



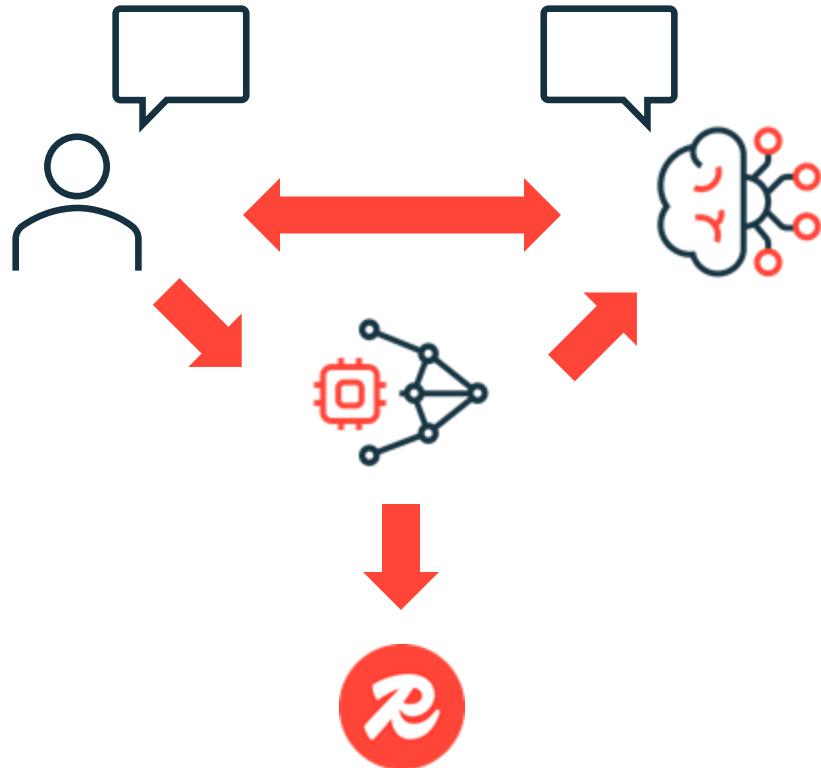
Podcast Chatbot with Agent Memory & Azure

Guy Royse

December 4th, 2025

What is it?

- It's a chatbot
- It remembers conversations
- It learns facts from those conversations
- It uses those facts to provide better recommendations
- Stores it all in Redis



Agent Memory Server

- Stores short-term conversation history
- Summarizes older conversation history
- Extracts long-term memories for use in future conversations
- Generates prompts for your LLM to use



I'm interested in podcasts about science



Great choice! Here are some podcasts about science.



Tell me more about Startalk



User expressed interest in podcasts about science and asked more about Startalk.



Likes podcasts about science.

Objectives

Azure

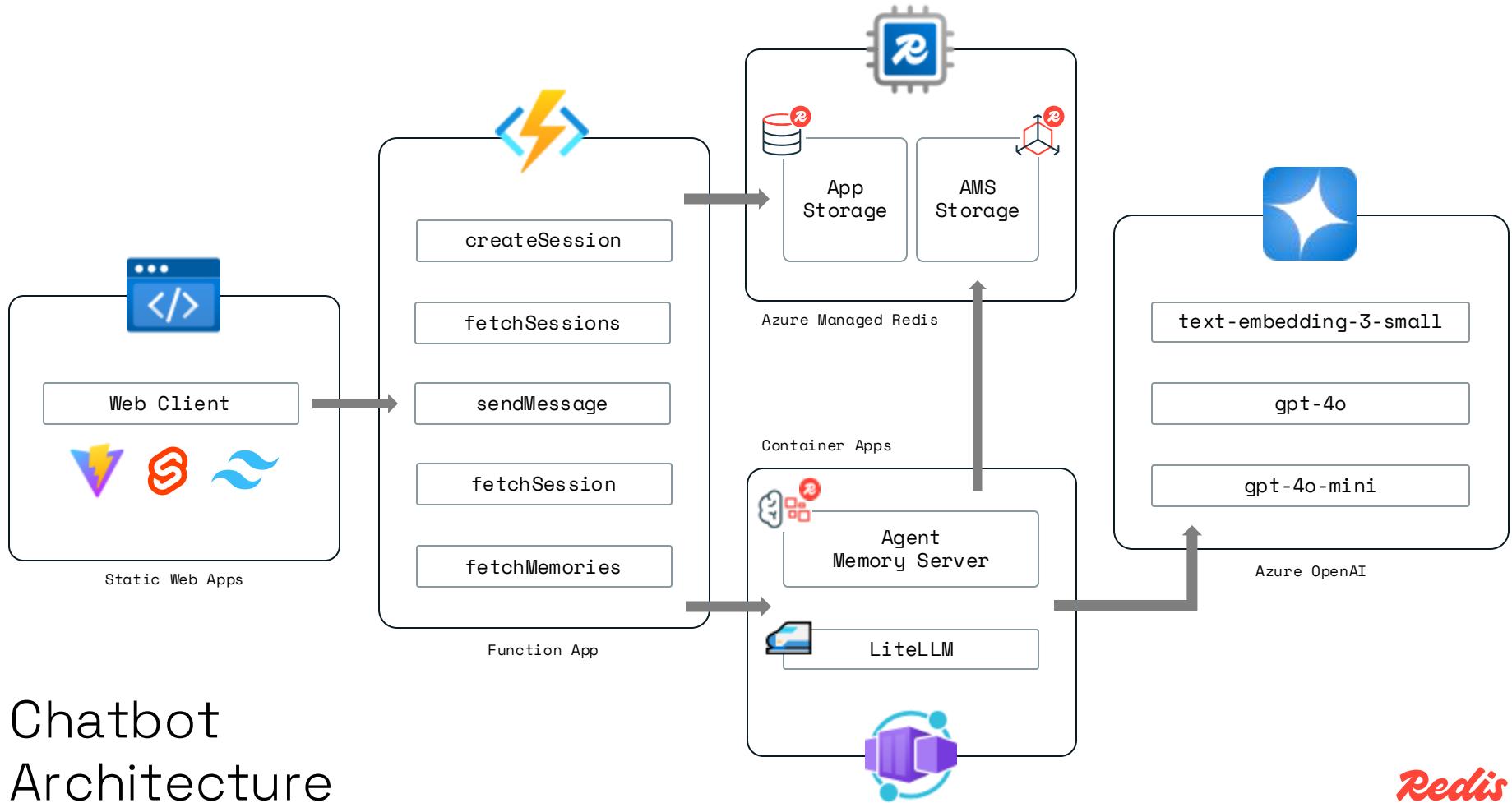
+

Agent Memory
Service

+

Azure Managed
Redis





Code & Demo

What was learned along the way

- Agent Memory Service is not optimized for Azure
 - Azure wants to use EntralID
 - Azure wants to use Azure OpenAI
- LiteLLM is a useful layer of abstraction over APIs
- Even though the system prompt guided the conversation, I could still get the LLM to talk about other things

What we could add

- Some sort of tool to search current info about podcasts
- Semantic routing to keep the agent on task and to
- LangCache for commonly asked questions



Thank you.