LLM and Deep Learning Aided Stock Portfolio Optimization and Management using RAG and LangChain

Umar Farooq Khan (22179780)
Artificial Intelligence Driven Decision Making – H9AIDM
MSCAI1
School of Computing
National College of Ireland

May 15, 2024

1 Background Research

The advent of Large Language Models (LLMs) Touvron et al. (2023), Wu et al. (2023) Ouyang et al. (2022) Yang et al. (2023) signifies a significant advancement in Generative AI, offering potential productivity enhancements across various domains. However, their widespread adoption faces hurdles, particularly in specialized sectors, due to issues such as hallucinations caused by a lack of domain-specific knowledge. To mitigate this, methods to imbue LLMs with specialized expertise are imperative.

In crafting a large language model, the initial step involves pre-training a transformer on a vast corpus of text data, typically non-domain-specific. While this equips LLMs with proficiency in general inquiries, it hinders their ability to address domain-specific queries effectively. Additionally, the model's efficacy is influenced by the frequency of relevant information in the training data, posing challenges for domains like finance where pertinent data is less..

To address these challenges, knowledge injection techniques are employed, allowing Large Language Models (LLMs) to access domain-specific information beyond their initial training data. One effective method is through additional training or fine-tuning using in-context learning methodologies like Retrieval Augmented Generation (RAG), which seamlessly integrates pre-existing knowledge to enhance LLM performance in specialized tasks such as financial analysis. The extensive complexity and nuances in finance, including specialized jargon and required domain expertise, make LLMs a valuable tool for handling such data, as highlighted in Yu et al. (2023).

The study explores AI-driven techniques, including Retrieval Augmented Generation (RAG), LangChain, and Deep Learning, to improve decision-making in stock portfolio optimization. These methods utilize advanced AI methodologies to efficiently process and analyze financial data, aiming to enhance stock price direction prediction and portfolio management.

Previous studies highlights the evolution of auto-regressive language models (LLMs) from GPT-3 to subsequent iterations, emphasizing their versatility across diverse tasks Brown et al. (2020). While early research primarily focused on general-purpose LLMs, recent studies underscore the importance of domain-specific models, particularly in fields like finance Xing et al. (2018). BloombergGPT Wu et al. (2023) emerges as a notable contribution, representing a hybrid approach with a 50 billion parameter model trained on a mixed dataset, showcasing superior performance in financial tasks while remaining competitive in general NLP benchmarks (Brown et al. (2020).

2 AI Driven Methodology

For this research, we made 4 models utilizing 2 different approaches of AI; first two ones are aiding Investors and quants with the help of LLM utilizing two new Retreival Approaches ie. RAG and LangChain Agents, and the second one is Deep Learning and ML. RAG with LLM incorporates information retrieval by selecting relevant passages based on cosine similarity between query and passage embeddings.

$$\text{cosine\_similarity}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \tag{1}$$

2.1 Retrieval augmented generation (RAG)

LLM has come to a stage where it achieved a lot in a very less timeframe but still if faces some challenges and limitations specially in knowledge intensive task, and domain specific and when it comes to numbers crunching and have outdated knowledge and always has a factor of hallucination.

To solve these issues, RAG comes into play that enchances the knowledge of LLMs by retreiving relevant documents chunks from any other external knowledge base through semantic similarity. The knowledge raw text has been converted into embeddings. These embeddings capture the semantic meaning of the text, enabling more nuanced analysis and understanding. In early day, RAG's emerged along the development of Transformers architecture, its architecture has been shown in Figure 1.

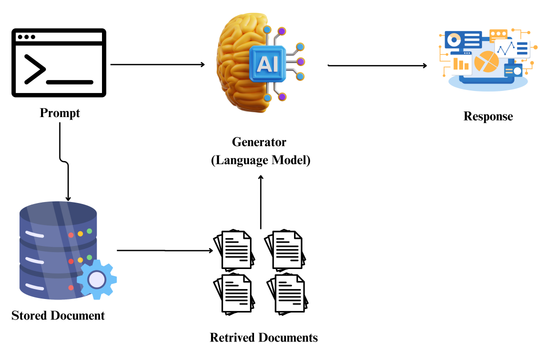


Figure 1: Architecture of RAG

2.2 LangChain utilizing Chain of Thoughts

Langchain is a open source framework used to develop applications that uses LLM in the background, and the purpose of it is to use resources and data sources and have the ability to interact with other applications such as Google search Engine. There are three main components of Langchain: Prompts,Models, Chains.

The most important component of LangChain which is the building block is its chain. It is coupled with LLM, a prompt, and with this chain and you can use multiple chains. A just just takes input and generates an output. Mutiple chains can run in sequential manner. Where the output of the first chain becomes the input of the second chain. You can chain mutiple inputs and takes one output as shown in Fig 2 Input from user is combined with the prompt that is already given and that is the description of the role and task of the chain. Architecture of our module 3 has been shown in Figure 2 on how tabular data which are merly numbers interacts with Agents.

In this module, we employed advanced tools like RAG and LangChain, to process and analyze information extracted from financial PDF documents. This methodology aims to facilitate the comparison and evaluation of different stocks based on their financial performance as outlined in their annual and quarterly reports.

The "Chain of Thought" encompasses a structured approach to processing, analyzing, and interpreting financial data from CSV files, leading to informed decision-making in the realm of stock market investment. This systematic approach ensures that numerical information is comprehensively understood and leveraged to derive actionable insights.

The insights obtained from the analysis and question-answering process are synthesized to draw conclusions about the stock's performance and potential investment opportunities. These insights can guide decision-making processes, such as whether to invest in the stock or not.

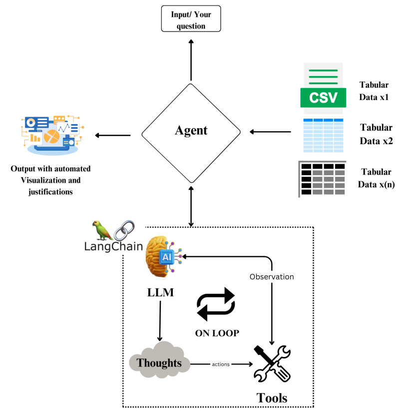


Figure 2: Architecture of Module 3

2.3 Deep Learning for the prediction of Stock Price Direction

Our third module, which used numerical tabular data in form of dataset which contains open, closing, high, low and volume of the stock. We applied feature engineering to introduce some more features as predicting closing price directly is straight out of discussion. After introducing some new features like tomorrows price, MACD, EMA12, EMA26, average gain and loss. We applied Machine learning as well as Deep Learning. In the end, we applied Explanable AI techniques for making our model easily explanable. Few Formulae of newly introduced decision variables are given below:

4. **Relative Strength (RS):**

$$\text{RS} = \frac{\text{Average Gain}}{\text{Average Loss}}$$

5. **Relative Strength Index (RSI):**

$$\text{RSI} = 100 - \frac{100}{1 + \text{RS}}$$

6. **12-day Exponential Moving Average (EMA):**

$$\text{EMA}_{12} = \text{Close}_t \times \alpha + \text{EMA}_{12_{t-1}} \times (1 - \alpha) \quad \text{where } \alpha = \frac{2}{13}$$

8. **Moving Average Convergence Divergence (MACD):**

$$\text{MACD} = \text{EMA}_{12} - \text{EMA}_{26}$$

9. **Signal Line (9-day EMA of MACD):**

$$\text{Signal Line} = \text{MACD} \times \alpha_9 + \text{Signal Line}_{t-1} \times (1 - \alpha_9) \quad \text{where } \alpha_9 = \frac{2}{27}$$

3 Implementation & Experiment

We used OpenAI GPT3.5 for this study, which is considered a benchmark and a baseline for results whenever we are using any LLM research, as Chatgpt is considered one of the most highly acclaimed LLM.

3.1 Datasets Used

A pleothra of around 30 different datasets of financial data has been fetched from Yahoo Finance platform including various types of data, including historical market data, actions (such as dividends and stock splits), share count, financial statements (income statement, balance sheet, cash flow statement), holders information, news, and dividends. Each type of financial data is stored in separate CSV files for easy access and analysis.

For the other module, quarterly and annual financial reports have been downloaded from the company official investors website and given to our system. For the 4th module in which we train our own model, we used just historical data of a stock which includes high, low, closing, volume etc.

3.2 Module 1 & 2: Annual and Quarterly Reports

Financial PDF documents of different stocks are initially gathered for comparison, containing detailed data on metrics like revenue, earnings per share, and operating expenses. Using the PyPDF2 library in Python, these PDF files are processed, extracting text to facilitate further analysis. To handle large volumes of text effectively, the extracted information is divided into smaller chunks to avoid overwhelming the LLM, given its context window limitation of 4096 tokens. This approach prevents token size inflation and ensures efficient processing of the data.

Next, embeddings are created using OpenAI's pre-trained language model. FAISS (Facebook AI Similarity Search) facilitates similarity search to locate pertinent documents in response to user queries. This process assists in pinpointing relevant sections within financial reports. For question answering tasks, the LangChain framework, integrating the RAG model, is utilized. It involves crafting queries regarding financial metrics and objectives and extracting accurate responses from the identified documents.

3.3 Module 3: Numeric Data using Langchain Agents

This module applies the Chain of Thought concept through the Langchain Python framework, utilizing its agents. It follows a systematic approach to process and analyze financial data from CSV files, aiming to derive meaningful insights and make informed decisions.

A large amount of financial data for a specific stock is gathered, in this case, "TEL," using the Yahoo Finance API (yfinance).

Then various analysis techniques can be applied to derive insights from the data. For example, historical market data can be used to analyze trends and patterns in stock prices, while financial statements can be analyzed to assess the company's financial health and performance. Using LangChain, a question-answering framework, queries can be formulated based on specific financial metrics or performance indicators. These queries are then processed and precise answers along with the justification in the forms of visualizations and graphs is shown to provide insights into the queried topics.

3.4 Module 4: DL on Stock Market Price Direction Prediction

The financial data is preprocessed and then we created several technical indicators that are mentioned previously. We are only predicting price direction which is more important. Then 40+ ML and DL models are applied on it to find out the best performing model which exceeds in all the metrics and all models are giving a variety of results showing good diversity which are also ensembled as an experiment.

4 Evaluation of AI Model

Students should present the results of the AI system experiments. Furthermore, students should evaluate and discuss the results appropriately.

The reliability of responses from LLMs like OpenAI ChatGPT 3.5 is assured by robust benchmarks such as HumanEval Chen et al. (2021) and BLEU, indicating exceptional reasoning capabilities. ChatGPT outperforms in many benchmarks and metrics when evualuated on variety of domains with other LLMs in the market Akter et al. (2023). Users trust the model's accuracy and insight across diverse topics, empowered by its remarkable ability to understand complexity, connect information, and generate coherent responses, enhancing reliability across applications.

However, the performance of LLMs are calculated on several benchmarks like HumanEval GLUE (General Language Understanding Evaluation), SuperGLUE (Super General Language Understanding Evaluation), SQuAD (Stanford Question Answering Dataset), RACE (Reading Comprehension from

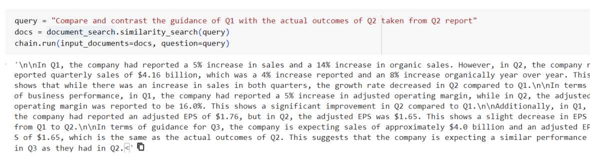
Examinations), WMT (Workshop on Machine Translation), CoQA (Conversational Question Answering) and lastly MMLU on which OpenAI GPT 3.5 models excels,

The study Akter et al. (2023) evaluates GPT-3.5 and GPT-4 in physics coding assignments, finding students averaged 91.9%, surpassing GPT-4's 81.1%. Prompt engineering significantly improved GPT-4 and GPT-3.5 , while authorship identification accuracy averaged 85.3% which shows there is no doubt on the reasoning capabilities of GPT3.5 and GPT4.

4.1 Module 1 & 2: Annual & Quarterly Reports

In every quarterly report, there is a term used called Guidance which refers to a company's projected performance for upcoming quarters. It's based on analysis by company management and analysts, including estimates like Earnings per Share (EPS), revenue, and expenses. Investors use this to make decisions about stocks, but it's not a guarantee. Unexpected factors can alter results, so investors consider guidance alongside other information for a clearer view of the company's future.

So we evaluate our model to by giving all quarterly financial reports of a fiscal year and try to compare, contrast and try to analyze what goals do they achieve, what KPIs do they successfully achieved what they claimed in the previous guidance. As a result, we or even by one more quick question from our model can quickly came to a conclusion, that how much reliable are their 'Guidance' in their reports, what is trend of goals they always achieved and what goals and numbers they couldnt achieve in every quarter, what went wrong etc type of things. So that, an investor can easily comprehend the quarterly reports and is well aware what is about to be happen to some common key metrics. We can see in the Figure 3 that a lot of KPIs and goals could not be achieved by TE Connectivity Company according to what they claim in their Q1 Quarterly Report and Figure 4 shows goals not achieved over the year by analyzing Annual Report.

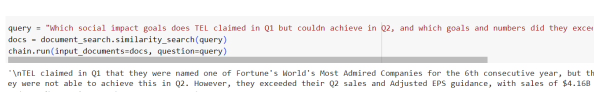


```
query = "Compare and contrast the guidance of Q1 with the actual outcomes of Q2 taken from Q2 report"
docs = document_search.similarity_search(query)
chain.run(input_documents=docs, question=query)
```

```
'\n\nIn Q1, the company had reported a 5% increase in sales and a 14% increase in organic sales. However, in Q2, the company r
eported quarterly sales of $4.16 billion, which was a 4% increase reported and an 8% increase organically year over year. This
shows that while there was an increase in sales in both quarters, the growth rate decreased in Q2 compared to Q1.\n\nIn terms
of business performance, in Q1, the company had reported a 5% increase in adjusted operating margin, while in Q2, the adjusted
operating margin was reported to be 16.0%. This shows a significant improvement in Q2 compared to Q1.\n\nAdditionally, in Q1,
the company had reported an adjusted EPS of $1.76, but in Q2, the adjusted EPS was $1.65. This shows a slight decrease in EPS
from Q1 to Q2.\n\nIn terms of guidance for Q3, the company is expecting sales of approximately $4.0 billion and an adjusted EP
S of $1.65, which is the same as the actual outcomes of Q2. This suggests that the company is expecting a similar performance
in Q3 as they had in Q2.'
```

Figure 3: Goals and KPIs not achieved in Q2

Second module is implemented giving user the ability to process and analyze Annual Reports of various stocks at a time. Due to the context limit, it can handle annual financial reports of 4-5 stocks. It works in the same way, just it has a different use case as compared to one, Module one has quarterly reports which is suitable for short term goals and visions of company's that are best for company and its investors, while Annual Reports shows a complete picture of their company throughout the year, which is more suitable for investors when they want to look forward to the long term vision and see if they align with the company's long term goals and performance. It is evaluated by comparing and contrasting previous annual reports with the current to understand how many times does they couldnt achieve their annual goals or throughout the year how much the company has improved or acheived their annual KPIs, because quarterly KPIS could be achieved but sustaining them is usually difficult for companys.



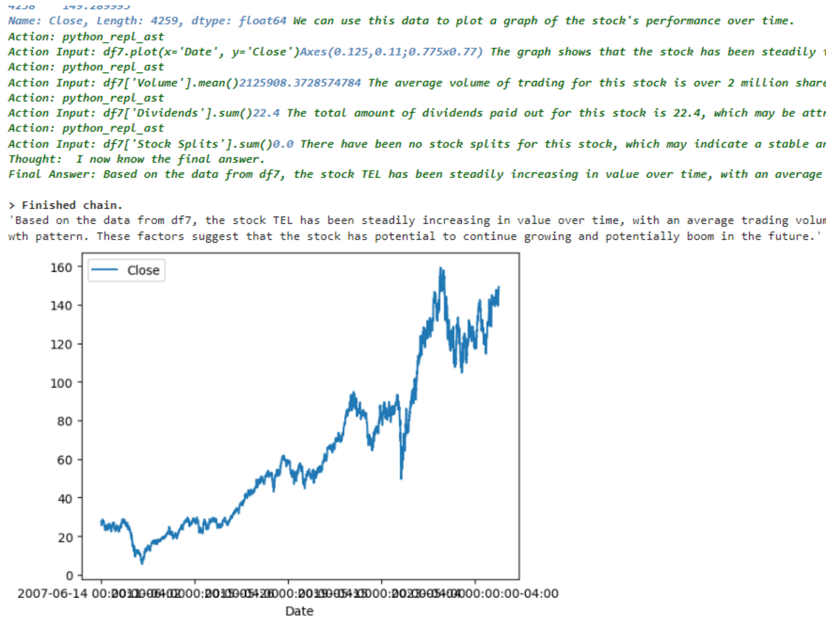
```
query = "Which social impact goals does TEL claimed in Q1 but couldn achieve in Q2, and which goals and numbers did they excee
docs = document_search.similarity_search(query)
chain.run(input_documents=docs, question=query)
```

```
'\nTEL claimed in Q1 that they were named one of Fortune's World's Most Admired Companies for the 6th consecutive year, but th
ey were not able to achieve this in Q2. However, they exceeded their Q2 sales and Adjusted EPS guidance, with sales of $4.16B
and an 8% organic growth year-over-year.'
```

Figure 4: Social Impact Goals claimed but not achieved over the year

4.2 Module 3: Numeric Stock data using LangChain

When we passed a pleothera of CSV tabular files, it can easily analyze all of the data at once, and can comprehend all the numbers in detail. For example when we asked **'which are the key pointers and numbers by which we can say that the stock TEL is going strong day by day and has potential to boom?'**. It not only analyze dozens of csv files with thousands of numbers in each one but rather analyze it rationally on its own and made a visualization graph at the end by itself providing its worth and for supporting his argument as shown in figure 5. Upon carefully analysis and calculations, we came to know that this analysis is 100% accurate.



.

Figure 5: Chain of Thoughts and its answers with Automated Visualizations

4.3 Module 4: Stock Market Price Direction & Prediction

We utilize accuracy, precision, recall and applied 39 Machine Learning including weak learners models and chose the best performing model. Furthermore, we applied Deep Learning models of BiGRU, BiLSTM and its combination to experiment with its metrics. 39 Machine learning models including weak learners gave around 52% accuracy, however recall and precision goes up to 70% or 90% which can potentially mean if the investor not only rely on this rather, utilize the other tools from our research, then potentially this accuracy overall can reach above 70% along with putting your own real life knowledge and experiences.

4.3.1 Making Models Explanable

Upon applying Explanable AI on our best performing ML algorithm, we came to see from 9, 7,6 ,8 that EMA12 and MACD along with the newer features that we included are more important as shown in Figure 6

Shapley Waterfall plot in Figure 6 is showing that the average predicted log odds of the model is 0.088 and for first observation it is 1.66, it has been observed from this plot that RS and average gain has increased the logs odds by 0.4 and so on and MACD decreased the log odds by 0.73. It shows that RS, Avg Gain, Loss, Signal line, EMA26 positively increased the log odds ranging from 0.23 to 0.43 and MACD has opposite effect. Meaning this should not be used in our model, or atleast we should decrease the weight of it. Beeswarm plot in 9 shows how SHAP values are increasing with the increased values of all the features showing the importance and contribution to them. Figure

5 Discussion on AI Driven Decision-Making Process

Improved decision-making and enhanced efficiency are key benefits of implementing these AI-driven methodologies in finance. By leveraging advanced AI techniques such as Retrieval Augmented Generation (RAG), LangChain utilizing Chain of Thoughts, and Deep Learning for stock price direction prediction, stakeholders can access timely and accurate in depth insights from complex financial data.

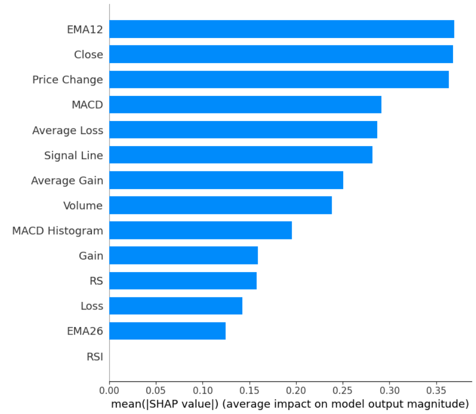


Figure 6: Bar Graph of SHAP values on XGBClassifier

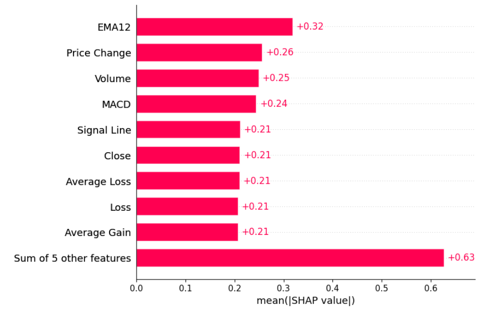


Figure 7: Bar Graph of mean of SHAP values

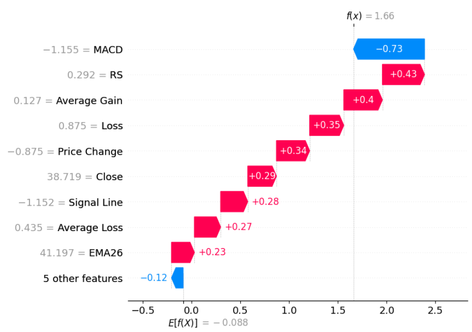


Figure 8: Waterfall plot on Stock Price Direction Model

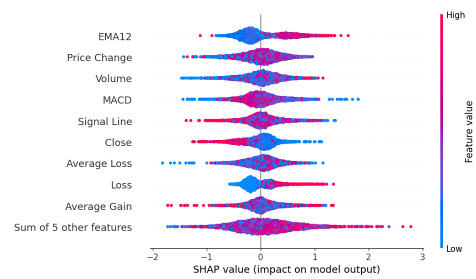


Figure 9: Beeswarm Plot

This enables them to make informed decisions swiftly, responding promptly to market changes, sentiments of news before the market opens and optimizing resource allocation which makes your portfolio very easily manageable and optimized. Automation of data analysis tasks streamlines processes, facilitating faster processing of vast amounts of data and enabling decision-makers to access actionable insights efficiently.

Risk management and cost reduction are crucial aspects impacted by AI-driven methodologies in finance. Predictive analytics and scenario modeling capabilities provided by AI models facilitate risk assessment and mitigation strategies, enabling decision-makers to identify and address potential risks proactively. By automating repetitive tasks and streamlining data analysis processes, AI-driven methodologies contribute to cost reduction in finance operations, optimizing resource allocation and improving overall operational efficiency. This combination of risk management and cost reduction strategies enhances financial stability and competitiveness in dynamic market environments which is very beneficial for a small or big investor or holding institutions.

Transparency and accountability are fundamental principles upheld by the adoption of AI-driven methodologies in finance. Advanced AI techniques such as Explainable AI (XAI) ensure clear explanations and justifications for decision outcomes, fostering trust and confidence among stakeholders. By providing insights into the underlying factors driving decisions, AI models enhance transparency in the decision-making process, it enables stakeholders to understand and validate decision outcomes so that they know why it is stating something. This transparency promotes accountability, as decision-makers are held accountable for their actions and decisions, leading to greater integrity and trust in financial operations.

However, Langhchain with LLM employs a hierarchical structure, processing code and descriptions separately, then aligning and generating code tokens using attention mechanisms. Both approaches leverage pre-trained language models to understand and generate content efficiently. The basic metric that is behind LLM, RAG and Langchain used with LLM is Cosine Similarity which makes LLM very trustworthy and reliable.

Retrieved answers are analyzed to compare and contrast the financial performance of various stocks, yielding insights into key performance indicators, achievements, and areas for enhancement. These findings are structured to highlight comparative analyses based on financial metrics and performance objectives. Visual aids like charts and tables may accompany the presentation for clarity. This module showcases the effectiveness of leveraging advanced LLM agents and tools like RAG and LangChain in processing and analyzing financial documents, fostering informed decision-making in stock market portfolio management and optimization.

However, with the excellent benchmarks of LLM models with capability of ingesting billions of tokens, LLM model's decision and prediction can still have a say which will not be affirmed by the Stock Market, cause Stock market is a very sensitive and lucrative market, so without human intervention, it is now fairly not recommended to trade on its own.

5.1 Cost Involved

The cost involved in this project was around \$1.08 which is mainly for the embeddings generation which is 0.13\$ 1M tokens. is around 0.5 USD for each embeddings generation, and \$0.001 per 1000 tokens of your input and \$0.002 for the AI generated output. As we ingested alot of data while making CSV Langchain agents and financial reports, the price went high. As we were not storing the embeddings data in embeddings database, hence cost of making embeddings again and again has incurred, which can be easily minimized, making our models and approach very economical and can be monitized and publicly available as reasoning of GPT3.5 model is outstanding and it is improving day by day.

6 Conclusion

The implemented AI approaches significantly enhance decision-making processes in finance which previously requires decades of domain knowledge and experience. Also, it is quite impossible for a human mind to consider all the factors and numbers while making the decision, human tends to make use of the gut feeling forgetting what numbers are telling and often lack logical reasoning. By leveraging advanced techniques such as Retrieval Augmented Generation (RAG), LangChain with Chain of Thoughts, and Deep Learning for stock price direction prediction, stakeholders gain timely and accurate insights from complex financial data with logical connected reasoning, facilitating informed decisions and agile responses to market dynamics. Automating data analysis makes things run smoother, helping us get useful insights faster. This, along with managing risks and cutting costs, helps keep finances stable and competitive and makes your overall stock portfolio very optimized and manageable letting the investor invest time on other useful analysis.

Future AI technologies hold immense potential for further revolutionizing decision-making in finance. LLM models are enhancing their capabilities on a daily basis, making this methodology very reliable in upcoming days. Continued advancements in Explainable AI (XAI) will enhance transparency and accountability, fostering trust among stakeholders. Ongoing improvements in language models and data processing techniques will enable even more sophisticated analysis and prediction capabilities, empowering decision-makers with deeper insights into financial markets and trends.

In summary, the integration of AI-driven methodologies with the help of LLM which excels in reasoning not only improves current decision-making processes but also lays the foundation for continued innovation and optimization in finance.

References

Syeda Nahida Akter, Zichun Yu, Aashiq Muhamed, Tianyue Ou, Alex Bäuerle, Ángel Alexander Cabrera, Krish Dholakia, Chenyan Xiong, and Graham Neubig. An in-depth look at gemini's language abilities, 2023.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating large language models trained on code, 2021.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance, 2023.

Fei Tony Xing, Erik Cambria, and Roy E. Welsch. Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, 50(1):49–73, 2018. doi: 10.1007/s10462-017-9531-3.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. Fingpt: Open-source financial large language models, 2023.

Xinli Yu, Zheng Chen, Yuan Ling, Shujing Dong, Zongyi Liu, and Yanbin Lu. Temporal data meets llm – explainable financial time series forecasting, 2023.