Forecasting Performance of Quantitative Strategies with OpenAI GPT-4

Ege Doğan Dursun[1*], Mansur Alp Toçoğlu[1], Emre Şatır[1]
[1] Department of Computer Engineering, Izmir Katip Celebi University, Turkey
ORCID: -, 0000-0003-1784-9003, 0000-0002-1950-5549
Emails: esa.ege@gmail.com, mansuralp.tocoglu@ikcu.edu.tr, emre.satir@ikcu.edu.tr
*Corresponding author.

***Abstract*— The advent of advanced language processing models like OpenAI's GPT-4 presents new opportunities for enhancing financial decision-making. This study aims to explore the potential of GPT-4 in forecasting the performance of quantitative trading strategies, with a focus on the application of specific indicators in a long-short portfolio over a time frame. To achieve this, we employ a novel approach that involves posing targeted questions to GPT regarding the effectiveness of the indicators. The responses of the model are then subjected to a comparison with backtesting results obtained from the corresponding timeframe, enabling an evaluation of the predictive accuracy of GPT-4. By leveraging the linguistic capabilities of GPT, we aim to extract predictions that can inform the optimization of strategies. The benchmarking results obtained from this comparison serve as the primary output of the study, offering an objective assessment of the model's performance in forecasting the behavior of markets.**

***Keywords*— Quantitative Trading Strategies, Forecasting Performance, GPT-4, Natural Language Processing, Financial Decision-Making, Large Language Models.**

I. INTRODUCTION

The landscape of financial markets has undergone significant transformations with the advent of technological advancements, particularly in the field of information technology (IT); such as artificial intelligence (AI), machine learning (ML), and deep learning. Among the various AI models, natural language processing (NLP) models, such as OpenAI's Generative Pre-trained Transformer 4 (GPT-4) [1], have emerged as powerful tools capable of understanding and generating human-like text, which opening up new possibilities for their application in financial analysis and decision-making. Moreover, with the recent advancements, these models are empowered with multi-modality capabilities, which helps the AI agents to run software functions, directly reach API tools, browse the web to learn about recent updates for a given topic, and even read and understand PDF files and word processor documents.

Quantitative trading strategies, which rely on mathematical and statistical models to identify trading opportunities, have been an important topic in modern finance. These strategies often utilize multiple indicators and metrics to predict market movements and make informed trading decisions. Yet, the dynamic nature of the financial markets, coupled with the vast amount of data generated, poses a challenge to traditional quantitative models in accurately predicting the volatile nature of the financial environment, which can be significantly impacted by political, legal, and social events. This is where GPT-4's multi-modality capabilities can be leveraged to enhance the forecasting the performance of these strategies.

GPT-4, with its advanced language understanding and generation abilities, can potentially provide insights into market trends and investor sentiment by analyzing multiple sources of unstructured financial data, such as news articles, earnings reports, and social media posts. This study aims to explore the feasibility of using GPT-4, which is further empowered by Retrieval-Augmented Generation (RAG) frameworks [2] to forecast the performance of quantitative trading strategies, with a particular focus on the application of specific indicators in a long-short portfolio over a designated time frame. To achieve this, we employ a novel approach that involves posing targeted questions to GPT-4 regarding the effectiveness of multiple financial indicators. The model's responses are then compared with backtesting results [3] obtained from the corresponding timeframe by using pure mathematical calculations, thereby enabling an evaluation of the predictive accuracy of GPT-4's market insights. By leveraging the linguistic capabilities of GPT-4, the aim is to extract predictions that can inform the optimization of trading strategies and create leverage in the financial industry.

The significance of this research lies in its potential to bridge the gap between traditional quantitative models and the evolving field of AI in finance. By integrating GPT-4's predictive capabilities into the complex decision-making processes of high-frequency trading (HFT) environments, traders and financial analysts can enhance their understanding of market dynamics and improve the robustness and reliability of their trading strategies. This study also aims to contribute to the broader discourse on the integration of AI in financial analysis and decision-making, providing insights into how language models can be effectively utilized in the context of quantitative finance.

The objective results obtained as a result of this study can serve as a benchmark for future academic research and industrial applications, possibly offering a quantitative assessment of GPT-4's performance in forecasting the behavior of financial markets. These results have implications for the development of AI-driven trading strategies and the future of financial decision-making, as they highlight the potential synergies between language models and quantitative finance. By examining the predictive accuracy of GPT-4's insights in comparison with backtesting results, the aim is to provide a more comprehensive understanding of the state-of-the-art LLM's current condition and potential in forecasting the performance of financial indicators and indicator-based financial trading strategies.

This paper is organized as follows: After the introduction, Section 2 discusses related studies and methods in literature. We describe the data set and propose our approach in Section 3. Section 4 presents experimental results. We conclude the paper in Section 5 with the last words about our work.

II. LITERATURE REVIEW

The integration of large language models (LLMs) such as the Generative Pre-trained Transformers (GPT) series into financial technology has marked a significant shift in the landscape of quantitative investment and trading strategies. Recent research has shed light on the multifaceted applications of LLMs in this domain, ranging from the innovative mining of trading signals to the enhancement of decision-making processes in multi-agent trading systems. These studies highlight the potential of LLMs not only in interpreting and translating quantitative researchers' ideas into actionable strategies but also in addressing inherent biases and improving the robustness of trading decisions through advanced memory structures and model architectures.

Wang et al. (2023) introduce a novel paradigm in the field of quantitative investment, focusing on the mining of new alphas (effective trading signals or factors) through the integration of human-AI interaction and LLMs. This approach aims to address the limitations of traditional alpha mining methods, which either rely on hand-crafted factor synthesizing or algorithmic factor mining, by leveraging the capabilities of LLMs to interpret and translate the ideas of quantitative researchers into actionable alphas. The Alpha-GPT system, as proposed by the authors, aims to serve as an interactive platform that enables a heuristic understanding of quantitative researchers' ideas and outputs creative, insightful, and effective alphas. The system has three main components: (I) the Dialog Box for user input and system feedback, (II) the Mining Session Manager for organizing user-generated data, and (III) the Alpha Mining Dashboard for detailed analysis and visualization of alpha mining results [4].

One of the key contributions of this study to the field is the introduction of a new alpha mining paradigm that enhances human-AI interaction to improve the effectiveness and efficiency of alpha research. The authors propose an algorithmic framework, AlphaBot, which employs an LLM as a mediator for human-AI interaction. This framework includes a knowledge compiler that converts researchers' intents into domain-specific prompts for the LLM, a thought decompiler that translates the LLM's natural language output into actionable configurations for alpha mining, and a knowledge library that incorporates additional domain-specific knowledge to enhance the LLM's performance. The experimental results demonstrate the effectiveness of Alpha-GPT in generating alphas consistent with trading ideas and enhancing seed alphas through search algorithms. The system also allows for efficient human-AI interaction, enabling users to guide the process and receive explanations for the generated alphas in natural language. This approach to integrating human-AI interaction and LLMs in alpha mining aligns with the aims and goals of our research in forecasting the performance of quantitative trading strategies using GPT-4. By leveraging the linguistic capabilities of GPT models, we aim to extract predictions that can inform the construction and optimization of trading strategies, similar to how Alpha-GPT interprets and translates quantitative researchers' ideas into effective alphas. The interactive nature of Alpha-GPT and its ability to provide natural language explanations for financial signals resonate with our goal of making quantitative strategy forecasting more accessible and interpretable through the use of advanced AI models.

Glasserman and Lin examine the challenges posed by look-ahead bias and the distraction effect in backtesting trading strategies based on sentiment analysis performed by LLMs such as GPT-3.5. Look-ahead bias refers to the LLM's potential knowledge of future events within its training data, which can artificially inflate backtesting performance [5]. The distraction effect, on the other hand, pertains to the LLM's general knowledge about companies, which may interfere with the accurate assessment of a text's sentiment. To mitigate these biases, the authors propose an anonymization procedure that removes company identifiers using Named Entity Recognition (NER) [6] from the news text. This approach allows for a more accurate evaluation of the LLM's sentiment analysis capabilities. The study finds that, surprisingly, anonymized headlines outperform the original ones in-sample, suggesting that the distraction effect has a more significant impact than look-ahead bias. This effect is particularly pronounced for larger companies, which are likely to have more information available in the model's training data. Out-of-sample, where look-ahead bias is not a concern, the study still finds evidence that anonymization can be beneficial, although the results are less statistically significant. The authors conclude that anonymizing news text can be a useful tool not only for de-biasing backtesting but also for potentially improving out-of-sample performance. This research is relevant to our study as it highlights the importance of addressing biases in LLM-based sentiment analysis for financial applications. The findings suggest that careful consideration of training data and potential biases is crucial for developing accurate and reliable quantitative trading strategies using natural language processing models. By incorporating these insights into our methodology, we can enhance the robustness and validity of our own research on forecasting the performance of quantitative trading strategies using GPT-4.

Li et al. (2023) present a novel approach to financial trading using a multi-agent system powered by LLMs, specifically the GPTs. The authors propose a framework that incorporates a layered memory structure, akin to human cognitive processes, to enhance the decision-making capabilities of trading agents [7]. This structure includes long-term, medium-term, and short-term memory layers, each with its own decay mechanism to prioritize relevant, up-to-date information for trading decisions. The framework also introduces distinct characters for each trading agent, allowing for a diversity of trading strategies and risk preferences. This diversity is expected to improve the robustness of decision-making and uncover latent market opportunities. The agents can communicate and debate with each other, leveraging collective wisdom to optimize trading outcomes. The system also integrates real-time multi-modal data from various sources, providing a comprehensive view of the financial landscape. This enables the agents to react promptly to market changes and execute high-frequency trading in appropriate conditions. The authors suggest that this LLM-based multi-agent system, with its human-like memory processing and character diversity, could significantly outperform existing automated trading strategies. The research aligns with our study on forecasting the performance of quantitative trading strategies using GPT-4. By understanding the mechanisms proposed in this study, we can explore the potential of layered memory structures and diverse agent characters in enhancing the predictive accuracy of trading strategies. Additionally, the integration of multi-modal data sources and real-time market insights could be crucial in developing more sophisticated and adaptive trading models using LLMs.

Some other different studies on LLMs in large language models (LLMs) have demonstrated significant potential in financial applications, too. For example, Romanko et al. explored the use of ChatGPT for investment portfolio selection [8], showcasing its ability to optimize allocation decisions based on diverse financial metrics. Wu evaluated the trading performance of GPT-4, highlighting both its predictive strengths and limitations in capturing market dynamics [9]. In other work This paper aims to project the holistic view of the intersection of AI chatbots ChatGPT, Gemini, and Copilot with pedagogical frameworks, especially their potential for generating educational questions within the context of Turkish [10]. Zhang et al. proposed BreakGPT, a multi-stage structured model specifically designed for financial breakout detection, emphasizing the importance of advanced architectures in enhancing predictive accuracy in volatile financial environments [11]. Additionally, the English-Turkish translation performances of Google Translate, Amazon Translate and OpenAI GPT translation services were analyzed [12].

III. MATERIALS AND METHODS

The methodology of our study is based on the integration of LLMs, specifically OpenAI's GPT-4, into the domain of quantitative finance to forecast the performance of trading strategies, and further empowering it with the RAG technology to enable it to have access to multiple different sources, which is also named as "multi-modality" in the artificial intelligence domain [13]. Our methodology involves a multifaceted approach.

A. Data Collection and Preprocessing

First, historical financial data, including cryptocurrency prices, their respective trading volumes, and several standard fundamental indicators were imported from reliable data sources such as CoinGecko [14] and open-source data libraries. The data were cleaned, normalized, and structured to be suitable for analysis where necessary to avoid potential biases in the processing and interpretation. We also allowed the RAG agent to have access to recent news articles and financial reports relevant to the stocks and sectors of interest, by using the Polygon.io [15] API to incorporate sentiment analysis into our models.

B. Multi-Modality and RAG

GPT-4 is already pre-trained on a corpus of financial literature, including research papers, market analysis reports, and historical trading data. However, the multi-modality step is still crucial to adapt the model to understand and generate finance-specific content. The RAG integration process involves using multiple software frameworks to divide the agents based on their specific tasks and allow them to cooperate to accomplish certain tasks such as retrieving data, applying mathematical calculations, running code in a sandbox environment, scraping the web to get information, approach HTTP request usage for API retrieval, and ensuring that each agent is properly capturing the nuances of financial markets and trading strategies.

C. Indicator Prediction and Strategy Formulation

This study uses the GPT-4 model to predict the relevance and potential impact of various technical and fundamental indicators on the performance of financial instruments. This includes indicators like moving averages, relative strength index (RSI), price-to-earnings ratio (P/E), and more sophisticated strategies such as factor-based alphas [16, 17]. Based on the model's predictions about certain quantitative trading strategies, for long-short portfolios, we then proceed to calculate the real-world performance of the indicators using benchmarking with backtesting.

D. Backtesting and Performance Evaluation

The formulated strategies are backtested on historical data to evaluate their performance in the real world. The backtesting process involves simulating the execution of the strategy on past data to assess its profitability and risk characteristics. This quantitative strategy and its performance are then compared with the predictive results of the RAG agent powered by GPT-4 as a natural language processor. Performance metrics such as Sharpe Ratio [18, 19], Max Drawdown [20], and annualized returns are calculated to gauge the effectiveness of the strategies.

IV. RESULTS AND DISCUSSION

In our study, we explore the application of OpenAI's GPT-4, enhanced with RAG technology, in forecasting the performance of quantitative trading strategies for long-short portfolios. The methodology involves posing targeted questions to GPT-4 regarding the effectiveness of various financial indicators and comparing the model's predictions with backtesting results. This section provides the results of the experiments. Table 1 provides a summary of the performance metrics and predicted values from the experiments for Bitcoin, Ethereum and Ripple.

Table 1 – The comparison of the performance metrics and the forecasted values for Bitcoin, Ethereum and Ripple

Metric	Bitcoin		Ethereum		Ripple	
	Actual	Forecasted	Actual	Forecasted	Actual	Forecasted
Sharpe Ratio	2.12	2.63	1.36	0.26	0.62	0.14
Max Drawdown (%)	20.06	22.65	23.27	44.58	42.19	32.62
Average Daily Return (%)	0.27	0.33	0.18	0.031	0.16	0.018

A. *Bitcoin*

The metrics provided offer a comprehensive comparison between the actual performance of Bitcoin and its forecasted performance according to GPT-4's model, within the identical time frame. Starting with the Sharpe Ratio, which is a measure of the return earned above the risk-free rate per unit of volatility or total risk. We observe that the actual Sharpe Ratio is 2.12, indicating that Bitcoin has provided investors with a robust risk-adjusted return over the observed period. In contrast, the forecasted Sharpe Ratio is even higher at 2.63. This suggests that GPT-4's model anticipated an even more favorable risk-adjusted return scenario for Bitcoin than what was actually realized. The forecasted improvement in the Sharpe Ratio points to a prediction of either higher returns for the same level of risk or a similar level of returns for a reduced risk, both of which would be attractive to investors.

Moving to the Max Drawdown, which assesses the maximum observed loss from a peak to a trough of a portfolio, before a new peak is attained. We find that the actual Max Drawdown for Bitcoin stands at 20.06%, while the forecasted Max Drawdown is slightly higher at 22.65%. The forecast indicates a slightly more pessimistic view of Bitcoin's potential price dip within the same period, suggesting that GPT-4's model braced for a more substantial downturn than what actually occurred.

Lastly, the Average Daily Return for Bitcoin shows an actual value of 0.27% compared to a forecasted average daily return of 0.33%. The higher forecasted daily return indicates that GPT-4's model expected Bitcoin to perform slightly better on a daily basis than it actually did. This optimistic forecast underscores the model's potential to identify positive trends, although it slightly overestimated the actual performance. In synthesizing these insights, it's evident that GPT-4's model held a somewhat more optimistic view of Bitcoin's market performance over the specified period. While the actual metrics depict a solid and attractive investment proposition, the forecasted metrics suggest that the model anticipated an even stronger performance.

B. Ethereum

The Actual Sharpe Ratio for Ethereum stands at 1.36, illustrating a commendable risk-adjusted return on investment. This ratio is indicative of Ethereum's ability to offer investors meaningful returns in consideration of the risks involved. It reflects a balanced scenario where the volatility inherent in Ethereum's market is sufficiently compensated by its performance, highlighting its appeal to a broad spectrum of investors who weigh risk and reward carefully. In stark contrast, the Forecasted Sharpe Ratio plunges to 0.26, suggesting a far less optimistic prediction of Ethereum's risk-adjusted returns by GPT-4's model. This drastic reduction signals a significant underestimation of Ethereum's market resilience and its potential to provide favorable returns against market volatilities. The model's conservative outlook points towards a perceived increase in risk without a commensurate increase in expected returns, suggesting areas where the model's predictive capabilities could be enhanced to more accurately reflect market realities.

The divergence becomes more pronounced when examining the Max Drawdown metrics. The Actual Max Drawdown recorded at 23.27% underscores the volatility that Ethereum investors navigated during this period. Although this figure denotes a considerable risk, it aligns with the expected market behavior for cryptocurrencies, which are known for their rapid price fluctuations. The Forecasted Max Drawdown forecasted by the model at a significantly higher 44.58%, reveals an expectation of a much more turbulent market downturn than actually occurred. This overestimation of potential loss heightens the perceived investment risk, potentially deterring investment based on the forecast alone. This discrepancy underscores the need for refining the model's approach to volatility and risk assessment to more closely mirror the complexities of the cryptocurrency market.

Lastly, the comparison of Average Daily Returns further elucidates the contrast between reality and prediction. The actual returns at 0.18% per day depict Ethereum as a profitable venture on average, albeit with the usual risks. The forecasted average daily return dwindling to 0.031% significantly downplays Ethereum's earning potential as envisaged by GPT-4's model, again pointing to the conservative bias in the model's market outlook.

C. Ripple

Sharpe Ratio Analysis reveals a considerable gap between actual and forecasted values. The Actual Sharpe Ratio of 0.62, though not as high as those seen with some other assets, still indicates a level of return on Ripple that compensates for the risk incurred. This ratio suggests that, despite the volatility inherent to cryptocurrencies, Ripple offered a moderately favorable risk-adjusted performance. The Forecasted Sharpe Ratio dramatically drops to 0.14. This sharp decline in the forecasted ratio underscores a pessimistic prediction, indicating that GPT-4's model anticipated significantly lower returns for the level of risk associated with Ripple. This suggests that the model might have overly focused on the risks or not fully appreciated the potential returns, highlighting an area for potential recalibration to better align with market behaviors.

Max Drawdown metrics further emphasize the differences in actual versus forecasted performance. The Actual Max Drawdown at 42.19% showcases the high level of risk and volatility that was present, indicating that Ripple's value saw a significant decline at its lowest point. This level of drawdown is substantial, reflecting the high-stakes environment of cryptocurrency investment. The Forecasted Max Drawdown is somewhat lower at 32.62%, suggesting that GPT-4's model expected a less severe worst-case scenario for Ripple than what actually occurred. This forecasted reduction in max drawdown could indicate a model expectation of a somewhat more stable market for Ripple, or it may reflect limitations in the model's ability to capture the full extent of market volatility.

Average Daily Return metrics offer a different comparison as well. The actual returns at 0.16% indicate that, on average, Ripple provided a modest but positive yield on a daily basis. In stark contrast, the forecasted daily returns plummet to just 0.018%, a significant downgrade that suggests the model predicted Ripple to be far less profitable than it was. This drastic reduction in predicted returns could reflect a conservative bias within the model or a potential underestimation of Ripple's market appeal and performance.

V. CONCLUSIONS, LIMITATIONS AND FUTURE DIRECTIONS

Bitcoin's Forecast Analysis highlights a notable optimism in GPT-4's predictions compared to actual market performance. While the actual Sharpe Ratio, Max Drawdown, and Average Daily Return for Bitcoin indicate a robust risk-adjusted return, GPT-4's forecasts suggest an even more favorable outlook. This optimistic bias in the model's predictions, particularly evident in the higher forecasted Sharpe Ratio and Average Daily Return, points towards GPT-4's potential to recognize positive market trends. The discrepancy, especially in the context of the Max Drawdown, suggests a need for recalibration to more accurately mirror market volatilities and risk factors.

Ethereum's Predictive Performance draws attention to GPT-4's predictive precision, which, while capturing the general market trends, occasionally missed rapid market shifts. This variance underscores the model's capacity to detect underlying market dynamics but also highlights the challenge of anticipating quick shifts in market sentiment. The divergence between actual and forecasted metrics—particularly in the Sharpe Ratio and Average Daily Return—reinforces the necessity for integrating broader datasets and adopting more sophisticated forecasting methodologies.

Ripple's Price Predictions elucidate the model's overarching tendency to either overestimate or underestimate market movements. The comparative analysis of Ripple's actual versus forecasted performance metrics further illuminates the model's conservative forecasting approach, as evidenced by a lower-than-actual forecasted Sharpe Ratio and Average Daily Return.

In synthesizing these findings, it's evident that GPT-4, with RAG technology, holds substantial promise as a tool for financial forecasting, particularly in the volatile realm of cryptocurrencies. The model's ability to closely track market trends in some instances highlights its potential utility. Nonetheless, the observed deviations and optimistic biases across Bitcoin, Ethereum, and Ripple forecasts point to the imperative for ongoing model enhancements. By refining the model's responsiveness to market sentiment shifts, broadening the scope of data inputs, and employing advanced machine learning techniques, the predictive accuracy and reliability of GPT-4 can be significantly improved.

The exploration of GPT-4's capabilities in forecasting the performance of cryptocurrencies has revealed insightful yet nuanced challenges inherent in this endeavor. A primary constraint is the reliance on the quality and scope of the input data. Cryptocurrency markets are highly volatile, influenced by a wide array of factors including regulatory changes, market sentiment, and global economic conditions. The model's predictions, therefore, hinge on the comprehensiveness and accuracy of the data fed into it. If the dataset lacks critical information or is skewed, predictions may not fully capture market realities.

Another limitation lies in the model's current understanding and integration of real-time market dynamics. The cryptocurrency market's fast-paced nature means that information becomes outdated rapidly. GPT-4's ability to incorporate the latest market developments into its forecasting process is crucial for accuracy. However, delays in data integration or the inability to process information in real-time can lead to discrepancies between predicted and actual market movements. The inherent unpredictability of cryptocurrency markets poses a significant challenge. Factors such as unexpected geopolitical events or sudden shifts in investor sentiment can drastically affect market behavior. While GPT-4 is designed to identify patterns and make informed predictions, the unpredictable nature of these markets can sometimes render even the most sophisticated models less effective. The complexity of modeling financial indicators and their impact on cryptocurrency prices also cannot be overlooked. Predictive models must navigate not just historical price data but also the intricate web of factors that influence market movements. This requires a deep and nuanced understanding of both financial theory and market psychology, areas where even advanced AI models like GPT-4 may encounter limitations.

While the application of GPT-4 in financial forecasting shows the potential of AI in understanding and predicting market trends, it also underscores the need for continuous refinement of these models. Enhancements in data quality, real-time processing capabilities, and a deeper understanding of market drivers are essential for overcoming the current constraints and limitations, thereby paving the way for more accurate and reliable predictive tools in the realm of cryptocurrency trading.

A critical area for future exploration is the enhancement of real-time data processing. Given the rapid pace at which the cryptocurrency markets move, developing methods that allow GPT-4 to ingest and analyze data in real-time could significantly improve prediction accuracy. This entails not just faster data processing but also the integration of live feeds from social media, news outlets, and market transactions, which could offer a more dynamic view of market sentiments and trends. Another promising direction involves the incorporation of more sophisticated machine learning techniques, such as deep learning and reinforcement learning. These methodologies could provide GPT-4 with a deeper, more nuanced understanding of complex market dynamics, enabling it to make more informed predictions. Specifically, reinforcement learning could allow the model to learn and adapt from its forecasting outcomes, refining its predictions over time based on feedback from actual market movements. Exploring the impact of global economic indicators and geopolitical events on cryptocurrency markets could also yield valuable insights. By understanding how these external factors influence market behaviors, researchers could enhance GPT-4's predictive model to account for such variables, potentially offering a more holistic view of the market's drivers.

Further research could also investigate the application of GPT-4's predictive capabilities across different cryptocurrencies and trading strategies. Each cryptocurrency operates within its own ecosystem and may be influenced by unique factors. Similarly, the effectiveness of various trading strategies may vary across market conditions. A more granified analysis could uncover specific patterns or indicators that are particularly predictive for certain assets or under certain conditions, contributing to more targeted and effective trading strategies. The development of a more interactive model, where traders and financial analysts can pose specific questions or scenarios to GPT-4, could revolutionize the way market predictions are utilized. This could transform GPT-4 from a forecasting tool into an advisory one, providing customized insights and recommendations based on the latest market data and trends.

REFERENCES

[1] OpenAI. (2023). GPT-4 technical report (Technical report). https://cdn.openai.com/papers/gpt-4.pdf

[2] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information *Processing Systems*, *33*, 9459-9474. https://arxiv.org/abs/2005.11401

[3] Pardo, R. (2011). Evaluation and optimization of trading strategies (2nd ed.). Wiley.

[4] Wang, S., Yuan, H., Zhou, L., Ni, L. M., Shum, H. Y., & Guo, J. (2023). Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment. arXiv. https://arxiv.org/abs/2308.00016

[5] Glasserman, P., & Lin, C. (2023). Assessing Look-Ahead Bias in Stock Return Predictions Generated By GPT Sentiment Analysis. arXiv. https://arxiv.org/abs/2309.17322

[6] Jurafsky, D., & Martin, J. H. (2009). Speech and language processing (2nd ed.). Pearson.

[7] Li, Y., Yu, Y., Li, H., Chen, Z., & Khashanah, K. (2023). TradingGPT: Multi-Agent System with Layered Memory and Distinct Characters for Enhanced Financial Trading Performance. arXiv. https://arxiv.org/abs/2309.03736

[8] Romanko, O., Narayan, A., & Kwon, R. H. (2023). ChatGPT-based investment portfolio selection. SS&C Algorithmics, University of Toronto, Indian Institute of Technology Bombay.

[9] Wu, B. (2023). Is GPT4 a Good Trader? Preprint, arXiv:2309.10982. https://arxiv.org/abs/2309.10982

[10] Atayolu, Y., Kutlu, Y. (2024). Similarity and Classification Analysis in Turkish Question Generation of AI Tools According to Bloom's Taxonomy. Tethys

Environmental Science, 1(2), 87-98, doi : 10.5281/zenodo.13138659

[11] Zhang, K., Yoshie, O., & Huang, W. (2024). BreakGPT: A large language model with multi-stage structure for financial breakout detection. *Proceedings of the 2024 ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '24)*. ACM, New York, NY, USA.

[12] Öncü, M., Kutlu, Y. (2024). Performance Comparison of Known AI Translation Tools for Turkish Language. Tethys Environmental Science, 1(3), 117-126, doi : 10.5281/zenodo.13269233.

[13] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. In Proceedings of the 28th International Conference on Machine Learning (pp. 689–696).

[14] CoinGecko (2024). CoinGecko API Documentation. https://www.coingecko.com/tr/api/documentation

[15] Polygon.io (2024). Polygon.io API Documentation. https://polygon.io/docs

[16] Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. New York Institute of Finance.

[17] Damodaran, A. (2012). Investment Valuation: Tools and Techniques for Determining the Value of Any Asset (3rd ed.). Wiley.

[18] Sharpe, W. F. (1966). Mutual Fund Performance. The Journal of Business, 39(1), 119–138.

[19] Bodie, Z., Kane, A., & Marcus, A. J. (2018). Investments (11th ed.). McGraw-Hill.

[20] Lhabitant, F.-S. (2004). Hedge Funds: Quantitative Insights. John Wiley & Sons.