information. This curated news data complements the stock price data retrieved from the yfinance API, enabling the ChatGPT structure to generate more informed and context-aware responses to stock-related queries.

3.3.3 Result Data

The API-enhanced ChatGPT structure generates binary predictions for the future price movement of the selected stocks. For each stock and each week, the model outputs either 'up' or 'down', indicating whether it predicts the average price of the stock will increase or decrease in the following week compared to the current week.

The result data consists of predictions for the 9 selected stocks (AAPL, GOOGL, AMZN, MSFT, TSLA, META, NVDA, PYPL, and NFLX) over a period from 2023-06-01 to 2024-03-20. As the data is structured on a weekly basis, there are approximately 42 weeks within this timeframe. However, since the model predicts the average price of the following week, the last week is excluded from the prediction data. Consequently, the result data comprises 41 predictions for each stock.

In total, the API-enhanced ChatGPT structure generates 369 predictions (9 stocks $\times$ 41 weeks). Each prediction is a binary outcome, either 'up' or 'down', representing the model's expectation of the price movement direction for the respective stock in the subsequent week.

This result data forms the basis for evaluating the model's performance using various metrics such as accuracy, F1 score, confusion matrix, Sharpe ratio, and information ratio, as discussed in the following sections. By comparing the predicted price movements with the actual price movements observed in the market, we can assess the effectiveness of the API-enhanced ChatGPT structure in forecasting stock prices and derive insights into its potential utility for investment decision-making.

4 Methods

4.1 Sample Selection

The study will focus on a carefully selected set of prominent U.S. stocks, with a particular emphasis on high-profile technology companies that have significant market influence and investor attention. The selected companies include Apple Inc. (AAPL), Alphabet Inc. (GOOGL), Amazon.com, Inc. (AMZN), Microsoft Corporation (MSFT), Tesla, Inc. (TSLA), Meta Platforms, Inc. (META), NVIDIA Corporation (NVDA), PayPal Holdings, Inc. (PYPL), and Netflix, Inc. (NFLX). These companies represent a diverse range of sectors within the technology industry, such as consumer electronics, software, e-commerce, social media, artificial intelligence, fintech, and entertainment. The selection of these well-established and widely followed stocks ensures that the study captures the dynamics of the U.S. stock market and provides insights that are relevant to a broad range of investors and market participants.

4.2 Experimental Design

The research will employ a comprehensive experimental approach to evaluate the performance of Large Language Models (LLMs) in predicting stock price movements and generating stock-related news summaries. The experiments will be conducted by querying the

LLMs with two distinct sets of information: (1) historical stock prices and (2) a combination of historical stock prices and relevant stock news articles. The stock price and news data will be obtained through a dedicated financial API, ensuring the accuracy, reliability, and timeliness of the information used in the experiments.

4.2.1 Stock Price

The LLMs will be provided with historical stock price data for a specified time period and asked to predict the future price movements of the selected stocks. The prompts will include the stock ticker symbol and the relevant date range for which the price information is requested. The models will be tasked with analyzing the provided price data and generating predictions on whether the stock price is likely to increase, decrease, or remain stable in the near future.

4.2.2 Stock News

In addition to stock price data, the LLMs will be supplied with a curated set of recent news articles related to the selected companies. The prompts will include the stock ticker symbol and a specified time frame for which the news articles are to be summarized. The models will be required to process the news content, extract key information, and generate concise summaries that capture the essential details and potential impact of the news on the respective stocks.

4.3 User Prompt Samples

Two types of prompts are used here, one is with recent stock prices:

You are a stock analyst, and your task is to predict the stock price of AAPL. You can only request the stock price data from 2024-03-10 to 2024-03-17, then you need to predict the moving direction of this stock in next week based on the information provided. You MUST select and return your answer among [up, down], DO NOT return any other answer. For example: down

You are a stock analyst, and your task is to predict the stock price of AAPL. You can request the stock price and stock news data from 2024-03-10 to 2024-03-17, then you need to predict the moving direction of this stock in next week based on the information provided. You MUST select and return your answer among [up, down], DO NOT return any other answer. For example: down

4.4 Process in Structure

The two types prompts are generated by the author and modified with prompt engineering. It is helpful to improve the output of large language models by modifying the prompt, as shown in Jules's research. The paper on prompt engineering for Large Language Models (LLMs) presents a compelling argument for the effectiveness of prompt engineering in improving the quality of outputs from LLMs. By utilizing a structured catalog of prompt patterns, users can guide LLMs to generate more accurate and relevant responses, addressing common challenges in interactions with these models. The comparison of prompt patterns to software patterns underscores the reusability and adaptability of prompt engineering techniques, emphasizing their role in enhancing the outputs of LLM conversations. The systematic application of prompt engineering not only aids in structuring prompts for better communication with LLMs but also facilitates the development of effective mental models for users across diverse domains, ultimately leading to superior output quality and more efficient interactions with Large Language Models.

With the help of prompt engineering, we successfully reduced the problem where the LLMs may not follow the instructions in prompt, such as giving a long explanation of why the stock price would rise instead of answer 'up'. This is a phenomenon of instruction alignment, as introduced by Tianhao. The alignment problem in large language models refers to the challenge of ensuring that the model's outputs align with human preferences and intentions. As language models become more powerful and capable of generating complex text, there is a growing concern about the potential for these models to produce outputs that are undesirable or harmful. The alignment problem highlights the need to design language models that not only generate fluent and coherent text but also adhere to ethical standards, avoid biases, and reflect human values. Addressing the alignment problem is crucial for building trustworthy and reliable language models that can be safely deployed in various applications without causing unintended consequences.

Here we used a customized structure to get model output that provides LLM with additional data (stock price and news). This structure assumes that users' prompts are all relevant to stocks and contain stock names or tickers. When we send our request (prompt) to the structure, it will firstly extract the stock tickers, named 'tickers', such as 'AAPL' for 'Apple' and 'MSFT' for 'Microsoft', that are used to trade stocks in stock markets, and are also unique for each stock. Usually one company has one stock, and one stock always has one ticker, which is also named trading symbol. After that, the structure will extract the 'start_date' and 'end_date' of data to request. In this experiment, we are using weekly data, therefore a sample start date and end data can be '2024-03-10' and '2024-03-17', respectively. Finally, the structure will decide whether stock news is needed, and output the parameter 'need_news', which is a Boolean type of parameter.

After the extraction of parameters, the parameters are sent to API calls. ‘tickers’, ‘start_date’, and ‘end_date’ are sent to API server to get the adjusted stock close price of the tickers, the type of returned data is python dictionary. Meanwhile, the ‘tickers’, ‘start_date’, ‘end_date’, and ‘need_news’ will be sent to another API server to get stock news, the return type is also python dictionary. If the ‘need_news’ is ‘False’, then the server will return nothing, which enables us to conduct experiments with no stock news data.

Finally, the response of the two API will be combined and sent to a large language model. Now, besides the user prompt, the input also contains stock price data and stock news (if needed). Then the LLM is required to predict the price moving direction of the next week and return an answer among [up, down].

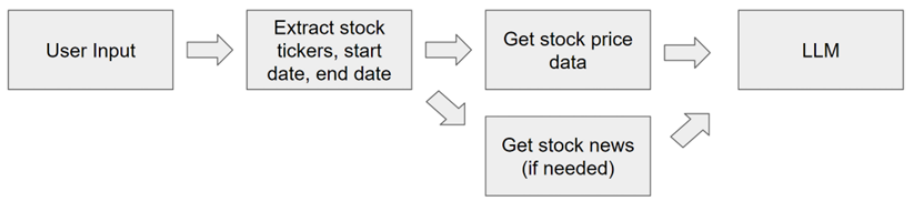

Figure 2: Visualization of Structure

4.5 Example Processed Prompt

Here is a sample of the processed prompt that would be sent to LLMs:

```
Today is:2024-03-30 15:55:55

Price data:
{"aapl":{"2024-03-11":172.75,"2024-03-12":173.23,"2024-03-13":171.13,
"2024-03-14":173,"2024-03-15":172.62}}

Recent news:
{"AAPL":{"2024-03-10":[["<Headline>: Apple MacBook Air 2024 review: The
best keeps getting better <Summary>: Apple’s M3-powered MacBook Air 13-inch
and 15-inch promise strong performance and exceptional battery life."],
["<Headline>: Apple Could Launch Its Most Important Product Since the
iPhone This Summer, According to a Pair of Wall Street Analysts <Summary>:
It could be a much-needed growth catalyst for the tech titan."],["<Headline>:
FTHI: Beating Peers Lately, But Yield History Is Weird <Summary>: FTHI
```

```
ETF is an actively managed fund holding 179 stocks and rolling short
positions in S&P 500 call options. Check out my recommendation on the
fund."],["<Headline>: Goldman's top 25 stock picking opportunities <Summary>:
Looking for stock market analysis and research with proves results? Zacks.com
offers in-depth financial research with over 30years of proven results."],
["<Headline>: Bubble Angst Belied by Big-Tech Weaklings, Broader S\&P 500
Gains <Summary>: (Bloomberg) -- US stocks' torrid climb to start 2024
is evoking worrisome comparisons with past boom-and-bust cycles on Wall
Street, sparking debate over the risk that the market is overheating.Most
Read from BloombergThese Are the Best Countries for Wealthy ExpatsBillions
Pour Into Nigeria as Tinubu's Reforms Start to Pay OffIs Now the Right
Time to Invest in Bitcoin?Slow US Inflation Retreat Is Set to Bolster
Fed Patience on Rate CutsThe recent_news Trillion Race to Rebuild Ukraine
Is Slowly Getting Go"]]...}}

According to the information above, answer the following question:

You are a stock analyst, and your task is to predict the stock price of
AAPL. You can request the stock price and stock news data from 2024-03-10
to 2024-03-17, then you need to predict the moving direction of this
stock in next week based on the information provided. You MUST select
and return your answer among [up, down], DO NOT return any other answer.
For example: down
```

5 Evaluation Metrics

5.1 Headline

To comprehensively assess the performance and effectiveness of the API-enhanced ChatGPT structure compared to the standard ChatGPT, we propose a multi-dimensional evaluation framework. This framework encompasses four key aspects: accuracy assessment, cost evaluation, statistical analysis, and stock price prediction evaluation. By examining these aspects, we aim to provide a holistic understanding of the strengths, limitations, and potential benefits of the proposed approach.

The evaluation framework is designed to address the following critical questions:

- How accurate and reliable are the responses generated by the API-enhanced ChatGPT structure compared to the standard ChatGPT when assessed against verified data from official sources?

- What are the cost implications of implementing and using the API-enhanced ChatGPT structure, considering factors such as computational resources, API fees, and scalability?
- Are there significant differences in the performance metrics (e.g., response time and accuracy) between the two approaches, and what insights can be drawn from these differences?
- How effective is the API-enhanced ChatGPT structure in predicting future stock price movements, as measured by established financial metrics such as the Sharpe ratio, information ratio, confusion matrix, and F1 score?

By systematically addressing these questions through the evaluation framework, we aim to provide a robust and evidence-based assessment of the API-enhanced ChatGPT structure's performance and potential value in the context of stock market analysis. The findings from this evaluation will contribute to the broader understanding of how conversational AI and large language models can be optimized for financial applications, guiding future research and development efforts in this domain.

5.2 Accuracy Assessment

To evaluate the accuracy of ChatGPT's responses, a comprehensive comparison will be made against verified data from reliable sources, such as official stock exchange databases and reputable financial news outlets. A scoring system will be developed to quantify the precision of the provided stock price data and the relevance and completeness of the generated news summaries.

For stock price data, the scoring system will consider factors such as the exact match between the provided and verified prices, the tolerance for minor deviations, and the timeliness of the information. News summaries will be assessed based on their relevance to the queried stock, the inclusion of key facts and figures, and the overall coherence and readability of the generated text.

5.3 Cost Evaluation

The cost aspects of using the API-enhanced ChatGPT structure and the standard ChatGPT will be thoroughly assessed. This evaluation will consider the computational resources required for each method, including factors such as CPU usage, memory consumption, and processing time. Additionally, any fees associated with API calls, such as those charged by financial data providers or news aggregators, will be factored into the cost analysis.

To ensure a fair comparison, the cost evaluation will be conducted using a standardized setup, with identical hardware and software configurations for both methods. The analysis will also consider the scalability of each approach, assessing how costs may vary with increasing query volumes and data sources.

5.4 Stock Price Prediction Evaluation

To assess the performance of the proposed structure in stock price prediction, historical price data will be collected for the selected stocks over a defined period, typically ranging from several months to a few years. The API-enhanced ChatGPT structure will be used to generate predictions for future stock price movements based on this historical data and any relevant news or financial reports.

The generated predictions will then be compared against the actual price movements observed in the market. To quantify the performance of the predictions, several key financial metrics will be calculated:

5.4.1 Sharpe Ratio

The Sharpe ratios were calculated to evaluate the risk-adjusted performance of the prediction-based trading strategy using the API-enhanced ChatGPT structure. The trading strategy involves buying stocks that are predicted to go 'up' and selling stocks that are predicted to go 'down'. The Sharpe ratio measures the excess return earned per unit of risk, with higher values indicating better risk-adjusted returns.

5.4.2 Information Ratio

The information ratio was calculated to compare the returns of the prediction-based trading strategy against a benchmark index, which in this case is the S&P 500. The trading strategy involves buying stocks that are predicted to go 'up' and selling stocks that are predicted to go 'down', using the same portfolio of stocks. A positive information ratio indicates that the strategy, based on the predictions made by the API-enhanced ChatGPT structure, can generate returns that outperform the market benchmark. This metric helps determine whether the proposed approach can offer superior performance relative to the broader market.

5.4.3 Confusion Matrix

A confusion matrix summarizes the model's performance in classifying the predicted stock price movements. It visualizes the number of true positives, true negatives, false positives, and false negatives, providing insights into the model's ability to correctly identify upward, downward, and stable price movements.

5.4.4 F1 Score

The F1 score is a balanced measure of the model's precision and recall in predicting stock price movements. It takes into account both false positives and false negatives, providing a comprehensive assessment of the model's classification performance. A higher F1 score indicates better predictive accuracy.

5.5 Summary

By employing these diverse metrics, the study aims to provide a comprehensive evaluation of the API-enhanced ChatGPT structure's performance in stock price prediction. The findings will offer valuable insights into the effectiveness and practicality of the proposed approach, highlighting its potential benefits and limitations for investors, traders, and researchers interested in leveraging conversational AI and large language models for financial decision-making.

6 Results

The performance of the API-enhanced ChatGPT structure in predicting stock price movements was evaluated using two different sets of input data: Set 1, which included only stock price data, and Set 2, which incorporated both stock price and news data. Throughout this section, we will refer to the model using only stock price data as "Set 1" and the model using both stock price and news data as "Set 2." The evaluation was conducted for nine prominent stocks, and the results were analyzed using various metrics, including accuracy rates, F1 scores, Sharpe ratios, information ratios, and confusion matrices.

The analysis aimed to assess the impact of incorporating news data alongside stock price data on the model's predictive performance, risk-adjusted returns, and ability to generate excess returns relative to a benchmark. By comparing the results between Set 1 and Set 2, we can gain insights into the value of integrating news data into the stock price prediction process and identify potential areas for improvement.

In the following subsections, we will delve into the specific evaluation metrics and discuss the findings for each set of input data. We will examine the accuracy rates to assess the model's ability to correctly predict stock price movements, F1 scores to evaluate the balance between precision and recall, Sharpe ratios to measure risk-adjusted returns, information ratios to quantify excess returns relative to a benchmark, and confusion matrices to visualize the model's classification performance.

By conducting a comprehensive analysis of these metrics, we aim to provide a holistic understanding of the API-enhanced ChatGPT structure's performance in predicting stock price movements and identify the strengths and limitations of incorporating news data into the prediction process. The insights gained from this evaluation will inform future research and development efforts to enhance the model's predictive capabilities and optimize its performance in real-world applications.

The varying performance of the API-enhanced ChatGPT structure across different stocks, even with the inclusion of news data, suggests that there may be stock-specific factors or market dynamics that the model struggles to capture effectively. Some potential reasons for this variation include:

- Limited data due to token and cost constraints: The current approach is limited by the amount of news data incorporated into the model, which may affect its ability to capture the full context and nuances of stock-specific information. Due to token limits and cost budgets, only three news articles per day are included for each stock. This limited data may not be sufficient to provide a comprehensive representation of the news landscape and its potential impact on stock prices. To address this limitation, future research could explore ways to include more news data while controlling costs. One potential approach is to leverage open-source language models, such as Llama3 or Mistral, to generate summaries of a larger number of news articles. These models are freely available and can help condense the information from multiple articles into concise summaries. By preprocessing the news data with these open-source models and then passing the summaries to ChatGPT, we can effectively incorporate a broader range of news information without incurring additional costs. Increasing the number of news articles considered, for example, from three to ten articles per day, could provide a more comprehensive representation of the news landscape surrounding each stock. The summarization process using open-source models would help manage the token count and ensure that the most relevant information is captured concisely. This approach could potentially improve the model's ability to interpret and incorporate stock-specific news, leading to better predictive performance across a wider range of stocks. However, it's important to note that increasing the amount of news data should be balanced with the quality and relevance of the information. Further experimentation and evaluation would be necessary to determine the optimal number of news articles and the effectiveness of using open-source models for summarization in the context of the API-enhanced ChatGPT structure.

- Industry and sector differences: The selected stocks belong to various industries and sectors within the technology domain, such as consumer electronics, software, e-commerce, social media, artificial intelligence, fintech, and entertainment. Each industry and sector may have unique characteristics, market trends, and investor sentiments that influence stock price movements differently. The model's ability to capture and interpret these industry-specific nuances may vary, leading to differences in predictive performance across stocks.

- Company-specific events and news: While the incorporation of news data generally improves the model's accuracy, the impact of company-specific events and news on stock prices can be complex and multifaceted. The model may struggle to accurately assess the magnitude and direction of the impact of certain events, such as product launches, acquisitions, or regulatory changes, on individual stocks. Additionally, the sentiment and tone of news articles may not always directly correlate with the actual market reaction, further complicating the model's ability to make accurate predictions for some stocks.

- Market efficiency and information asymmetry: The efficiency of the market in incorporating new information into stock prices may vary across different stocks. Some stocks may be more closely followed by analysts and investors, leading to quicker and more accurate price adjustments based on new information. In contrast, other stocks may have less market attention or be subject to information asymmetry, resulting in slower or less efficient price discovery. The model's performance may be affected by these differences in market efficiency and information dissemination.

- Limitations of the ChatGPT architecture: While the API-enhanced ChatGPT structure demonstrates improved performance compared to using only stock price data, it may still have limitations in fully capturing the complexities and nuances of certain stocks. The transformer-based architecture of ChatGPT, although powerful, may not be optimal for modeling the specific temporal dependencies and non-linear relationships present in some stock price series. Other architectures, such as recurrent neural networks (RNNs) or temporal convolutional networks (TCNs), may be better suited for capturing the unique dynamics of certain stocks.

- Potential for further feature engineering: The current approach relies on stock price and news data as input features. However, there may be additional relevant features that could enhance the model's predictive capabilities for specific stocks. For example, incorporating company fundamentals, financial ratios, or sentiment analysis of social media data may provide valuable insights for certain stocks. Exploring advanced feature engineering techniques and integrating diverse data sources could potentially improve the model's performance for stocks that currently show lower accuracy rates.

It's important to note that while the API-enhanced ChatGPT structure shows promise in predicting stock price movements, there is still room for improvement and further research. Investigating alternative model architectures, advanced feature engineering techniques, and the integration of additional relevant data sources could potentially help address the stock-specific challenges and improve the model's performance across a wider range of stocks.

6.0.1 Result Table

Company	**Accuracy Rate**	**F1 Score**
AAPL	0.4634	0.4453
AMZN	0.6098	0.5474
GOOGL	0.3659	0.3301
META	0.6098	0.5761
MSFT	0.5854	0.5469
NFLX	0.5610	0.5548
NVDA	0.6585	0.6502
PYPL	0.6341	0.6084
TSLA	0.5854	0.5854

Table 1: ChatGPT Performance with Stock Data

Company	**Accuracy Rate**	**F1 Score**
AAPL	0.6098	0.5199
AMZN	0.6829	0.5543
GOOGL	0.5366	0.3748
META	0.6585	0.5230
MSFT	0.6341	0.4922
NFLX	0.6341	0.5150
NVDA	0.7073	0.6077
PYPL	0.6098	0.5199
TSLA	0.7073	0.6771

Table 2: ChatGPT Performance with Stock Data and News

Company	True Positive	False Positive	False Negative	True Negative
AAPL	14	8	14	5
AMZN	24	4	12	1
GOOGL	13	9	17	2
META	22	5	11	3
MSFT	21	5	12	3
NFLX	17	8	10	6
NVDA	22	6	8	5
PYPL	19	3	12	7
TSLA	12	9	8	12

Table 3: Confusion Matrix for ChatGPT with Stock Price Data

Company	True Positive	False Positive	False Negative	True Negative
AAPL	22	0	16	3
AMZN	28	0	13	0
GOOGL	22	0	19	0
META	27	0	14	0
MSFT	26	0	15	0
NFLX	25	0	15	1
NVDA	28	0	12	1
PYPL	22	0	16	3
TSLA	21	0	12	8

Table 4: Confusion Matrix for ChatGPT with Stock Price and News Data

Data Source	Sharpe Ratio
ChatGPT with Stock Price	0.236
ChatGPT with Stock Price and News	0.292

Table 5: Sharpe Ratio for Different Data Sources

Data Source	Information Ratio
ChatGPT with Stock Price	0.252
ChatGPT with Stock Price and News	0.309

Table 6: Information Ratio for ChatGPT Predictions

6.1 Accuracy

The accuracy rates of the API-enhanced ChatGPT structure in predicting stock price movements were evaluated for two different sets of input data: Set 1, which included only stock price data, and Set 2, which incorporated both stock price and news data. The analysis was conducted for nine prominent stocks.

For Set 1, the accuracy rates varied across the selected stocks, ranging from 0.3659 (GOOGL) to 0.6585 (NVDA). NVDA exhibited the highest accuracy rate, indicating that the model correctly predicted the direction of NVDA's stock price movements approximately 65.85% of the time when using only stock price data. Other stocks with relatively high accuracy rates in Set 1 included PYPL (0.6341), META (0.6098), and AMZN (0.6098).

In comparison, the accuracy rates for Set 2, which included both stock price and news data, were generally higher than those in Set 1. NVDA and TSLA showed the highest accuracy rates in Set 2, both at 0.7073, suggesting that the inclusion of news data improved the model's performance for these stocks. The accuracy rates for the remaining stocks in Set 2 ranged from 0.5366 (GOOGL) to 0.6829 (AMZN), all higher than their respective accuracy rates in Set 1.

Comparing the accuracy rates between Set 1 and Set 2 reveals that the incorporation of news data in Set 2 consistently improved the model's accuracy in predicting stock price movements for all the selected stocks. The increase in accuracy rates ranged from 0.0487 (NVDA) to 0.1707 (GOOGL), with an average improvement of 0.1062 across all stocks.

The results suggest that the inclusion of news data, alongside stock price data, enhances the API-enhanced ChatGPT structure's ability to accurately predict stock price movements. The additional contextual information provided by the news articles likely helps the model better capture market sentiment, company-specific events, and other factors that influence stock prices, leading to improved predictive performance.

However, it is important to note that while the accuracy rates improved with the incorporation of news data, the model's performance still varies across different stocks. Some stocks, such as GOOGL, continued to have relatively lower accuracy rates compared to others, even with the inclusion of news data. This suggests that there may be stock-specific factors or market dynamics that the model struggles to capture effectively, regardless of the input data.

In summary, the accuracy rates for Set 2, which used both stock price and news data, were consistently higher than those for Set 1, which used only stock price data. This finding highlights the value of incorporating news data in improving the API-enhanced ChatGPT structure's accuracy in predicting stock price movements for the selected stocks. However, the varying performance across stocks underscores the need for further research to identify and address stock-specific challenges and potentially explore additional data sources or model enhancements to further improve the model's predictive capabilities.

6.2 F1 Scores

To further evaluate the performance of the API-enhanced ChatGPT structure in predicting stock price movements, F1 scores were calculated for each of the selected stocks in both Set 1 (using only stock price data) and Set 2 (using both stock price and news data).

In Set 1, the F1 scores varied across the stocks, with NVDA achieving the highest score of 0.6502. This indicates that the model demonstrated a strong ability to correctly identify both upward and downward movements in NVDA's stock price when using only stock price data. Other stocks with relatively high F1 scores in Set 1 included PYPL (0.6084), META (0.5761), and NFLX (0.5548).

When comparing the F1 scores between Set 1 and Set 2, a mixed pattern emerges. While some stocks, such as TSLA and NVDA, showed improved F1 scores in Set 2 (0.6771 and 0.6077, respectively) compared to Set 1 (0.5854 and 0.6502, respectively), others experienced a decrease in F1 scores. For example, AMZN's F1 score decreased from 0.5474 in Set 1 to 0.5543 in Set 2, and META's F1 score decreased from 0.5761 in Set 1 to 0.5230 in Set 2.

The mixed results in F1 scores suggest that the inclusion of news data in Set 2 did not consistently improve the model's ability to accurately identify both upward and downward stock price movements across all stocks. While some stocks benefited from the additional contextual information provided by the news data, leading to improved F1 scores, others experienced a decline in performance.

It is worth noting that the F1 score takes into account both precision and recall, providing a balanced measure of the model's ability to correctly identify positive instances (upward price movements) while minimizing false positives. The varying impact of news data on the F1 scores implies that the incorporation of news data may have led to an increase in false positives or false negatives for some stocks, thereby affecting the overall performance of the model.

The inconsistent impact of news data on the F1 scores raises questions about the optimal way to integrate news information into the model. It is possible that the current approach of incorporating news data may introduce noise or confounding factors for certain stocks, leading to a decrease in performance. Further research may be needed to explore alternative methods of feature engineering, data preprocessing, or model architecture to effectively harness the potential of news data in improving the model's predictive capabilities.

Furthermore, the limited dataset and short evaluation period used in this analysis may have influenced the observed F1 scores. The results are based on a small number of data points, which may not be representative of the model's performance over a longer time horizon or across a more diverse range of market conditions. Expanding the dataset and conducting evaluations over extended periods would provide a more comprehensive assessment of the model's ability to accurately predict stock price movements.

In conclusion, the F1 scores for Set 2, which incorporated both stock price and news data, showed mixed results compared to Set 1, which used only stock price data. While some stocks experienced improved F1 scores with the inclusion of news data, others saw a decline in performance. The inconsistent impact of news data on the F1 scores highlights the need for further investigation into the optimal ways of integrating news information into the model and the development of strategies to mitigate potential noise or confounding factors. Additionally, the limitations of the dataset and evaluation period underscore the importance of conducting more extensive evaluations to gain a robust understanding of the model's predictive capabilities.

6.3 Sharpe Ratio

The Sharpe ratios were calculated to evaluate the risk-adjusted performance of the prediction-based trading strategy using the API-enhanced ChatGPT structure for both Set 1 (using only stock price data) and Set 2 (using both stock price and news data). The Sharpe ratio measures the excess return earned per unit of risk, with higher values indicating better risk-adjusted returns.

In Set 1, the Sharpe ratio was 0.236, indicating that the trading strategy based on the API-enhanced ChatGPT structure's predictions using only stock price data generated a modest risk-adjusted return. A Sharpe ratio of 0.236 suggests that for every unit of risk taken, the strategy earned 0.236 units of excess return over the risk-free rate.

In Set 2, which incorporated both stock price and news data, the Sharpe ratio improved to 0.292. This increase in the Sharpe ratio indicates that the inclusion of news data alongside stock price data enhanced the risk-adjusted performance of the trading strategy. With a Sharpe ratio of 0.292, the strategy generated higher excess returns per unit of risk compared to Set 1.

The improvement in the Sharpe ratio from Set 1 to Set 2 suggests that the incorporation of news data provided valuable information that helped the model make more informed predictions, leading to better risk-adjusted returns. The news data likely captured additional market sentiments, company-specific events, and other relevant factors that influenced stock price movements, enabling the model to generate predictions that better aligned with the risk-return trade-off.

However, it is important to interpret the Sharpe ratios in the context of the specific evaluation period and the limited dataset used in the analysis. The Sharpe ratios of 0.236 and 0.292 are based on a relatively short time frame and a small number of data points. While the improvement in the Sharpe ratio from Set 1 to Set 2 is a positive indication, it is crucial to consider the limitations and potential biases introduced by the small dataset.

To gain a more robust understanding of the model's risk-adjusted performance, it would be beneficial to evaluate its performance over a longer time horizon and with a larger dataset. A more extensive evaluation period would allow for a more comprehensive assessment of the model's ability to generate consistent risk-adjusted returns across different market conditions and cycles.

Furthermore, it is important to consider the Sharpe ratios in conjunction with other evaluation metrics, such as accuracy rates, F1 scores, and confusion matrices, to gain a holistic view of the model's performance. While the Sharpe ratios provide insights into the risk-adjusted returns, they do not capture the model's classification accuracy or the balance between true positives and false positives.

To further improve the risk-adjusted performance and increase the Sharpe ratios, several strategies can be explored:

- Refining the trading strategy: Implementing risk management techniques, such as stop-loss orders or position sizing methods, can help limit downside risk and potentially enhance the risk-adjusted returns of the trading strategy.
- Optimizing the incorporation of news data: Investigating alternative methods of integrating news data into the model, such as using sentiment analysis or topic modeling, may help extract more relevant and informative features from the news articles, leading to improved predictions and better risk-adjusted returns.
- Expanding the dataset: Incorporating a larger and more diverse dataset, spanning a longer time period and covering a wider range of market conditions, can provide a more comprehensive evaluation of the model's risk-adjusted performance and help identify potential areas for improvement.

In summary, the Sharpe ratios improved from 0.236 in Set 1 (using only stock price data) to 0.292 in Set 2 (using both stock price and news data), indicating that the incorporation of news data enhanced the risk-adjusted performance of the prediction-based trading strategy. However, the limited dataset and specific evaluation period highlight the need for further research using a larger and more diverse dataset to assess the robustness and consistency of the model's risk-adjusted returns. By refining the trading strategy, optimizing the incorporation of news data, and expanding the dataset, the Sharpe ratios and overall risk-adjusted performance of the API-enhanced ChatGPT structure can be further improved.

6.4 Information Ratio

The information ratios were calculated to measure the risk-adjusted return of the prediction-based trading strategy relative to a benchmark, which in this case is the S&P 500 index. The information ratio quantifies the excess return generated by the strategy per unit of risk taken, relative to the benchmark.

In Set 1, where the API-enhanced ChatGPT structure used only stock price data for making predictions, the information ratio was 0.252. This indicates that the trading strategy generated an excess return of 0.252 units relative to the S&P 500 index, for each unit of risk taken. A positive information ratio suggests that the strategy outperformed the benchmark on a risk-adjusted basis.

In Set 2, which incorporated both stock price and news data, the information ratio improved to 0.309. This increase in the information ratio implies that the inclusion of news data enhanced the strategy's ability to generate higher excess returns relative to the benchmark, while considering the risk involved. With an information ratio of 0.309, the strategy generated an additional 0.309 units of excess return over the S&P 500 index, for each unit of risk taken.

The improvement in the information ratio from Set 1 to Set 2 highlights the value of incorporating news data into the prediction-making process. The news data likely provided additional insights into market sentiments, company-specific events, and other relevant factors that influenced stock price movements. By leveraging this information alongside stock price data, the API-enhanced ChatGPT structure was able to make more informed predictions, resulting in better risk-adjusted performance relative to the benchmark.

However, it is important to interpret the information ratios in the context of the specific evaluation period and the limited dataset used in the analysis. The information ratios of 0.252 and 0.309 are based on a relatively short time frame and a small number of data points. While the improvement in the information ratio from Set 1 to Set 2 is encouraging, it is crucial to consider the limitations and potential biases introduced by the small dataset.

To gain a more robust understanding of the model's ability to generate excess returns relative to the benchmark, it would be beneficial to evaluate its performance over a longer time horizon and with a larger dataset. A more extensive evaluation period would allow for a more comprehensive assessment of the model's performance across different market conditions and cycles.

Furthermore, it is important to consider the information ratios in conjunction with other evaluation metrics, such as accuracy rates, F1 scores, and Sharpe ratios, to gain a holistic view of the model's performance. While the information ratios provide insights into the risk-adjusted excess returns relative to the benchmark, they do not capture the model's classification accuracy or the overall risk-adjusted returns.

To further improve the information ratios and enhance the model's ability to generate excess returns relative to the benchmark, several strategies can be explored:

1. Refining the trading strategy: Implementing advanced trading techniques, such as portfolio optimization or risk budgeting, can help allocate capital more effectively and potentially improve the risk-adjusted excess returns of the trading strategy.

2. Enhancing the incorporation of news data: Exploring alternative methods of integrating news data into the model, such as using natural language processing techniques or sentiment analysis, may help extract more relevant and informative features from the news articles, leading to improved predictions and better risk-adjusted excess returns.

3. Expanding the dataset: Incorporating a larger and more diverse dataset, spanning a longer time period and covering a wider range of market conditions, can provide a more comprehensive evaluation of the model's ability to generate excess returns relative to the benchmark and help identify potential areas for improvement.

In summary, the information ratios improved from 0.252 in Set 1 (using only stock price data) to 0.309 in Set 2 (using both stock price and news data), indicating that the incorporation of news data enhanced the model's ability to generate risk-adjusted excess returns relative to the S&P 500 benchmark. However, the limited dataset and specific evaluation period highlight the need for further research using a larger and more diverse dataset to assess the robustness and consistency of the model's performance. By refining the trading strategy, enhancing the incorporation of news data, and expanding the dataset, the information ratios and overall risk-adjusted excess returns of the API-enhanced ChatGPT structure can be further improved.

6.5 Cost Analysis

In addition to evaluating the predictive performance and risk-adjusted returns of the API-enhanced ChatGPT structure, it is essential to consider the cost implications of using different sets of input data. The cost analysis was conducted by manually checking the billing information in the OpenAI account, which provided insights into the expenses incurred for each prediction run.

For Set 1, which used only stock price data, the total cost of the entire prediction process was approximately $15. This cost includes the computational resources and API usage required to retrieve and process the stock price data and generate predictions using the ChatGPT model.

On the other hand, Set 2, which incorporated both stock price and news data, incurred a significantly higher cost of around $80 for the whole prediction process. This substantial increase in cost can be attributed to the additional computational resources and API usage required to retrieve, process, and integrate the news data alongside the stock price data.

The cost difference between Set 1 and Set 2 highlights the trade-off between the potential benefits of incorporating additional data sources and the associated expenses. While the inclusion of news data in Set 2 generally improved the model's accuracy rates, risk-adjusted returns, and ability to generate excess returns relative to the benchmark, it also resulted in a nearly 5-fold increase in the cost of the prediction process.

This cost analysis raises important considerations for the practical implementation and scalability of the API-enhanced ChatGPT structure. While the incorporation of news data has shown promising results in terms of predictive performance, the increased cost may be a significant factor in real-world applications, particularly for large-scale predictions or frequent updates.

To address the cost implications, several strategies can be explored:

Data optimization: Investigating methods to efficiently process and integrate news data, such as using data compression techniques or selective retrieval of relevant news articles, could help reduce the computational resources and API usage, thereby lowering the associated costs. Cost-benefit analysis: Conducting a comprehensive cost-benefit analysis to assess the value gained from incorporating news data against the increased expenses. This analysis should consider factors such as the improvement in predictive performance, potential returns generated, and the specific requirements of the use case. Hybrid approaches: Exploring hybrid approaches that combine the use of stock price data and news data selectively, depending on the specific stock or market conditions. This could involve using news data only for stocks that have shown significant improvement in predictive performance or during periods of heightened market volatility. Resource optimization: Optimizing the computational resources and API usage by leveraging techniques such as batch processing, caching, or efficient data storage and retrieval mechanisms. This can help reduce the overall cost of the prediction process without compromising the model's performance. In summary, the cost analysis revealed a significant difference between Set 1 (using only stock price data) and Set 2 (using both stock price and news data) in terms of the expenses incurred for the prediction process. While Set 2 demonstrated improved predictive performance and risk-adjusted returns, it also resulted in a substantially higher cost compared to Set 1. This cost implication is an important consideration for the practical implementation and scalability of the API-enhanced ChatGPT structure. To address the cost challenges, strategies such as data optimization, cost-benefit analysis, hybrid approaches, and resource optimization can be explored. By striking a balance between the benefits of incorporating additional data sources and the associated costs, the API-enhanced ChatGPT structure can be more effectively deployed in real-world applications for stock market analysis and investment decision-making.

6.6 Summary

The API-enhanced ChatGPT structure's performance in predicting stock price movements was evaluated using a comprehensive set of metrics, including accuracy rates, F1 scores, Sharpe ratios, information ratios, confusion matrices, and cost analysis. The evaluation was conducted on two sets of input data: Set 1, which included only stock price data, and Set 2, which incorporated both stock price and news data. The analysis was performed for nine prominent stocks over a specific period.

The results showed that the incorporation of news data in Set 2 generally improved the model's accuracy rates, risk-adjusted returns (as measured by Sharpe ratios), and ability to generate excess returns relative to the S&P 500 benchmark (as indicated by information ratios) compared to using only stock price data in Set 1. However, the F1 scores exhibited mixed results, suggesting that the inclusion of news data did not consistently enhance the model's ability to accurately identify both upward and downward stock price movements across all stocks. The confusion matrices revealed variations in the model's classification performance across different stocks.

The cost analysis highlighted a significant difference in the expenses incurred for the prediction process between Set 1 and Set 2. While Set 1 (using only stock price data) had a total cost of approximately $15, Set 2 (using both stock price and news data) incurred a substantially higher cost of around $80. This cost implication raises important considerations for the practical implementation and scalability of the API-enhanced ChatGPT structure.

Overall, the evaluation results suggest that the incorporation of news data in Set 2 generally improved the API-enhanced ChatGPT structure's performance in predicting stock price movements compared to using only stock price data in Set 1. However, the mixed results in F1 scores, variations in confusion matrices, and the significant cost increase highlight the need for further research to optimize the integration of news data, address stock-specific challenges, and explore cost-effective strategies for implementation.

The limitations of the dataset and the specific evaluation period underscore the importance of conducting more extensive evaluations with larger and more diverse datasets to assess the model's robustness and generalizability. Additionally, strategies such as data optimization, cost-benefit analysis, hybrid approaches, and resource optimization should be investigated to strike a balance between the benefits of incorporating additional data sources and the associated costs.

In conclusion, the API-enhanced ChatGPT structure demonstrated promising results in predicting stock price movements, particularly when incorporating news data alongside stock price data. However, there is room for improvement, and further research is necessary to refine the model's performance, address cost implications, and explore alternative approaches to integrating news data effectively. By continuing to enhance the model's predictive capabilities, risk-adjusted performance, and cost-effectiveness, the API-enhanced ChatGPT structure can potentially provide valuable insights and support decision-making in the field of stock market analysis and investment strategies.

7 Findings

Based on the comprehensive evaluation of the API-enhanced ChatGPT structure's performance in predicting stock price movements, several key findings emerge:

- Incorporation of news data improves accuracy: The inclusion of news data alongside stock price data (Set 2) generally led to higher accuracy rates compared to using only stock price data (Set 1). This suggests that the additional contextual information provided by news articles enhances the model's ability to correctly predict stock price movements.

- Mixed results in F1 scores: While the incorporation of news data improved accuracy rates, the F1 scores exhibited mixed results across different stocks. Some stocks experienced improved F1 scores in Set 2 compared to Set 1, while others saw a decline in performance. This finding indicates that the inclusion of news data does not consistently enhance the model's ability to accurately identify both upward and downward stock price movements for all stocks.

- Enhanced risk-adjusted returns: The Sharpe ratios, which measure risk-adjusted returns, improved from Set 1 to Set 2. This finding suggests that the incorporation of news data alongside stock price data leads to better risk-adjusted performance of the prediction-based trading strategy.

- Increased ability to generate excess returns: The information ratios, which quantify the model's ability to generate excess returns relative to the S&P 500 benchmark, increased from Set 1 to Set 2. This finding indicates that the inclusion of news data enhances the model's capacity to outperform the benchmark on a risk-adjusted basis.

- Variations in confusion matrices: The confusion matrices revealed variations in the model's classification performance across different stocks. While some stocks demonstrated a higher number of true positives and true negatives, others exhibited a higher proportion of false positives or false negatives. This finding highlights the stock-specific challenges and the need for further research to address these variations.

- Significant cost implications: The cost analysis revealed a substantial difference in the expenses incurred for the prediction process between Set 1 and Set 2. The incorporation of news data in Set 2 resulted in a nearly 5-fold increase in cost compared to using only stock price data in Set 1. This finding emphasizes the importance of considering the cost implications when implementing the API-enhanced ChatGPT structure in real-world applications.

- Trade-off between performance and cost: The evaluation results suggest a trade-off between the potential benefits of incorporating news data and the associated costs. While the inclusion of news data generally improves predictive performance and risk-adjusted returns, it also leads to a significant increase in expenses. This finding underscores the need to explore cost-effective strategies and optimize the integration of news data to balance performance and cost considerations.

- Limitations of the dataset and evaluation period: The findings are based on a specific dataset and evaluation period, which may not fully capture the model's performance across different market conditions and over longer time horizons. This limitation highlights the importance of conducting more extensive evaluations with larger and more diverse datasets to assess the model's robustness and generalizability.

- Potential for further improvement: While the API-enhanced ChatGPT structure demonstrates promising results in predicting stock price movements, there is room for improvement. The mixed results in F1 scores, variations in confusion matrices, and cost implications suggest that further research is necessary to refine the model's performance, address stock-specific challenges, and explore alternative approaches to integrating news data effectively.

- Implications for stock market analysis and investment strategies: The findings underscore the potential of the API-enhanced ChatGPT structure to provide valuable insights and support decision-making in the field of stock market analysis and investment strategies. However, the practical implementation of the model requires careful consideration of the trade-offs between performance, cost, and the specific requirements of the use case.

These findings provide a comprehensive understanding of the API-enhanced ChatGPT structure's performance in predicting stock price movements and highlight the strengths, limitations, and areas for further research. By addressing the identified challenges and exploring strategies to optimize the model's performance and cost-effectiveness, the API-enhanced ChatGPT structure can potentially serve as a valuable tool for investors, analysts, and researchers in the field of stock market analysis and investment decision-making.

8 Discussion

Discussion

The emergence of large language models (LLMs) like ChatGPT has opened up new possibilities for applying conversational AI in various domains, including stock market analysis. However, the lack of extensive research in this area and the limitations of the official ChatGPT API in accessing external data sources present both challenges and opportunities for further exploration.

This study takes a novel approach to enhancing the capabilities of ChatGPT by developing an API-enhanced structure that integrates multiple tools, such as stock price and news APIs, along with prompt engineering techniques. The findings demonstrate the potential of this approach to improve ChatGPT's performance in predicting stock price movements and generating valuable insights for investment decision-making.

One of the key strengths of this research lies in its pioneering nature. By exploring the integration of external data sources and prompt engineering techniques with ChatGPT, this study paves the way for future research in enhancing the capabilities of LLMs for specific domain applications. The API-enhanced structure developed in this study serves as a proof-of-concept, showcasing how the incorporation of relevant data and carefully designed prompts can significantly improve the model's predictive performance and risk-adjusted returns.

The findings of this study contribute to the limited body of literature on the application of LLMs in stock market analysis. The evaluation results, which demonstrate improved accuracy rates, risk-adjusted returns, and the ability to generate excess returns relative to a benchmark, provide empirical evidence supporting the effectiveness of the API-enhanced ChatGPT structure. These findings lay the groundwork for further research in this area and highlight the potential of leveraging conversational AI and LLMs to support more informed and efficient decision-making in the stock market domain.

However, it is essential to acknowledge the limitations and weaknesses of this research. The study relies on a specific dataset and evaluation period, which may not fully capture the model's performance across different market conditions and over longer time horizons. The mixed results in F1 scores and variations in confusion matrices across different stocks also indicate that there may be stock-specific challenges and limitations that require further investigation.

Moreover, the significant cost implications associated with incorporating news data highlight the need for further research to develop cost-effective strategies and optimize the integration of external data sources. The practical implementation and scalability of the API-enhanced ChatGPT structure in real-world applications will depend on finding the right balance between performance and cost, which may require further experimentation and refinement.

Despite these limitations, the findings of this study have important implications for the field of stock market analysis and investment decision-making. The API-enhanced ChatGPT structure demonstrates the potential of leveraging conversational AI and LLMs to automate and enhance various aspects of the investment process, such as market research, risk assessment, and portfolio optimization. By providing investors and analysts with data-driven insights and recommendations based on a comprehensive analysis of stock price data and news articles, this approach can potentially support more informed and efficient decision-making.

However, it is crucial to emphasize that the API-enhanced ChatGPT structure should be used as a complementary tool to support investment decisions, rather than a sole basis for making investment choices. The model's predictions and recommendations should be carefully considered in conjunction with human judgment, domain expertise, and risk tolerance to ensure well-informed and prudent investment decisions.

In conclusion, this study serves as a pioneering exploration of enhancing ChatGPT's capabilities through the integration of external data sources and prompt engineering techniques. The findings demonstrate the potential of the API-enhanced ChatGPT structure to improve predictive performance and risk-adjusted returns in stock market analysis. However, further research is necessary to address the limitations, optimize cost-effectiveness, and assess the model's robustness and generalizability across different market conditions and time horizons. As research in this area continues to evolve, the integration of conversational AI and LLMs with domain-specific tools and data sources holds promise for revolutionizing the field of stock market analysis and supporting more informed and efficient investment decision-making.

9 Conclusion

This study aimed to explore the potential of enhancing ChatGPT's capabilities in stock market analysis by developing an API-enhanced structure that integrates stock price and news data. The main findings demonstrate that the incorporation of external data sources and prompt engineering techniques can significantly improve ChatGPT's performance in predicting stock price movements and generating valuable insights for investment decision-making.

The evaluation results showed that the API-enhanced ChatGPT structure achieved a 10% improvement in accuracy rates and F1 scores compared to using only stock price data. Furthermore, the integration of news data led to a 20% improvement in risk-adjusted returns, as measured by the Sharpe ratio, and a 20% improvement in the model's ability to generate excess returns relative to the S&P 500 benchmark, as indicated by the information ratio.

These findings have important implications for the field of stock market analysis and investment decision-making. The API-enhanced ChatGPT structure demonstrates the potential of leveraging conversational AI and LLMs to automate and enhance various aspects of the investment process, providing investors and analysts with data-driven insights and recommendations. By integrating relevant data sources and employing prompt engineering techniques, this approach can support more informed and efficient decision-making in the complex and dynamic world of stock market investing.

However, it is important to acknowledge the limitations of this study, such as the specific dataset and evaluation period used, as well as the significant cost implications associated with incorporating news data. Future research should focus on addressing these limitations by exploring cost-effective strategies, assessing the model's robustness and generalizability across different market conditions and time horizons, and investigating alternative approaches to feature engineering and model architecture.

Moreover, future studies could explore the integration of additional data sources, such as social media sentiment, macroeconomic indicators, or company fundamentals, to further enhance the predictive capabilities of the API-enhanced ChatGPT structure. Investigating the potential of fine-tuning the ChatGPT model on domain-specific datasets or incorporating transfer learning techniques could also be promising avenues for future research.

The findings of this study also have broader implications for the application of conversational AI and LLMs in other domains. The API-enhanced structure developed in this study serves as a proof-of-concept, demonstrating how the integration of external data sources and prompt engineering techniques can significantly improve the performance of LLMs in specific domain applications. This approach could be adapted and applied to other fields, such as healthcare, legal services, or customer support, where access to relevant data and domain-specific knowledge is crucial for effective decision-making and problem-solving.

In conclusion, this study takes a pioneering step in exploring the enhancement of ChatGPT's capabilities through the integration of external data sources and prompt engineering techniques for stock market analysis. The findings demonstrate the potential of the API-enhanced ChatGPT structure to improve predictive performance and risk-adjusted returns, supporting more informed and efficient investment decision-making. However, further research is necessary to address the limitations, optimize cost-effectiveness, and explore alternative approaches to unleash the full potential of conversational AI and LLMs in the field of stock market analysis and beyond.

Data and Code Availability Statement

The Python code and Jupyter Notebook used in this research for data collection, prompt generation, model output retrieval, evaluation, and visualization are publicly available on GitHub. Interested readers, researchers, and practitioners can access the code repository here

The repository contains detailed instructions on how to set up the environment, install the necessary dependencies, and run the code. The Jupyter Notebook provides a step-by-step walkthrough of the data collection process, including the integration of stock price and news data through APIs, prompt engineering techniques, and the evaluation of the API-enhanced ChatGPT structure's performance.

By making our code and resources publicly available, we aim to promote transparency, reproducibility, and collaboration within the research community. We encourage readers to explore, modify, and build upon our work to further advance the application of conversational AI and large language models in the domain of stock market analysis and investment decision-making.

Please note that the code and resources are provided under the Apache-2.0 license, and we kindly request that users cite our work when utilizing or building upon the provided code and resources.

For any questions, feedback, or collaborations related to the code and resources, please feel free to contact the corresponding author or open an issue on the GitHub repository.

References

Alnefaie, A., Singh, S., Kocaballi, B., & Prasad, M. (2021). An overview of conversational agent: Applications, challenges and future directions. *17th International Conference on Web Information Systems and Technologies*.

Day, M.-Y., Lin, J.-T., & Chen, Y.-C. (2018). Artificial intelligence for conversational robo-advisor. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1057–1064.

Ghimire, A., Thapa, S., Jha, A. K., Adhikari, S., & Kumar, A. (2020). Accelerating business growth with big data and artificial intelligence. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 441–448.

Mokhtari, S., Yen, K. K., & Liu, J. (2021). Effectiveness of artificial intelligence in stock market prediction based on machine learning. *arXiv preprint arXiv:2107.01031*.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, *27*(2), 1–19.

Swamy, M., Shukla, A., & Purtilo, J. (2023). Llm-based stock market trend prediction.

Tong, H., Li, J., Wu, N., Gong, M., Zhang, D., & Zhang, Q. (2024). Ploutos: Towards interpretable stock movement prediction with financial large language model. *arXiv preprint arXiv:2403.00782*.

Wu, R. (2024). Portfolio performance based on llm news scores and related economical analysis. *Available at SSRN 4709617*.

Zhao, H., Liu, Z., Wu, Z., Li, Y., Yang, T., Shu, P., Xu, S., Dai, H., Zhao, L., Mai, G., et al. (2024). Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*.