

# Agentic Voice-to-Voice AI Assistant for Product Discovery



**Prepared By:** Soham Mandal, Bruna Medeiros, John Melel,  
Bradley Stoller, Aida Sarinzhipova, Jerry Gong

**Date:** December 12, 2025

# Agenda

01

Architecture

02

Product Demo

03

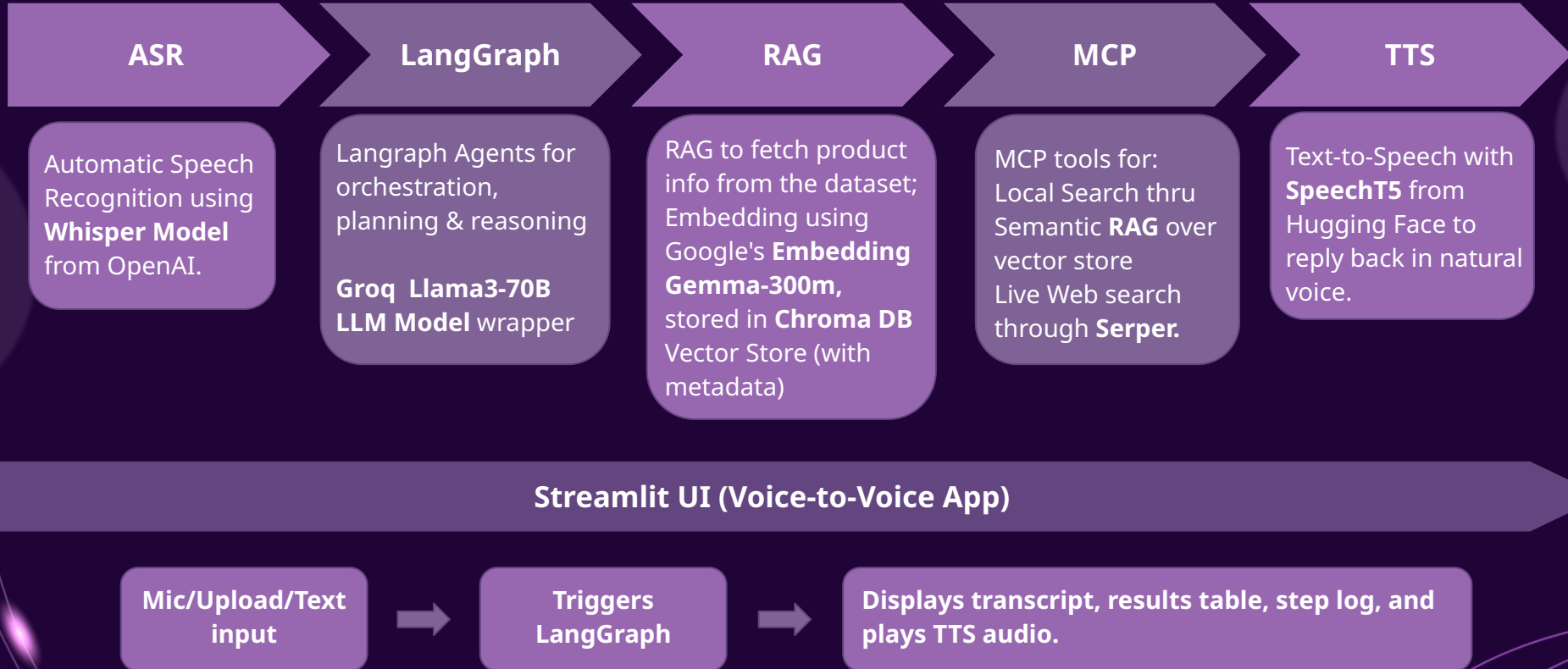
Limitations



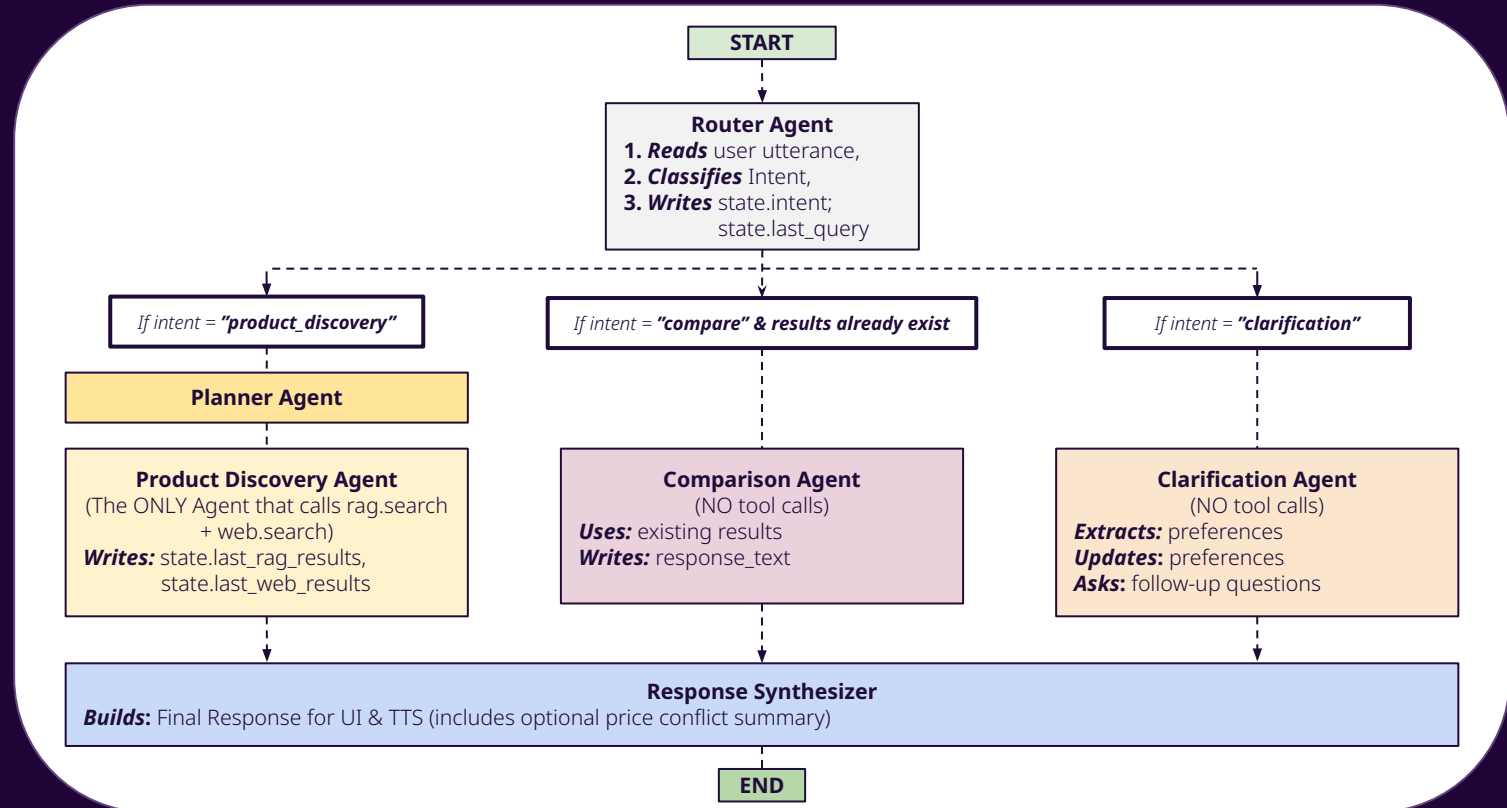
01

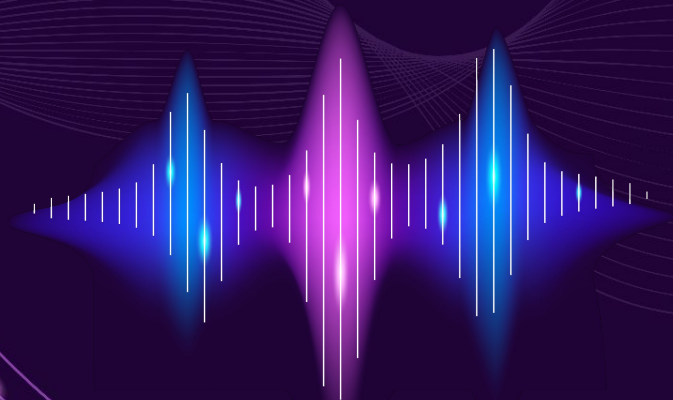
Architecture

# Project Architecture



# Project Architecture

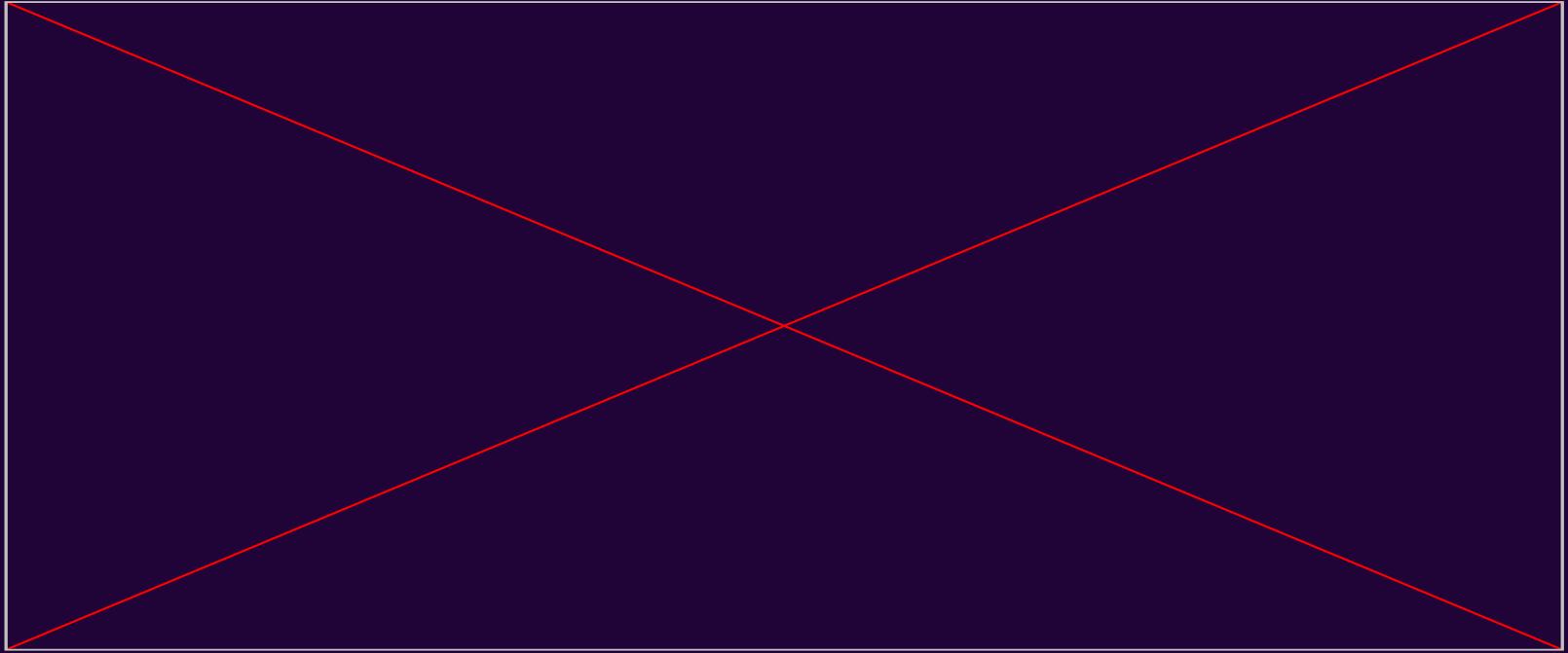




02

Product Demo

# Product Demonstration





03

Limitations



# Limitations (Known Tradeoffs)

Web Search

**Live web data** adds real-world market context **without** overwriting catalog prices, as real-time price matching is intentionally excluded to maintain modularity.

**Intent classification** relies on LLM reasoning combined with heuristic rules rather than supervised models, introducing occasional misclassification for short or ambiguous queries.

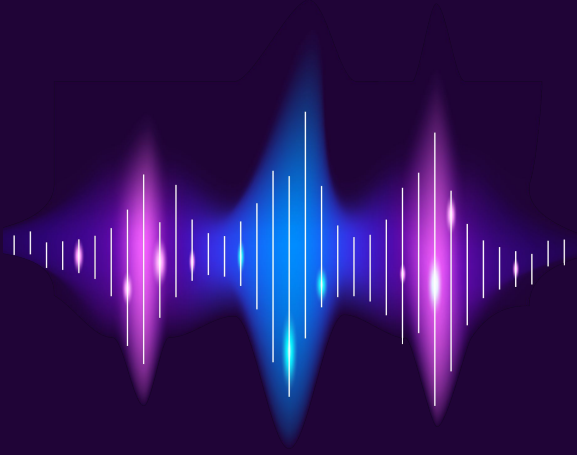
Intent

Embeddings

**Fixed embeddings** may lose semantic alignment over time, requiring re-embedding to stay accurate in production systems.

**Latency** increases due to multi-agent orchestration, speech processing, and MCP tool calls, which would require optimization for real-world deployment.

Latency



# Thank You!

Questions?

[GitHub Link](#)