2024 IEEE Pune Section International Conference (PuneCon)
Pune, India. Dec 13-15, 2024

Equity Research Chatbot Using LLM: A Responsive Agent for Investment Research.

1st Nikhil Dongare
Undergraduate student in the Artificial Intelligence and Data Science Department at the Vishwakarma Institute of Information Technology,
Pune-411048, India
nikhildongare2003@gmail.com

2nd Amey Bhirange
Undergraduate student in the Artificial Intelligence and Data Science Department at the Vishwakarma Institute of Information Technology,
Pune-411048, India
bhirangeamey@gmail.com

3rd Shreyas Kharche
Undergraduate student in the Artificial Intelligence and Data Science Department at the Vishwakarma Institute of Information Technology,
Pune-411048, India
kharcheshreyas03@gmail.com

4th Amar Buchade
Associate Professor in the Artificial Intelligence and Data Science Department at Vishwakarma Institute of Information Technology,
Pune-411048, India
amar.buchade@viit.ac.in

5th Parikshit Mahalle
Professor in the Artificial Intelligence and Data Science Department at Vishwakarma Institute of Information Technology,
Pune-411048, India
parikshit.mahalle@viit.ac.in

2024 IEEE Pune Section International Conference (PuneCon) | 979-8-3315-2782-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/PuneCon63413.2024.10895230

***Abstract*— Our research paper includes the combination of chatbot development and research on various companies' fundamentals (P&L, Balance sheet, and Cash Flow) for giving end users precise investment advice.**

The literature on technologies like machine learning (ML), neural networks, and Generative AI in financial analysis is surveyed, emphasizing the growing body of research using Large Language Models (LLMs) and their applications using chatbot assistance for investment opportunities. Additionally, previous studies on chatbots in finance are reviewed to contextualize their role in equity research.

The methodology outlines the research approach, data collection method, and usage of LLMs in this study. Here the data used are the pdf files containing company financial reports which are then fed into the chatbot for further processing. The data regarding current market trends and overall fundamental analysis provides a clear idea to chatbot for suggesting precise investments. Here NLP processes include text chunking, building semantics, and knowledge base for proceeding to further steps. Further, our results include categories of questions for which we have also showcased its graph that mentions the response time and accuracy versus the type of question asked to the chatbot. This analysis gives an efficient understanding of the chatbot. Overall, this research contributes to the knowledge of how chatbots can transform equity research, paving the way for more efficient and accessible financial analysis in the digital age.

Keywords*— *Equity Research, Natural Language Processing, Artificial Intelligence, Chatbot, Lang chain, Generative AI, Large Language Models.

I. INTRODUCTION

Equity research stands as a main component of financial markets, offering invaluable insights to investors, analysts, and institutions seeking to navigate their investment decisions. In the domain of finance, equity research involves the thorough analysis of stocks and other financial instruments to give precise investment recommendations that are both informed and strategic. Investing in a fundamentally strong business can lead to potential multi-bagger gains that can beat the returns of Fixed Deposit, Gold, or any other form of investments over the years. Equity researchers have the task of identifying such businesses, thoroughly analyzing them and suggesting those businesses to their clients to make investments. Fundamental investing involves a three-step process of researching the business model, its financials, and conference call analysis. Researching business models involves evaluating growth orientation, industry growth potential, competitive advantage, potential leadership in the industry, and the impact of government interference, among other factors. The final and most challenging step is reviewing recent quarterly conference call transcripts, where company management addresses investor questions on the company's status, market outlook, and future goals.

The purpose of this chatbot is to give you an answer quickly to the question by finding it from the PDFs (conference call transcripts). This chatbot is selective based on the type of questions you ask. It is a Large Language model (LLM) that gives answers to any of your questions regarding the fundamental research of the business. This research serves as a benchmark for the potential of NLP and Generative AI in revolutionizing investment research and advice for people who want to invest in various businesses.

II. LITERATURE REVIEW

This study explores an innovative investment strategy using NLP on Korean equity research reports and ML algorithms for binary classification, predicting stock price movements with 33 features. Leveraging KoNLPy and MeCab, the approach outperforms pre- and mid-COVID-19 benchmarks, especially in the post-COVID-19 period. However, the study lacks real-time adaptability. Our equity research chatbot can bridge this gap by providing timely, personalized investment recommendations, further enhancing user experience and returns.[1]

This study examines the shift from traditional interfaces to those powered by NLU, enabling real-time processing of text and numerical data. By using LSTM and RNN models for stock value prediction, it focuses on improving accuracy through hyperparameter tuning. However, it lacks user-specific customization. Our chatbot can fill this gap by

delivering personalized advice thereby enhancing user engagement and decision-making.[2]

This study explores models that grasp user intent and generate human-like responses, focusing on generative and retrieval-based approaches using multi-layer neural networks. These methods are key for developing an equity chatbot capable of delivering accurate, context-aware investment advice. However, it lacks real-time adaptability. Our chatbot can enhance this by offering personalized responses, further improving the user experience.[3]

This study introduces a method that combines LSTM networks with a mean-variance model to optimize portfolio performance by capturing financial time series dependencies and focusing on high-return assets. These techniques could enhance an equity chatbot's ability to recommend optimized portfolios. However, the study doesn't address real-time user-specific customization. Our chatbot can fill this gap by providing personalized, numeric data answers according to individual user needs.[4]

This study examines chatbot deployment, focusing on user adoption, attitudes, and overall acceptance, along with user experience aspects like perception, trust, and engagement. These insights are crucial for designing an equity chatbot with a more intuitive and user-friendly interface. However, the study doesn't address real-time adaptability. Our chatbot can enhance this by continuously learning from user interactions.[5]

This research delves into Generative AI and Pre-trained Transformers, highlighting the capabilities of Large Language Models (LLMs) in addressing real-world challenges and serving as a basis for future innovations. These technologies are vital for developing an equity chatbot that delivers accurate, contextually relevant investment advice. However, it lacks personalization. This can be done by providing different investment PDFs to the chatbot to improve relevance and effectiveness.[6]

This study addresses the challenges of developing a chatbot with linked data, focusing on interactive interfaces, machine learning for classification, and NLP. It underscores the need for semantic data exploration, a robust Knowledge Base, and accurate data-driven decision-making. However, the study doesn't cover real-time updates. Our chatbot can enhance this by providing up-to-date, precise investment recommendations based on the latest market data and the type of questions asked by users.[7]

The study highlights the crucial role of Natural Language Processing (NLP), demonstrated through the use of NLTK for analyzing speech input and generating human-like responses. This has led to an increased demand for virtual assistants such as Siri, Cortana, Google Assistant, and Alexa. The paper also discusses the architecture needed for effective query response generation. These insights are valuable for developing an equity chatbot, enabling it to process user queries accurately and deliver relevant investment advice.[8]

III. METHODOLOGY

Our proposed methodology consists of developing an equity research chatbot wherein we give PDF files(consisting of company financials and equity research reports) as input to our chatbot which then further generates required responses based on provided prompts to it. Fig. 2 is our system architecture for the backend working of the chatbot. According to this drawn architecture, we have explained every step in depth. Also, we have divided the whole workflow into two phases where the first phase consists of extracting text to create a Knowledge Base and the second phase includes the user query response generation.

PHASE I

Step1: - PDF Data Extraction

This is the first step of the process where we give PDF data as input to our system. After inputting these PDF files, we extract data using PyPDF (Python library for merging, splitting, and transforming the pages of PDF files). Here we are first splitting text and then extracting this text as data for further analysis and processing.

Step 2: - Chunking

Chunking refers to the process we use to break down large text data into smaller, more manageable segments or chunks. These segments serve as the input units that our LLMs process, analyze, and generate responses for. The Gemini Pro 1.0 Model can handle 32,760 tokens at once. Therefore, we need to split the input data into chunks. In our approach, we create chunks of 10,000 characters each so that the model can process each chunk individually.

Step 3: - Embedding

In this process, we are converting text chunks into embeddings, which involves representing the text in a high-dimensional vector space. For clarity, we can refer to the diagram in Fig. 1. Each part in the diagram represents a distinct feature of the text. This embedding process allows us to capture the semantic meaning of words, and in our system, these captured semantics focus on key metrics of equity fundamentals. These metrics are crucial for analyzing and developing investment strategies. To create these high-dimensional numerical vectors, we use techniques like Word2Vec and Doc2Vec, which streamline the analysis process.

Step 4: - Building Semantic Index

After creating embedded vectors that capture the semantic meaning of words, our next step is to conduct a similarity search among these vectors to combine them into a Knowledge Base. We carry out this process using the FAISS (Facebook AI Similarity Search) library from Langchain. This library is particularly effective for performing similarity searches and creating clusters of vectors. It includes algorithms that search through batches of vectors of any size. By finding the nearest neighbor vector, we can determine similarity and feed this data into the Knowledge Base. The nearest-neighbor similarity is crucial because combining text related to one another enhances the chatbot's understanding.

Step 5: - Knowledge Base

It serves as a storage place where we store semantically indexed text, which we can use when a user submits a relatable search query. So, when a user writes a prompt related to the content we've indexed, the chatbot responds with the most suitable result.

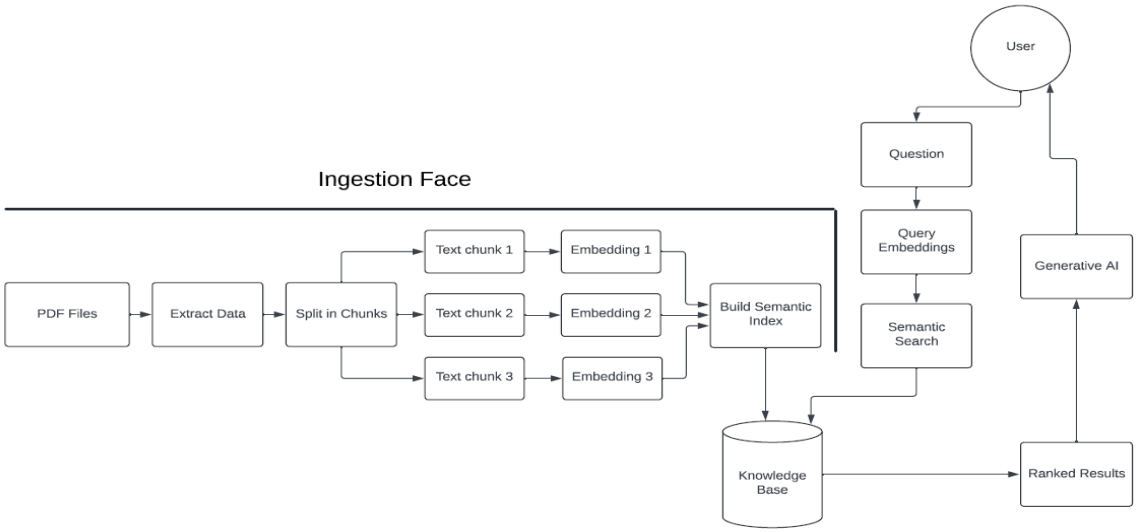

Fig. 1. Architecture Flow of Chatbot Development.

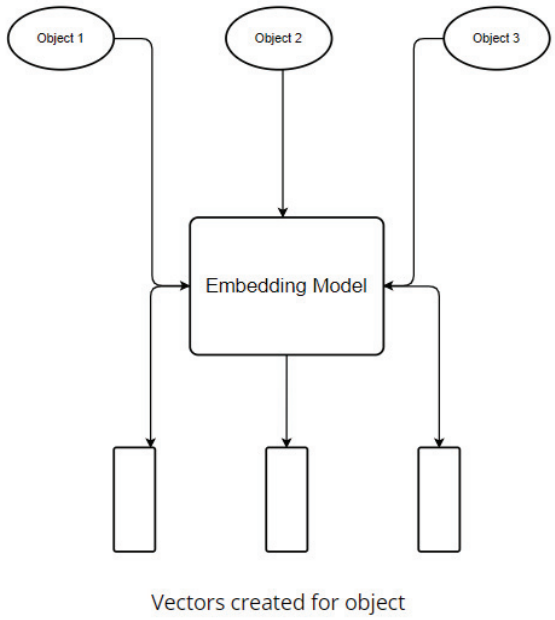

Fig. 2. Embedding Model Architecture

PHASE II

Step 1: - User Prompt

When a user asks a question to our model, it creates vector embeddings of the user's prompt. The model separates the input into two parts: context and question. The context provides essential background information that the model needs to grasp about the topic, while the question directs the search within the knowledge base, ensuring an accurate and relevant response. In this case, the context could include financial statements, news articles, analyst reports, or any other data source that provides background information. The question placeholder will contain the specific questions we want the chatbot to answer based on the provided context. We can refer to Fig. 3 for understanding user prompt requirements.

Step 2: - Query Embedding

It is the process where we break down the question asked by the user into chunks, making it easier to provide a suitable response to their prompt. During this process, we also perform vectorization, meaning we create numerical vectors and store them in a semantic search. These vectors are then compared with the data from our Knowledge Base to generate the most suitable results.

Step 3: - Actual Response Generation

In this step, when we compare the query's embedded data with the data from the Knowledge Base, a list of related and suitable responses is generated. These responses are then sorted based on their confidence level. The chatbot presents us with the highest-ranked response, which has the highest confidence. Here, we can observe the use of Generative AI, which enables us to generate responses. By obtaining ranked results, Generative AI provides us with improved accuracy, as it compares the context from both the query and the Knowledge Base to deliver an accurate response.

```
prompt_template = """
    **Answer the question as detailed as possible,

    **Context format:** The context will be provide
analyst's questions
    and management's answers format, financial repo

    **Question understanding:** Identify the type o

    **Answer generation:**

    * If the question asks for a fact or number-bas
        * Use your knowledge base or find the answe
        * Cite the source of the information.

    * If the question is behavioral and the answer

    * If the user asks for investment advice, state
      Please consult a financial advisor before mak

    **Answer:**\n\n
    Context:\n {context}?\n
    Question: \n{question}\n

    Answer:
    """
```

Fig. 3. User Prompt Template

IV. RESULTS AND DISCUSSIONS

By studying the text data from PDF, we have derived our results into 3 types of questions that are asked to the model to generate answers.

1) **Fact / Number-based questions:** Answers to these questions are certain to be given correctly they are present in the input files and are fact-based. The question in Fig. 4 asking about the operating margin is an example of this type of question.
2) **Behavioral brief questions:** These types of questions are only answered if the answers are given by management in the conference and are available in the pdf. If the answer to these types of questions is not present in the PDF, it is not answered saying management has not answered this question. For example. Fig. 7, is a question related to a plan of the company to beat the competition.
3) **Investment advice questions:** If how the user asks for any kind of investment advice, the model will deny the response as it is not appropriate to give such advice in public according to government rules. For simplicity, we have given the model 3 conference call transcripts of Infosys company, for the first 3 quarters of the financial year 2023-24. We have asked different types of questions to the model to see the answers given by the model.

Following are the outputs that are generated for each type of question.

Fig. 4. Fact-based questions

Fig. 4 represents the question regarding the reason behind the slowed-down growth of the company and the response accurately and concisely.

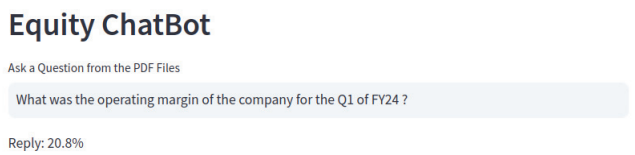

Fig. 5. Number-based questions

Fig. 5 gives the one-line answer to a number-based question. Since, the equity researcher would only want to know the number here, and not anything else to read, the model is giving only one line answer.

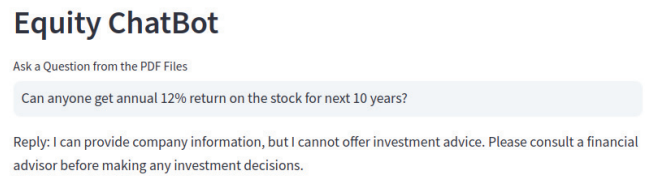

Fig. 6. Investment advice questions

If the user asks any type of direct question that would directly impact the user's investment decision, you would not want that to happen. In Fig. 6 a user asks if he could get a certain return for the next 10 years, and you are never sure about the stock market and the future so you cannot give any kind of answer to this question. So, the model declines the response and asks the user to consult their financial advisor.

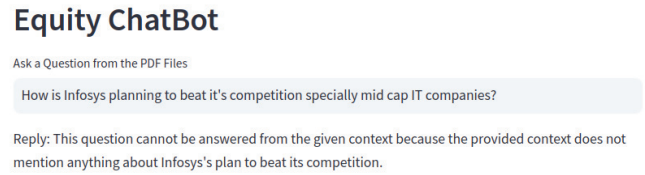

Fig. 7. Behavioral brief questions

Fig. 7 shows a user asking the question about management's plans to beat the competition of mid-cap growing companies the model responds that such a question has not been answered in the conference and it is not available in the pdf. Here we do not want to give any wrong answer because which user's investment thesis might change.

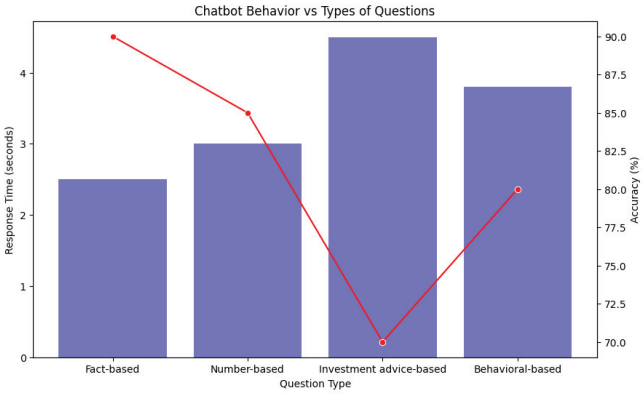

Fig. 8. Analysis of Chatbot Accuracy

Fig. 8 explains our chatbot behavior across four types of questions: fact-based, number-based, investment advice-based, and behavioral-based. The bar graph represents response time (in seconds), with investment advice-based questions taking the longest. The line graph represents accuracy, with investment advice-based questions having the lowest accuracy, while fact-based questions are the most accurate. Overall, there is an inverse relationship between response time and accuracy for most question types. Therefore, this analysis validates our work and the efficiency of our chatbot.

V. CONCLUSION

In summary, our study explores the development of a conversational agent designed to assist with the fundamental analysis of stocks, focusing on generating accurate financial advice. By integrating domain knowledge in finance with

advanced technologies such as LLMs, GPTs, NLTK, and Google Gemini Pro, we aim to create a chatbot with high response accuracy. Our analysis emphasizes the importance of asking the right types of questions to obtain reliable answers, particularly from credible sources like conference calls and annual reports. The chatbot is intended to support equity researchers by providing accurate, trustworthy information without offering speculative or low-credibility responses.

VI. FUTURE SCOPE

The current work of this chatbot in generating investment advice is promising, but there's always room for growth. The first step in future development is to connect a reliable data source, automating data retrieval and eliminating manual uploads. Integrating with dynamic databases will ensure real-time access, enhancing the speed and accuracy of analyses. This setup allows continuous updates, keeping the model aligned with the most current data for precise insights.

By incorporating direct data fetching for specific companies upon user request, the model's functionality will be significantly enhanced. This capability will enable instant access and analysis of company-specific data, providing tailored insights on demand. Automation will increase efficiency, offering users a seamless experience for quicker decision-making and more informed strategies.

ACKNOWLEDGEMENT

We would appreciate the teamwork of our team for this project. Their commitment and dedication helped us reach our goals. We also thank Vishwakarma Institute of Information Technology, Pune for supporting us. Lastly, we thank Amar sir and Parikshit sir for their great mentorship and support.

REFERENCES

[1] Cho, P., Park, J. H., & Song, J. W. (2021). Equity research report-driven investment strategy in Korea using binary classification on stock price direction. *IEEE Access*, *9*, 46364-46373.

[2] Lauren, P., & Watta, P. (2019, December). A conversational user interface for stock analysis. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 5298-5305). IEEE.

[3] Vamsi, G. K., Rasool, A., & Hajela, G. (2020, July). Chatbot: A deep neural network based human to machine conversation model. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

[4] Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*, *143*, 113042.

[5] Wube, H. D., Esubalew, S. Z., Weldesellasie, F. F., & Debelee, T. G. (2022). Text-based chatbot in financial sector: a systematic literature review. *Data Sci. Financ. Econ*, *2*(3), 232-259.

[6] Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., ... & Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*.

[7] Ait-Mlouk, A., & Jiang, L. (2020). KBot: a Knowledge graph based chatBot for natural language understanding over linked data. *IEEE Access*, *8*, 149220-149230.

[8] Karri, S. P. R., & Kumar, B. S. (2020, January). Deep learning techniques for implementation of chatbots. In *2020 International conference on computer communication and informatics (ICCCI)* (pp. 1-5). IEEE.