# **Introduction**

The rise of Generative Artificial Intelligence (AI) has revolutionized how humans interact with machines, enabling systems that can understand, generate, and respond to human language in increasingly sophisticated ways. Among the most impactful advancements are Large Language Models (LLMs), which are trained on massive corpora of text and capable of performing a wide range of natural language understanding and generation tasks. This project leverages the power of generative AI through the integration of advanced LLMs to build an intelligent assistant for stock market analysis.

The goal of this project is to develop a conversational AI chatbot that can respond to user queries related to stock performance, market trends, company analysis, and investment concepts. Unlike traditional financial information systems, which require manual data lookups or static dashboards, this chatbot provides an interactive, natural language interface capable of delivering contextual and real-time insights. The chatbot is implemented as a web-based application using Streamlit and integrates LLMs through Groq’s high-performance language model API, currently supporting models like LLaMA 3.3 70B and Gemma 2 9B.

This system is enhanced further by incorporating real-time financial data through LangChain tools such as Google Finance and the SERP API. The chatbot processes user input, extracts relevant financial keywords, and formulates dynamic prompts that combine static language understanding with up-to-date market information. The result is a powerful financial assistant that bridges the gap between generative AI and data-driven finance.

This project sits at the intersection of natural language processing, financial data analysis, and interactive software design. It not only demonstrates the practical application of generative AI in a real-world domain but also highlights the flexibility and intelligence of LLM-based agents when extended with external tools and APIs.

In the context of the Generative AI subject, this project provides a hands-on exploration of:

* Prompt engineering and context injection for LLMs
* Multi-model deployment using LLM APIs
* Tool-augmented agents for real-time information retrieval
* Ethical considerations in AI-generated financial advice

The development of this chatbot not only showcases technical proficiency in working with generative models but also underscores the growing role of AI in transforming decision-making across industries such as finance.

# **Project Aim**

The primary aim of this project is to develop a conversational AI chatbot that harnesses the capabilities of advanced large language models to deliver contextual and real-time insights into stock market trends, company performance, and investment analysis. This chatbot is designed to integrate generative AI with live financial data retrieval, enabling users to interact naturally while receiving informed responses that combine static language understanding with dynamic market information.

Specifically, the project seeks to:

* Leverage state-of-the-art generative AI models to interpret and generate human-like responses.
* Integrate external financial data sources through tools such as Google Finance and SERP API to ensure that the responses reflect current market conditions.
* Enhance user experience by employing prompt engineering techniques that merge static AI-generated content with live financial insights.
* Provide a secure and interactive web interface that simplifies access to complex financial data and analysis for both novice and expert users.

By achieving these objectives, the project demonstrates the practical application of generative AI in financial analysis and decision support, illustrating the potential for AI-driven tools to transform traditional methods of accessing and interpreting financial information.

# **Literature Review**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Title** | **Authors** | **Publication Year** | **Key Insights** |
| 1 | [Stock Market Prediction Through a Chatbot: A Human-Cantered AI Approach](https://www.researchgate.net/publication/362249264_Stock_Market_Prediction_Through_a_Chatbot_A_Human-Centered_AI_Approach) | [Anoushka Halder](https://www.researchgate.net/profile/Anoushka-Halder?_sg%5B0%5D=tng9b6kcOhdha3qX8YdTFxuAJfPC2tJSBJXj4hXKmVfRKpGH5FqVldDrKYiMhKq1DfWL0EI.YZv1-_wMprWHAo9d9NZibUJblr75ySg48fVbhmpTU021hnsVTmyLMI2S0SNS2uQ5-7fXxNYA9DQlGd_Yy2FzUw&_sg%5B1%5D=IWoKrbORmfr-lOXeSWQw3Bj88vBC73cH0m02HFO-CMzkES4j_LrvxFeQP360VwJtt-qcJ7o.uQTP7Gr4nctzDeSElltkwempbaBs-czl7fbx1EAxNaqE_8dQTI2RA-SAfmw5xgg2T2RD3Xy5_RHx5C9EsNQQog&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24iLCJwb3NpdGlvbiI6InBhZ2VIZWFkZXIifX0) , [Aayush Saxena](https://www.researchgate.net/profile/Aayush-Saxena-7?_sg%5B0%5D=tng9b6kcOhdha3qX8YdTFxuAJfPC2tJSBJXj4hXKmVfRKpGH5FqVldDrKYiMhKq1DfWL0EI.YZv1-_wMprWHAo9d9NZibUJblr75ySg48fVbhmpTU021hnsVTmyLMI2S0SNS2uQ5-7fXxNYA9DQlGd_Yy2FzUw&_sg%5B1%5D=IWoKrbORmfr-lOXeSWQw3Bj88vBC73cH0m02HFO-CMzkES4j_LrvxFeQP360VwJtt-qcJ7o.uQTP7Gr4nctzDeSElltkwempbaBs-czl7fbx1EAxNaqE_8dQTI2RA-SAfmw5xgg2T2RD3Xy5_RHx5C9EsNQQog&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24iLCJwb3NpdGlvbiI6InBhZ2VIZWFkZXIifX0) , [Priya Milton](https://www.researchgate.net/profile/Priya-Milton?_sg%5B0%5D=tng9b6kcOhdha3qX8YdTFxuAJfPC2tJSBJXj4hXKmVfRKpGH5FqVldDrKYiMhKq1DfWL0EI.YZv1-_wMprWHAo9d9NZibUJblr75ySg48fVbhmpTU021hnsVTmyLMI2S0SNS2uQ5-7fXxNYA9DQlGd_Yy2FzUw&_sg%5B1%5D=IWoKrbORmfr-lOXeSWQw3Bj88vBC73cH0m02HFO-CMzkES4j_LrvxFeQP360VwJtt-qcJ7o.uQTP7Gr4nctzDeSElltkwempbaBs-czl7fbx1EAxNaqE_8dQTI2RA-SAfmw5xgg2T2RD3Xy5_RHx5C9EsNQQog&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24iLCJwb3NpdGlvbiI6InBhZ2VIZWFkZXIifX0) | 2022 | Proposes a chatbot that utilizes deep learning models to predict stock closing values and responds to user inquiries, emphasizing a user-friendly interface for financial data interaction. ​[ResearchGate](https://www.researchgate.net/publication/362249264_Stock_Market_Prediction_Through_a_Chatbot_A_Human-Centered_AI_Approach?utm_source=chatgpt.com) |
| 2 | [A Survey of Large Language Models for Financial Applications](https://arxiv.org/html/2406.11903v1) | Not specified | 2023 | Explores the application of LLMs in various financial tasks, highlighting their potential to transform traditional practices and drive innovation in financial analysis. ​[arXiv+1arXiv+1](https://arxiv.org/html/2406.11903v1?utm_source=chatgpt.com) |
| 3 | [Artificial Intelligence Techniques in Financial Trading: A Systematic Literature Review](https://www.sciencedirect.com/science/article/pii/S1319157824001046) | Fatima Dakalbab, Manar Abu Talib, Qassim Na, Tracy Saroufil | 2024 | Reviews 143 research articles on AI techniques in financial trading, providing a comprehensive overview of AI applications in stock market prediction and trading strategies. ​[ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1319157824001046?utm_source=chatgpt.com) |
| 4 | [Large Language Models for Financial and Investment Management](https://www.pm-research.com/content/iijpormgmt/51/2/211) | Yaxuan KongYuqi, NieXiaowen DongJohn, M. MulveyH, Vincent PoorQingsong, WenStefan Zohren | 2023 | Discusses the integration of LLMs in financial services, focusing on applications like fraud detection and generative AI, and their impact on investment management. ​[PM Research](https://www.pm-research.com/content/iijpormgmt/51/2/211?utm_source=chatgpt.com) |
| 5 | [The Effect of Implementing Chatbot Customer Service on Stock Returns: An Event Study Analysis](https://www.researchgate.net/publication/359400831_The_effect_of_implementing_chatbot_customer_service_on_stock_returns_an_event_study_analysis) | [Michael A Wiles](https://www.researchgate.net/profile/Michael-Wiles-4?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19), [Darima Fotheringham](https://www.researchgate.net/profile/Darima-Fotheringham?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19) | 2022 | Analyses the impact of AI chatbot announcements on stock returns, finding a 0.22% abnormal return, indicating positive investor response to AI customer service implementations. ​[Texas Tech University+2Arizona State University+2ResearchGate+2](https://asu.elsevierpure.com/en/publications/the-effect-of-implementing-chatbot-customer-service-on-stock-retu?utm_source=chatgpt.com) |
| 6 | [Trading Insight AI: An Intelligent Research Tool for Equity Analysts](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID5065576_code7207861.pdf?abstractid=5065576&mirid=1) | Komal Kokare, Sakshi Pimpale, Nikita Kapadnis, Pranali Navghare, Dipika Raigar, Kimaya R. Urane | 2024 | Introduces a platform utilizing NLP technologies like LangChain and the Groq API to automate data gathering and analysis, providing real-time, accurate insights for financial decision-making. ​[SSRN](https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID5065576_code7207861.pdf?abstractid=5065576&mirid=1&utm_source=chatgpt.com) |
| 7 | [Re(Visiting) Large Language Models in Finance](https://papers.ssrn.com/sol3/Delivery.cfm/4963618.pdf?abstractid=4963618&mirid=1) | Eghbal Rahimikia, Felix Drinkall | 2023 | Evaluates specialized LLMs developed for accounting and finance, demonstrating their superior performance in return prediction compared to general-purpose LLMs. ​[SSRN](https://papers.ssrn.com/sol3/Delivery.cfm/4963618.pdf?abstractid=4963618&mirid=1&utm_source=chatgpt.com) |
| 8 | [FinBloom: Knowledge Grounding Large Language Model with Real-time Financial Data](https://arxiv.org/abs/2502.18471) | Ankur Sinha, Chaitanya Agarwal, Pekka Malo | 2025 | Introduces FinBloom 7B, a 7-billion-parameter LLM trained on financial news and SEC filings, fine-tuned to handle financial queries using real-time text and tabular data, enhancing LLM capabilities in dynamic financial tasks. ​[arXiv+1PM Research+1](https://arxiv.org/abs/2502.18471?utm_source=chatgpt.com) |
| 9 | [An Agent Framework for Real-Time Financial Information Searching with Large Language Models](https://arxiv.org/abs/2502.15684) | Jinzheng Li, Jingshu Zhang, Hongguang Li, Yiqing Shen | 2024 | Presents FinSearch, a novel agent-based search framework designed for financial applications, interfacing with diverse financial data sources and utilizing LLMs for structured query decomposition and dynamic information retrieval. ​[arXiv](https://arxiv.org/abs/2502.15684?utm_source=chatgpt.com" \t "_blank) |
| 10 | [Sentiment Trading with Large Language Models](https://arxiv.org/abs/2412.19245) | Kemal Kirtac, Guido Germano | 2024 | Investigates the efficacy of LLMs in sentiment analysis of U.S. financial news and their potential in predicting stock market returns, highlighting the superior performance of advanced LLMs in financial market prediction. ​[arXiv](https://arxiv.org/abs/2412.19245?utm_source=chatgpt.com" \t "_blank) |

# **Methodology**

The development of the Stock Market Analysis Chatbot followed a modular and layered methodology, combining techniques from generative AI, natural language processing, financial data integration, and web application development. The methodology can be divided into the following key components:

**1. Problem Definition and Scope**

The project was initiated with the objective of creating a generative AI system capable of understanding and responding to stock market-related queries. The chatbot was designed to serve both as an educational and analytical tool, delivering responses that blend real-time financial data with AI-generated explanations.

**2. Selection of Generative AI Models**

To ensure high-quality and context-aware responses, the project employed state-of-the-art Large Language Models (LLMs) using Groq's low-latency LLM infrastructure. Two primary models were selected:

* **LLaMA 3.3 70B**: A versatile and powerful model for deep financial reasoning and contextual analysis.
* **Gemma 2 9B**: A lighter model used for faster, concise responses where efficiency is prioritized.

These models were accessed through the LangChain-Groq integration, allowing seamless invocation within the chatbot's backend.

**3. Tool Augmentation with Real-Time Financial Data**

To supplement LLM-generated responses with live market data, the chatbot integrated LangChain's toolset for external API interaction. Specifically:

* The **Google Finance tool** was used via the LangChain load\_tools() function.
* The **SERP API** was configured to fetch live stock and market information.
* A **FakeListLLM** was employed to trigger tool-based responses without invoking generative outputs, ensuring tool use was isolated to data retrieval.

This setup enabled the chatbot to retrieve current stock prices, market indices, and financial trends based on the user’s query.

**4. Prompt Engineering and Contextualization**

To improve the accuracy and usefulness of the LLM responses:

* The system first extracted **financial keywords and stock symbols** from the user’s query using custom logic.
* If financial terms or symbols were found, relevant queries were automatically generated and passed to the tool agent.
* The retrieved real-time data was **embedded into the final LLM prompt** to ground the response in current market context.

This technique allowed the generative model to operate with up-to-date data while retaining its natural language generation capabilities.

**5. User Interface Development**

A user-friendly interface was created using **Streamlit**, enabling seamless interaction between users and the backend AI system. Key UI components included:

* A sidebar for model selection and API key status
* A dark-themed chat window for interactive querying
* Contextual cards that display welcome messages, usage tips, and disclaimer notes

The app was styled using custom CSS to enhance readability and modernize the design.

**6. Environment and API Configuration**

All API keys and configurations were securely managed using a .env file and the python-dotenv library. This ensured modularity and allowed for quick reconfiguration during testing and deployment.

**7. Iterative Testing and Enhancement**

The chatbot was tested across various stock-related queries, including:

* Direct stock symbol lookups (e.g., "Price of AAPL")
* Broader market questions (e.g., "How is the S&P 500 doing?")
* Conceptual analysis (e.g., "How do interest rates affect tech stocks?")

Testing led to improvements in:

* Keyword extraction accuracy
* Prompt structure
* Model/tool coordination
* Error handling and fallback mechanisms

# **Preliminary Results**

The Stock Market Analysis Chatbot was evaluated through a series of functional tests and user interaction scenarios to assess its effectiveness, accuracy, and usability. These preliminary results highlight the chatbot’s ability to combine generative AI with real-time financial data to deliver coherent and relevant responses.

**1. Functional Accuracy**

* The chatbot consistently identified key financial terms and stock symbols from a wide range of user queries.
* It correctly invoked LangChain tools for relevant queries, retrieving accurate real-time data (e.g., current stock prices, index performance).
* Prompts generated for the language models effectively blended user intent with up-to-date financial context, improving response quality.

**2. Model Response Quality**

* The LLaMA 3.3 70B model produced rich, contextual explanations suitable for intermediate to advanced financial understanding.
* The Gemma 2 9B model responded faster with concise and accurate summaries, making it suitable for casual or quick inquiries.
* In cases where live data was unavailable (e.g., invalid stock symbols), the system gracefully defaulted to general model-based responses.

**3. Real-Time Data Integration**

* Integration with Google Finance tools via LangChain enabled the chatbot to fetch:
  + Stock prices (e.g., AAPL, GOOGL, NVDA)
  + Market indices (e.g., Nasdaq, Dow Jones)
  + Company-specific performance metrics
* The system was able to embed this data into the prompt context, resulting in more informative and grounded replies.

**4. User Interface and Usability**

* The Streamlit interface allowed seamless user interaction, with users able to:
  + Choose between available models
  + View system messages, responses, and warnings clearly
  + Enter natural language queries without requiring technical input
* The dark-themed UI and structured chat layout improved readability and engagement.

**5. Observed Limitations (Preliminary)**

* Occasional API rate limits or delays in real-time data retrieval (dependent on SERP API responsiveness).
* The LLM sometimes generated overly verbose responses when minimal answers were sufficient.
* Financial advice disclaimers needed to be reinforced in model prompts to align with responsible AI usage.

# **References**

1. K. Kirtac and G. Germano, “Sentiment Trading with Large Language Models,” *arXiv preprint arXiv:2412.19245*, 2024. [Online]. Available: <https://arxiv.org/abs/2412.19245>
2. “Large Language Models for Financial and Investment Management,” *The Journal of Portfolio Management*, vol. 51, no. 2, pp. 211, 2023. [Online]. Available: <https://www.pm-research.com/content/iijpormgmt/51/2/211>
3. “A Survey of Large Language Models for Financial Applications,” *arXiv preprint arXiv:2406.11903v1*, 2023. [Online]. Available: <https://arxiv.org/html/2406.11903v1>
4. “Stock Market Prediction Through a Chatbot: A Human-Centered AI Approach,” *ResearchGate*, 2022. [Online]. Available: <https://www.researchgate.net/publication/362249264>
5. “Trading Insight AI: An Intelligent Research Tool for Equity Analysts,” *SSRN*, 2024. [Online]. Available: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN\_ID5065576\_code7207861.pdf?abstractid=5065576
6. “LangChain: Building Applications with LLMs through Composability,” *LangChain Docs*. [Online]. Available: https://docs.langchain.com
7. “Streamlit – The fastest way to build and share data apps,” *Streamlit.io*. [Online]. Available: <https://streamlit.io>
8. “Groq API Documentation,” *GroqCloud*. [Online]. Available: https://docs.groq.com
9. “Google Finance API via SERP,” *SERP API Documentation*. [Online]. Available: https://serpapi.com/google-finance-api
10. “yFinance: Yahoo! Finance for Python,” *yFinance GitHub Repository*. [Online]. Available: <https://github.com/ranaroussi/yfinance>