Application of GenAI on Financial Transcripts

Jayraj Dulange, Kushal Mehta & Manmohan Singh

Prof. Ravi Hegde

Department of Electrical & Computer Science Engineering

PROBLEM & MOTIVATION

Public companies regularly release a vast amount of **financial text** data through **earnings calls**, **investor presentations**, and **regulatory filings**. Analyzing this information manually can be time-consuming and overwhelming for financial analysts, especially when **accuracy** and **timeliness** are critical.

This project aims to explore how **Generative AI** can be applied to **financial transcripts** to streamline the analysis process. By automating initial insights and summarization, we hope to significantly reduce the time analysts spend on **data processing**, allowing them to **focus** on deeper strategic tasks—ultimately **improving** both the **speed and quality of financial decision-making**.

METHODOLOGY

Our pipeline begins by programmatically fetching conference call (concall) transcripts of publicly listed companies from the **Bombay Stock Exchange (BSE)**. Using a large language model (**Gemma3**), we segment and annotate the transcript based on speaker roles—distinguishing between **company officials**, **moderators**, **analysts**, and **investors**.

```
Chunk_id:7
Speaker: Gaurav Kumaar Dua
Role: Company Official (CEO, CFO, Executive)
Text: Hi. This is Gaurav. We are seeing a pretty poor footfall

Chunk_id:8
Speaker: Videesha Sheth
Role: Analyst
Text: So Gaurav, where I'm coming from is that during FY23 als
```

Figure 1. Example Annotated Transcripts

Once annotated, we generate **question-answer (Q&A)** pairs from **analyst queries** and **company official** responses.

```
--------------------------------------------------------------------------------
QA ID: concall_1_Q27
Questioner: Dixit Doshi
Question: So 15%, 16% EBITDA margin you feel is a sustainable margin?
Responders: R. Jayachandran
Answer: Yeah, it is not impossible, it can be achieved. We are not saying that it
--------------------------------------------------------------------------------
```

Figure 2. Example QA Pairs

These pairs are then analyzed using a fine-tuned **RoBERTa-based sentiment classification model**, originally trained on **financial news**, to determine the **sentiment** of each exchange. Sentiment is categorized into **four** combinations:

Analyst Question ↓ / Company Answer →	Positive Answer	Negative Answer
Positive Question	True Positive (TP)	False Negative (FN)
Negative Question	False Positive (FP)	True Negative (TN)

We also identify **false positives**—instances where companies provide **evasive**, **misleading**, or **overly optimistic responses** that do not align with the sentiment of the question. Finally, we **correlate** this sentiment data with **stock performance** metrics such as **monthly percentage change** and **trading volume**, as displayed in the results. This integrated analysis aims to reveal hidden patterns in **corporate communication** and its potential impact on **investor perception** and **market behavior**.

REFERNCES

[1] M. Hartmann, "Financial Sentiment Model", HuggingFace, 2021. [Online]. Available: https://huggingface.co/mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis

[2] DeepSeek R1 Model, Ollama Docs, https://ollama.com/library/deepseek

[3] LangChain Documentation, https://docs.langchain.com

[4] BSE Corporate Announcements, https://www.bseindia.com

[5] LLaMA 3.1 Model Card, Meta AI Research, 2024

PIPELINE

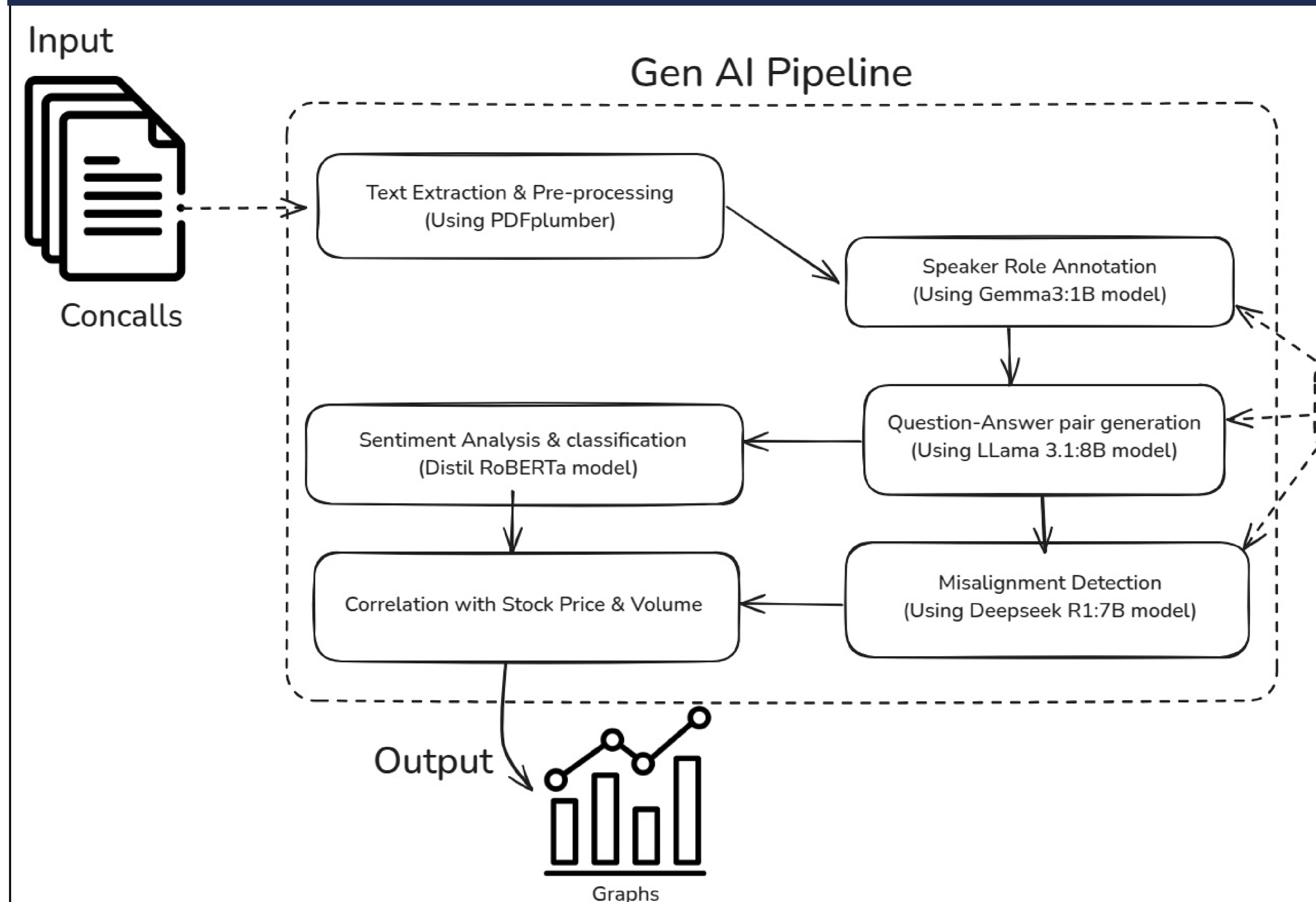

Figure 3. Pipeline

The pipeline employs **Gemma3 (1B)** and **LLaMA 3.1 (8B)** for its efficient **instruction-following** capabilities, making it suitable for speaker role annotation in financial transcripts.

For sentiment analysis, **DistilRoBERTa** fine-tuned on the **Financial PhraseBank** dataset is utilized, achieving an accuracy of **98.23%** in financial news **sentiment classification** .

DeepSeek (7B) is selected for its proficiency in structured **reasoning tasks**, such as identifying **evasive responses** in earnings calls.

This modular approach ensures each component is optimized for its specific task, balancing performance and efficiency for comprehensive analysis of financial communications.

```
"qa_id": "concall_1_Q6",
"concall": "concall_1",
"questioner": "Videesha Sheth",
"question": "My first question was on the commentary that you made in the press release regarding consumers down-trading to lower price c
"question_sentiment": "negative",
"responders": "Gaurav Kumaar Dua",
"answer": "Hi. This is Gaurav. We are seeing a pretty poor footfall happening across the country what we have noticed. And there is a mus
"answer_sentiment": "negative",
"analysis": {
  "question_addressed": "partially_addressed",
  "reasoning": "The answer from Gaurav Kumaar Dua partially addresses Videesha Sheth's question about consumer down-trading and its impac
}
```

Figure 4. QA Pair and their misalignment

RESULTS & INFERENCES

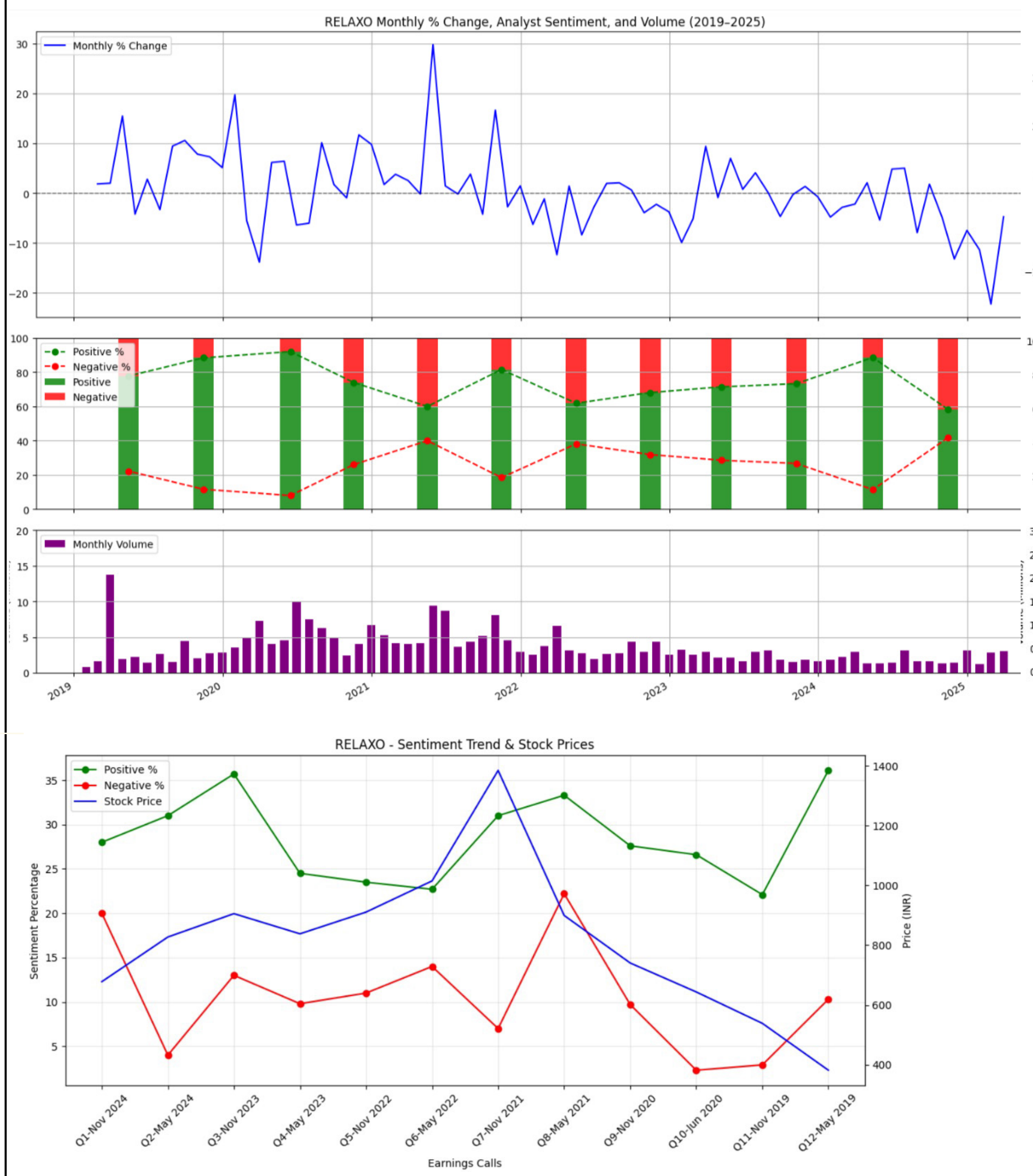

Figure 5. Results for RELAXO Footwears

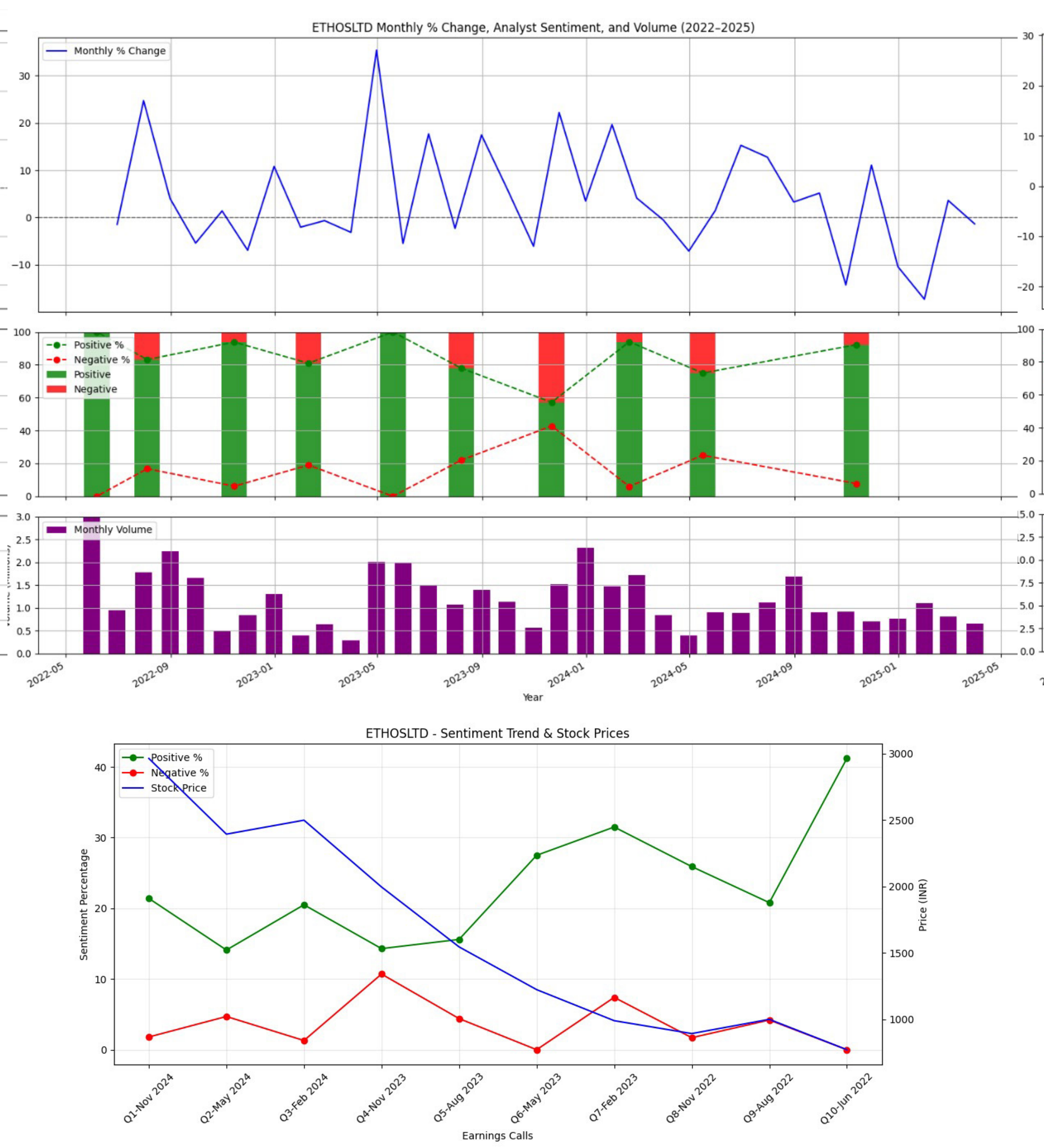

Figure 6. Results for Ethos

Figure 7. Results for Globus Spirits

The sentiment analysis methodology exhibits **partial correlation** with market movements, offering directional insights in select instances (e.g., GLOBUS' positive sentiment aligning with price surge), but fails to **generalize** across **all firms** and **time periods**. This underscores that while earnings call sentiments can serve as **useful indicators**, they are not **definitive predictors** of stock performance. Going forward, we aim to expand across **sectors**, and build **real-time tools** for analyst-ready insights.