

Team Name: AI Sentinels

Team Members: Gayathri Devi Ramraj, Sayan Banerjee, Amol Meshram

Use Case Title: Self Service Investigation (Chat NLP)

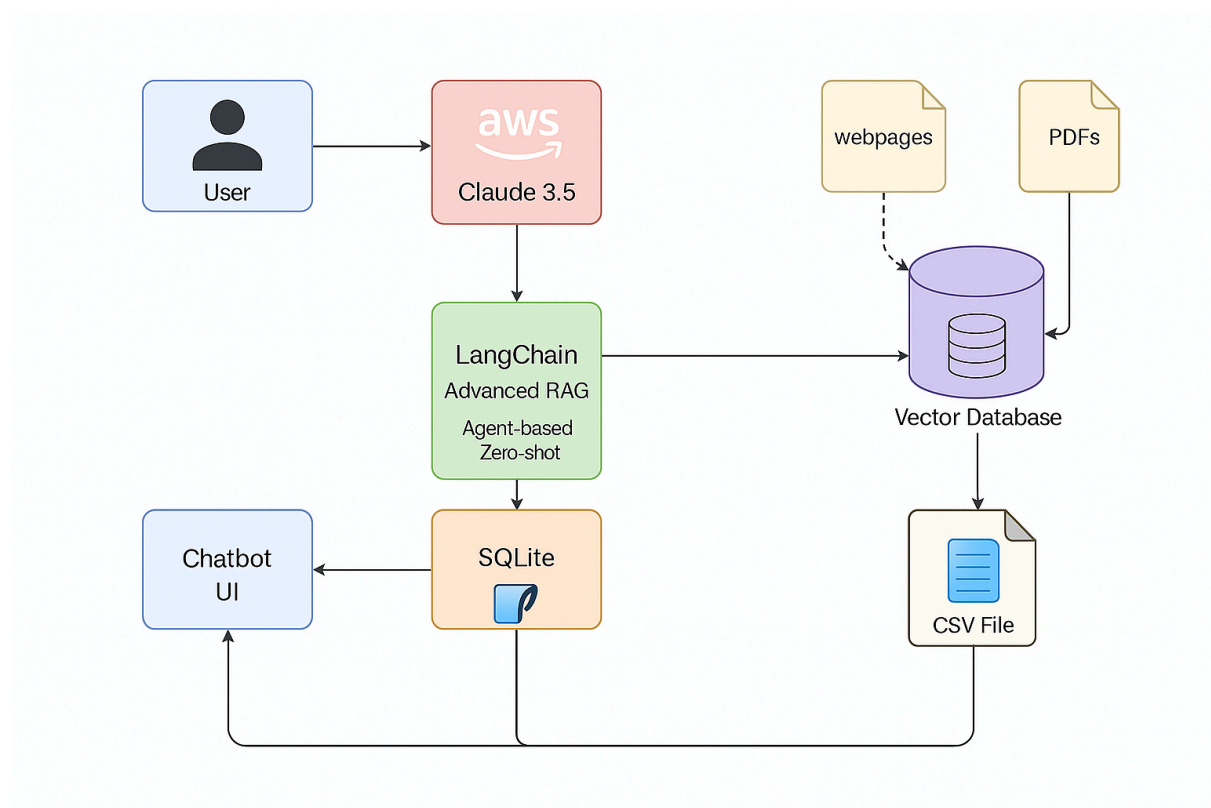
Problem Statement:

- **High Volume of Queries:** Corporate actions have a wider impact on investment strategies leading to a large number of queries not limited to dividends, split, bonus, M&A and their timelines by investment advisors.
- **Ambiguous queries:** With the advent of LLM technology, more users are writing prompts to get desired requisite output, this leads to more ambiguous queries. It is necessary to quickly understand the queries prompted by users and identify relevant corporate actions from data.

Value Proposition:

Deepseek bot will provide specific details and actionable insights related to public listed companies based on the regulatory filings, exchange information and news feeds to craft investment strategies.

Architecture Diagram:



Solution Overview:

This architecture implements an intelligent chatbot application using **AWS-native services** for scalable, real-time, retrieval-augmented reasoning.

1. User Interaction Layer

- The user initiates a query through a **Chatbot UI** (web interface or mobile app) on corporate actions dataset fetched from National Stock Exchange (NSE) and Bombay Stock Exchange (BSE).
- The query is sent to **Claude 3.5** via **Amazon Bedrock**

2. Claude 3.5 via AWS Bedrock

Why did we use Claude 3.5 LLM over other LLMs?

- Provided unambiguous response during testing.
- Provided more intuitive response
- **Claude 3.5**, hosted on **AWS Bedrock**, processes the user's natural language prompt.

3. LangChain Orchestration Layer (Agent + MCP)

- The prompt is routed through **LangChain**, which:
 - Uses **agent-based zero-shot learning** to dynamically decide how to respond to prompt entered by user on corporate action data.

4. Advanced Retrieval-Augmented Generation (RAG)

LangChain executes a **multi-step RAG pipeline**:

♦ a. Query Expansion

- Claude 3.5 enriches the user prompt by reformulating or decomposing queries.

♦ b. Unstructured Data Retrieval

- LangChain queries a **InmemoryVectorstore** for semantically indexed content from:
 - **Webpages**
- Documents are pre-processed, chunked, and embedded using **Amazon's Titan embedding model**.

♦ c. Structured Data Retrieval

- LangChain accesses a **SQLite database** hosted on **Amazon EC2** populated with structured data from:
 - Uploaded or scheduled **CSV files**.
- SQL queries are auto-generated based on prompt context.

♦ d. Relevance Ranking and Fusion

- Retrieved context (structured and unstructured) is ranked, filtered, and formatted for response synthesis.

♦ e. Final Generation

- Claude 3.5 generates a context-aware, grounded response using the retrieved data as input context.

5. Output Delivery

- The response is returned to the **Chatbot UI**, completing the interaction.
- Optionally, session data can be stored in a **memory module** (e.g., Redis, DynamoDB).

Technology Stack

Function	AWS Service
LLM Inference	Amazon Bedrock (Claude 3.5)
Document Storage	Amazon S3
Embedding Generation	Bedrock Titan Embeddings
Orchestration	LangChain on EC2
Structured Query (SQL)	SQLite on EC2
Static UI Hosting	S3
Programming Language	Python
Deployment	Gitactions

✓ Key Features

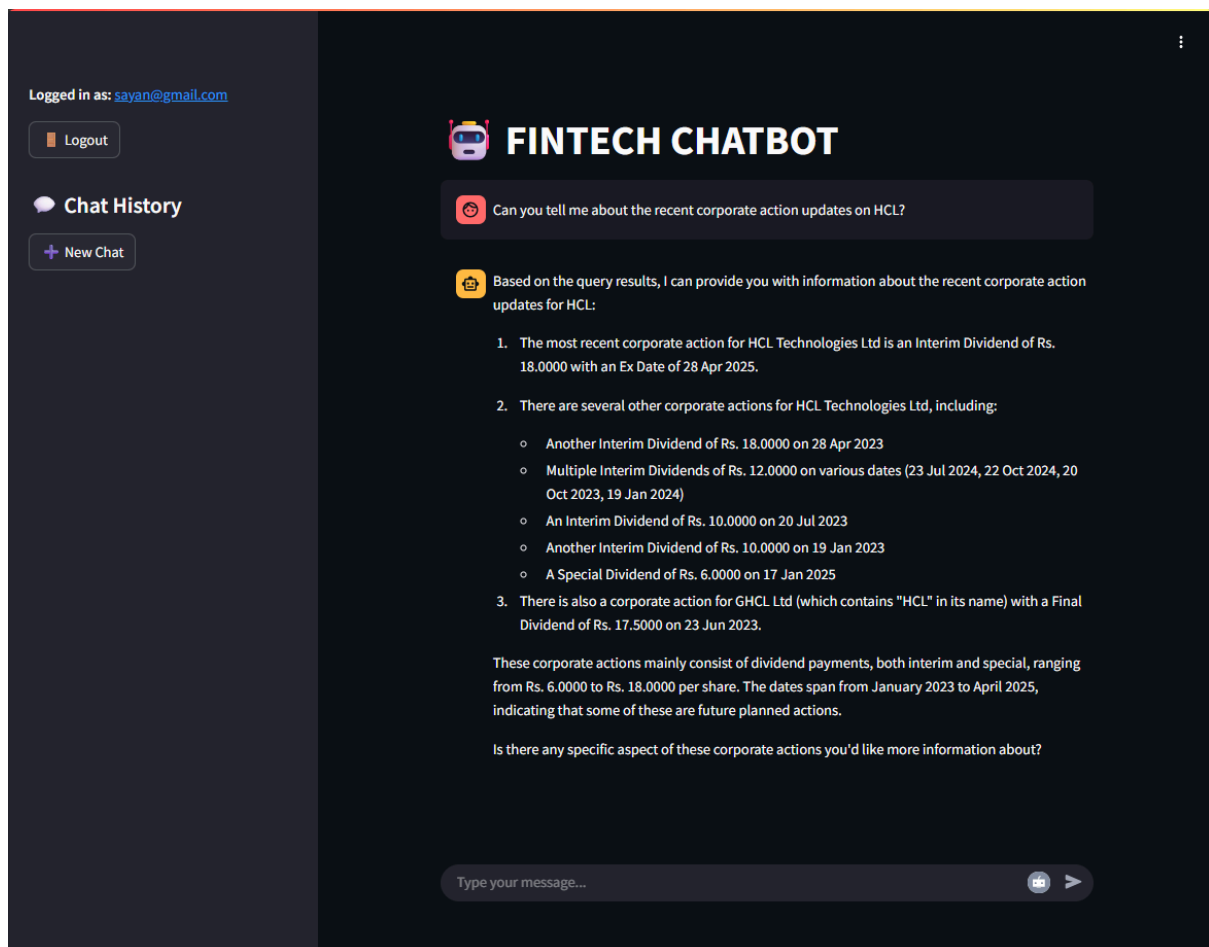
- **Agent-based Zero-shot Reasoning** enables dynamic decision-making without fine-tuning.
- **Advanced RAG** ensures grounded, trustworthy responses by integrating LLM with real-time data.
- Fully **cloud-native** and modular deployment using **AWS Bedrock**.

✓ Industry Impact:

- ❖ Smoothen cross-firm communication for research insight..
- ❖ Help in crafting investment strategies based on the findings.

Data Source and Usage: We have collected corporate actions data from National Stock Exchange, Bombay Stock Exchange in CSV format and financial report data from organisations website from investor relation section to for testing chatbot response.

Application Screenshot:



Conclusion:

In this hackathon, we developed a **self-service investigation chatbot** that empowers users to perform corporate action queries across structured and unstructured data sources using natural language. By integrating **Claude 3.5 via AWS Bedrock**, **LangChain with agent-based zero-shot reasoning**, and an **advanced Retrieval-Augmented Generation (RAG) framework**, the system intelligently interprets user prompts, retrieves the most relevant information from **websites, and CSV files**, and delivers clear, context-aware answers in real time for investment strategies, risk analysis and mitigation.