

Report - AI Financial Chatbot

Author: Akeem Asiru | Date: June 4, 2025

Introduction

The AI Financial Chatbot project is a generative AI-powered assistant designed to answer financial questions based on user-uploaded documents and real-time market data. The system leverages Retrieval-Augmented Generation (RAG) to combine context from proprietary files with external financial sources.

Objectives

- Automate financial document understanding and Q&A.
- Deliver accurate, real-time financial insights.
- Provide a scalable interface for users via web deployment.

Technologies Used

- LangChain for orchestrating RAG pipelines
- OpenAI's GPT-based LLM for language generation
- Pinecone as the vector store for document retrieval
- Streamlit for interactive frontend
- FastAPI to serve backend endpoints
- Financial market APIs for real-time data augmentation

System Architecture

The architecture includes four main components:

1. Document Upload & Parsing
2. Vectorization & Indexing
3. Question Processing with RAG

Report - AI Financial Chatbot

Author: Akeem Asiru | Date: June 4, 2025

4. Response Generation and Web Interface Delivery

Evaluation Metrics

The chatbot was evaluated based on internal test questions and user simulations. Metrics included:

- Accuracy: 91%
- Average Response Time: ~1.2 seconds
- Document Parsing Success Rate: 98%

Results & Impact

The chatbot demonstrated strong performance in answering complex financial queries, with over 90% accuracy in internal tests. It successfully processed various document types (e.g., PDFs, earnings reports) and offered context-aware responses using real-time market data.

Future Work & Limitations

While the chatbot performs well, it currently relies on static prompt engineering and single-language support. Future improvements may include multilingual capability, enhanced user feedback mechanisms, and broader market integration.

GitHub Repository

You can find the full source code and setup instructions at: <https://tinyurl.com/RAGFinanceBot>

Contact Information

For more information or collaboration, contact Akeem Asiru via LinkedIn or GitHub.