

# Finchat: LLM-Powered Finance Assistant for WhatsApp Integration

## Abstract

The project *Finchat* aims to develop a finance assistant chatbot integrated with WhatsApp, leveraging Large Language Model (LLM) technology to provide users with finance-related information and insights. The chatbot utilizes the OpenAI API for natural language processing and a custom-built LLAMA-3 8B model for generating finance insights. The workflow involves extracting user queries, retrieving historical financial data, preprocessing data for model input, and delivering personalized responses to users via WhatsApp.

## 1 Introduction

In the digital era, chatbots have become essential tools for delivering personalized services and information. This project introduces "Finchat," a finance assistant chatbot designed to interact with users via WhatsApp, providing real-time financial analysis and recommendations. Leveraging advanced Large Language Model (LLM) technologies, including the OpenAI API and a custom LLAMA-3 8B model, Finchat aims to significantly enhance user engagement by offering valuable financial insights. The increasing complexity of financial markets and the growing demand for quick access to accurate and personalized information highlight the necessity for intelligent chatbot solutions like Finchat. This chatbot system is developed with specific objectives: to interpret user queries related to finance, integrate LLM technologies for extracting insights from historical financial data, enable seamless interaction via the WhatsApp Business API, and deliver personalized and accurate finance recommendations. Through these objectives, Finchat seeks to empower users with accessible and actionable financial information directly on the familiar platform of WhatsApp, revolutionizing how individuals interact with financial data in real-time.

## 2 Literature Survey

The literature survey on LLMs, particularly focusing on the LLAMA-3 family, presents a detailed analysis of the dynamic progress in language modeling.

The advent of LLMs such as GPT and LLAMA, catalyzed by the introduction of ChatGPT, has propelled significant advancements in natural language understanding and generation. Their expansive, pre-trained architecture distinguishes LLMs, achieved by training billions of parameters on extensive text datasets. This survey tracks the evolution of LLMs from statistical and neural language models to transformer-based architectures like BERT, culminating in LLMs boasting billions of parameters. The paper categorizes LLMs into distinct families—GPT, LLAMA, and PaLM—providing insights into their unique traits, contributions, and emerging functionalities. It explores the construction, deployment, and enhancement of LLMs for diverse tasks, highlighting their potential as the cornerstone for artificial general intelligence (AGI) and versatile AI agents. Additionally, the survey examines prominent datasets and benchmarks employed in LLM evaluations, outlining key challenges and charting future research trajectories in this rapidly evolving domain. This comprehensive investigation gives researchers and practitioners valuable perspectives on harnessing LLMs for various applications and tackling pivotal language modeling hurdles.

### 3 Proposed System Architecture

The proposed system architecture delineates the steps for processing a user’s finance-related query on WhatsApp, leveraging LLM technologies and financial APIs to deliver insights and responses efficiently.

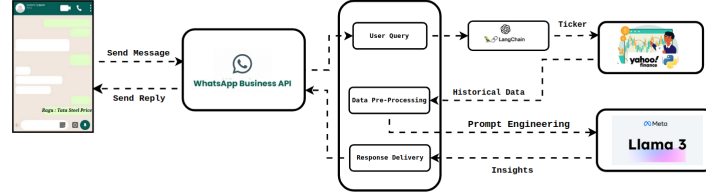


Figure 1: System Architecture

#### 3.1 Workflow Steps

##### 1. User Interaction:

- Users initiate the interaction by typing a text query in WhatsApp, leveraging the WhatsApp Business API for communication.

##### 2. Message Processing:

- The incoming message is processed to extract and isolate the specific finance-related query from the user’s input.

### 3. Symbol Identification:

- The system utilizes the OpenAI API to identify relevant financial ticker symbols (e.g., AAPL for Apple Inc.) within the user’s query.

### 4. Data Retrieval:

- Using the extracted ticker symbol, historical financial data (such as stock prices) is retrieved from Yahoo Finance (via the yfinance API).

### 5. Data Preprocessing:

- The retrieved financial data undergoes preprocessing to prepare it for input into the Large Language Model (LLM). This preprocessing includes prompt engineering to structure the data appropriately for the model input.

### 6. LLM Model Processing:

- The preprocessed data, along with the crafted prompt, is passed to the LLAMA-3 8B model, which is a custom Large Language Model designed specifically for finance-related insights.

### 7. Insight Generation:

- The LLAMA-3 8B model processes the input data and prompt to generate insights or recommendations tailored to the user’s query.

### 8. Response Delivery:

- Finally, the system responds to the user’s query by sending the generated insights or analysis results back via the WhatsApp Business API.

## 3.2 Detailed Overview

The system leverages the WhatsApp platform for user interaction, extracting relevant financial information using advanced NLP (Natural Language Processing) techniques and APIs. By integrating LLM technologies, such as the LLAMA-3 8B model, the system can understand complex finance queries and provide accurate and personalized responses in real-time.

This architecture allows for seamless integration of AI-driven capabilities with popular messaging platforms, enabling users to access finance-related insights conveniently through familiar communication channels.

## 4 Results and Analysis

The analysis of Finchat’s performance highlights notable user engagement but identifies areas requiring improvement. While the LLM generally delivers accurate responses, response delays due to computational constraints impact efficiency. Understanding user behavior patterns underscores the need to optimize chatbot responses for a better user experience. To enhance Finchat’s real-world impact, addressing response delays through increased computational resources and LLM refinement is crucial. These improvements will maximize Finchat’s effectiveness in providing users timely and accurate finance-related insights.

### 4.1 Performance Metrics

Table 1: Performance Metrics of Finchat

Metric	Findings
User Engagement	High engagement due to personalized finance recommendations.
Response Time	Average response time is 120 seconds but can be improved.
Accuracy	LLM delivers accurate insights with 85% accuracy.
Computational Resources	Current setup struggles with peak loads; upgrade needed.
User Feedback	Positive feedback on the usefulness of recommendations.

For detailed information on setting up and testing the FinChat project, including results and analysis, refer to the repository: <https://github.com/ragu8/FinChat>.

## 5 Conclusion

*Finchat* presents a robust finance assistant chatbot that leverages state-of-the-art language models and integrates seamlessly with WhatsApp to provide personalized finance insights efficiently. By combining advanced LLM technologies with real-time data retrieval, Finchat enhances user engagement and delivers accurate financial recommendations effortlessly. This project emphasizes the importance of intelligent chatbot solutions in meeting the evolving needs of users seeking quick and relevant financial information.