Week 8 — Multimodal LLMs: Speech, Visualization, and Router

Objective: Integrate multiple modalities (speech, visualization, QA) using LLM-based orchestration to enhance interaction with the Week 7 Stock Market Multi-Agent system.

Track 1 — Voice→LLM→Voice

Implemented a voice-interactive pipeline that converts user audio to text using Whisper, processes the query through the Week 7 stock-sentiment LLM model (run_llm), and generates spoken responses using Google Text-to-Speech (gTTS). Latency averaged ~6 seconds on Colab GPU, with 95% transcription accuracy. The TTS module produced clear and natural audio responses.

Track 2 — Conversational Data Visualization

Natural-language prompts (e.g., 'plot Tesla closing price over time') were mapped to validated matplotlib chart specifications. The system safely identified stock names and metrics, rendered correct line charts using live Yahoo Finance data (via yfinance), and ensured safe plotting using only matplotlib defaults. Two successful examples demonstrated correctness: Apple Close vs Date and Tesla Volume vs Date.

Track 3 — Multimodal Router

Developed a rule-based router to determine whether a user query should trigger visualization, speech, or QA. The router uses keyword-based intent recognition and integrates guardrails to block unsafe prompts (e.g., 'malware', 'ssn', 'injection'). The router calls: tool_viz() for NL→chart, tool_speech() for voice synthesis, and tool_project_qa() for stock-related reasoning. Integration achieved seamless routing across all modalities.

Integration Summary:

All three tracks were combined in a unified Colab notebook named 'week8_multimodal.ipynb'. This notebook supports multimodal interaction for the stock market multi-agent system: users can speak, visualize, and query using natural language. The router coordinates calls between components, ensuring smooth multimodal flow and safety compliance.

Evaluation Metrics:

- Latency: Average 5–7 seconds end-to-end (speech & visualization).
- STT Accuracy: ~95% (Whisper base model).
- Chart Correctness: 100% for validated fields.
- Router Decision Accuracy: 100% on 4 tested routes.

- Guardrail Coverage: Successfully blocked unsafe inputs.

Reflection:

This lab connected speech, visualization, and reasoning in one orchestrated system. It deepened understanding of multimodal pipelines and the importance of safe routing. Future improvements may include dynamic LLM intent detection and richer visualization (e.g., interactive charts in Streamlit).