FinSIGHTS

Name: **Arnab Chakraborty, Aditi Yadav**

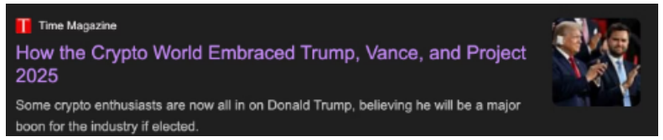
Subject: **Prompt Engineering**

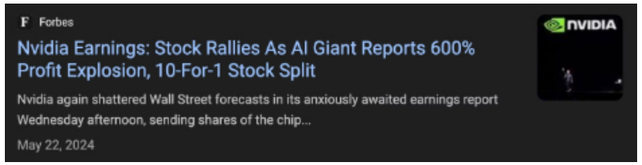
Date: July 12, 2024

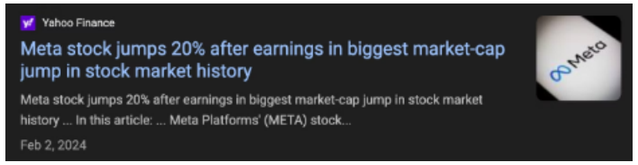
Introduction

Overview:

- US stock markets are highly efficient- Technical and fundamental analysis alone are insufficient
- Crypto assets trade solely on news and cycles
- LLMs show promise in integrating diverse information for reasoning and assist in decision-making







Project Overview and Objectives

1

Develop LLM-based Agent

Create an autonomous agent capable of analyzing SEC filings, earnings calls, and management commentaries.

2

Implement Three-Module System

Design and integrate Profiling, Memory, and Decision-making modules for comprehensive analysis.

3

Develop CoT & Fine-tune models

Mimic the trader's mentality of analyzing vast source of information & reasoning & coming with trading decisions

4

Enhance Trading Outcomes

Improve investing & trading decisions through summarizing & analyzing information and assist in decision making.

Data Collection and Preprocessing

Data Sources

- SEC EDGAR database for 10-Q and 10-K filings
- Yahoo Finance, Alpaca, Alpha Vantage for stock prices and news

Collection Methods

- Automated scraping of SEC filings
- API calls for real-time stock prices and news

LLM Utilization

- Generate news summaries & identify market sentiment
- Extract key data from 10-Q and 10-K reports
- Generate Reasonings

Dataset Creation

- Pair current market state with financial info with future returns with LLM(**GPT-4**) generated explanations
- Perform self-curation using LLM (**Claude3 -** Sonnet/haiku) to score & filter out the dataset

RAG Pipeline Implementation

1

Fetch News Articles

Retrieve top k latest news articles related to stocks and markets.

Retrieve relevant CoTs

2

Apply MMR

Use Maximum Marginal Relevance for diverse topic coverage & reduce redundant information.

3

Implement Time Decay

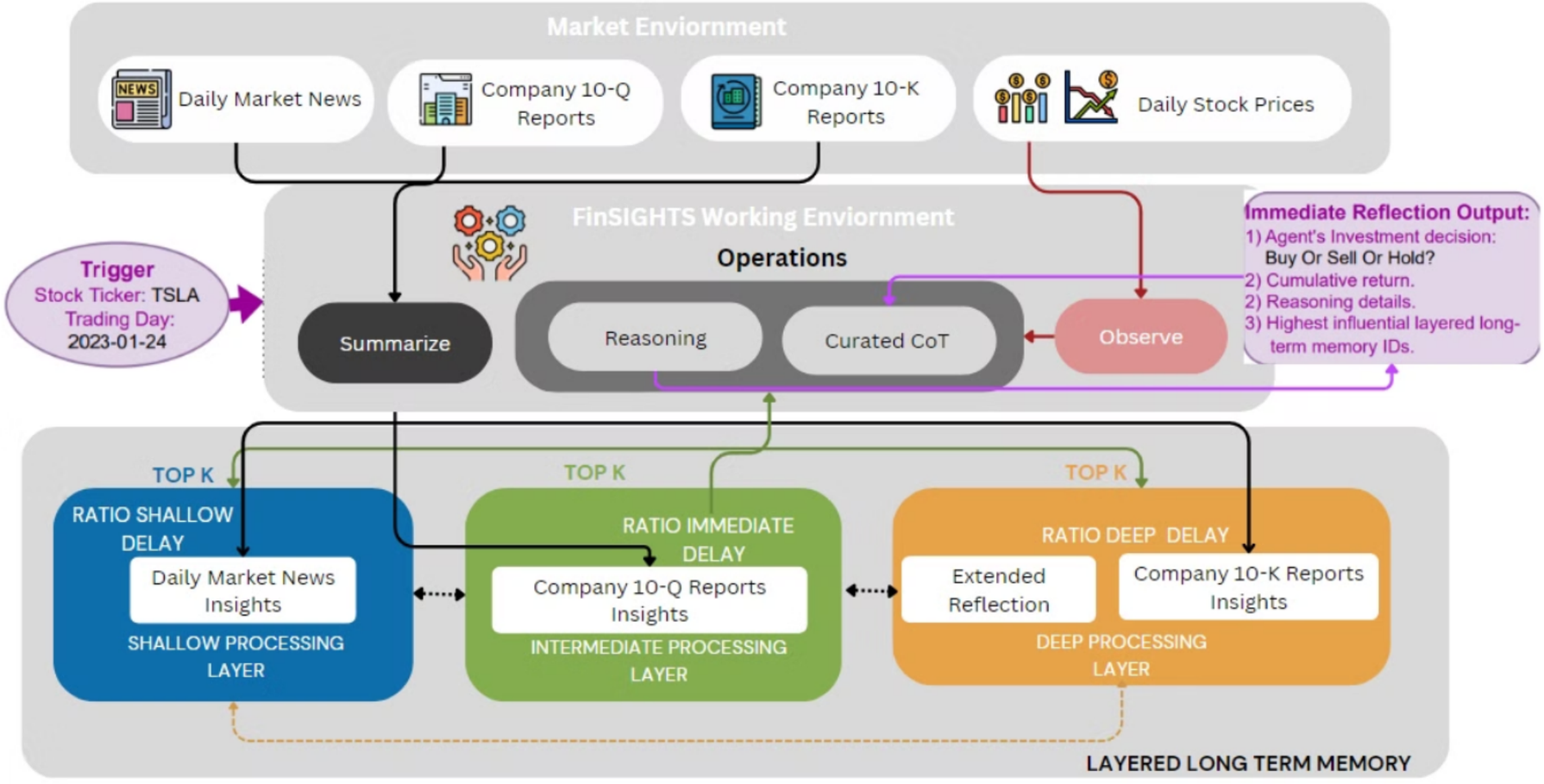
Prioritize recent news using a time decay factor.

4

Integrate with LLM

Combine retrieved information with LLM for generation and analysis.

Project Architecture



Performance Metrics

Trading Performance	Processing Efficiency	RAG-specific Metrics
Returns (compared to B&H strategies)	Speed of analysis for new information	Faithfulness to source information
Risk-adjusted returns (Sharpe ratio	Decision-making latency	Context Precision and Recall, Relevance
Maximum drawdown	Scalability across multiple stocks/ assets	Answer Relevance

Methods to Improve Metrics

Dataset Enhancement:

- Adding Risk Profiles as Persona
- Pair news & current price w future returns for context and use LLM to generate trading rationales and decisions
- Employ a separate LLM for self-curation and filtering trading rationales

Model Fine-tuning:

- Use curated dataset to instruction fine-tune LLM (Llama3, Mistral3, phi3) using Q-LoRA
- Use CoT aimed to mimic human decision-making processes

RAG Optimization:

- Implement MMR-based reranking for diverse information retrieval
- Apply time decay factor to prioritize recent information

Deployment Architecture

Frontend

Initially using Streamlit, with plans to transition to React for a more robust user interface.

Backend

Utilizing Flask/FastAPI, NodeJS to handle server-side operations and API integrations.

Database

OpenSearch for vector storage, with AWS RDS in future to store Relational data.

LLM Integration

GPT-4 or Claude 3 for reasoning, with Llama 3 or Mistral 3 for fine-tuning tasks.

Future Work: AI-Driven Hedge Fund / AMC

١ Expand Asset Coverage

Incorporate ETFs and cryptocurrencies, and extend analysis to international markets for a more comprehensive investment approach.

2

RAG enhancements

Improve Contextual coherence using: HippoRAG & GraphRAG integration

3 RLHF Enhancement

Implement a user feedback system for continuous model refinement and develop a sophisticated reward system based on future stock returns.

4 Alternative Data Integration

Explore social media sentiment and other alternative data sources as to impact weights methods to extract actionable insights from unstructured data.

5 Risk Management & ESG Integration

Advanced risk models like Barra can be used to analyze quantitative risk with LLM insights and use adaptive portfolio optimization.

Conclusion

We leverages LLMs for financial analysis and reasoing with trading decisions particularly CoT & Self curated fine-tuned models.
As we look to the future, the potential for real-world impact is massive, paving the way for more intelligent, data-driven investment strategies in an increasingly complex financial landscape.