

Week 8 Final Report — Multimodal Florida Hurricane Irma Pipeline

Student: Adrija Ghosh | Course: COMP-SCI 5597 (Capstone/Research Practicum)

Project Theme: Voice-Interactive & Multimodal AI for Disaster Data Understanding (Hurricane Irma, 2017)

1. Overview

This week's deliverables extend the Week 6–7 QA/RAG model into a voice-interactive multimodal system capable of handling three input modes:

Speech → Text → LLM → Speech (Track A), Natural-Language → Validated Visualization (Track B), and Routing across Speech + Visualization + QA tools (Track C).

2. Track A — Voice-Interactive LLM (Speech → LLM → Speech)

Objective: Demonstrate end-to-end speech processing and reasoning using OpenAI Whisper for STT, local Irma QA for reasoning, and gTTS for speech synthesis.

Workflow: Audio → Transcription → Reasoning → TTS → Latency logging.

Results: Total roundtrip latency ≈ 2.1 s, WER ≈ 0.05 , accuracy ≈ 0.82 .

Example: "When did Hurricane Irma make landfall in Florida?" → Landfall at Cudjoe Key (Cat 4) and Marco Island (Cat 3) on Sept 10, 2017.

3. Track B — Conversational Data Visualization

Objective: Translate natural-language queries into validated matplotlib charts using Irma's Florida dataset (Sept 5–12, 2017).

Functions: `nl_to_plot_spec()` maps keywords (wind, pressure, rainfall, category) to safe specs; matplotlib renders line/bar/scatter plots.

Queries: "Show wind speed trend over time" and "Compare rainfall and pressure across days."

Evaluation: 90%+ correctness; latency ≈ 1.4 s; instant visual feedback; no unsafe code execution.

4. Track C — Multimodal Router (Speech + Viz + QA)

Objective: Integrate all tools with routing logic, guardrails, and trace logging.

Router Keywords: "plot/chart/trend" → Viz; "record/audio" → Speech; "what/when/how" → QA.

Guardrails: Blocklist (self-harm, illegal, hate), input limit (2000 chars).

Results:

- "show trend for wind speed" → viz (0.63 s)

- "plot rainfall vs pressure as bar" → viz (1.32 s)

- “record: ./sample.wav” → speech (0.01 s)
- “When did Irma make landfall?” → qa (0.00 s)
- “explain damages and costs” → qa (0.00 s)

5. Ablation Study Summary

Variant, Latency(s), Accuracy, Visualization Quality, Speech Accuracy

Baseline_TextOnly, 1.21, 0.82, N/A, N/A

With_Speech, 1.63, 0.82, N/A, 0.93

With_Visualization, 1.38, 0.85, Good, N/A

Full_Multimodal, 1.92, 0.86, Very Good, 0.91

6. Reproducibility and Environment

Environment: openai-whisper, gTTS, SpeechRecognition, pandas, matplotlib, streamlit.

Files: week8_florida_voice_trackA.ipynb, week8_florida_conversational_viz.ipynb, week8_florida_multimodal_router.ipynb, ablation_results_week8.csv.

Instructions: Run each notebook in Colab, upload /content/irma.wav, view saved outputs in /audio_outputs/, /visual_outputs/, and logs/router_traces.jsonl.

7. Discussion and Future Work

- Add real-time microphone + TTS playback in Streamlit.
- Replace rule-based router with LLM intent classifier.
- Connect NOAA APIs for forecast visualization.
- Introduce reinforcement-based optimization for latency/accuracy trade-offs.

8. Conclusion

The Week 8 implementation satisfies all Tracks (A, B, C) using the Florida Hurricane Irma dataset. It demonstrates a full multimodal AI system with speech, visualization, and reasoning integration. The pipeline provides an accessible educational and analytic framework for disaster data exploration.