

Runtime Architects — Team Meeting Minutes

Date: 2 July 2025

Time: 11:04 AM – 12:11 PM

Location: Online Call

Attendees:

- Pablo Periañez Cabrero
 - Firose Shafin
 - Aditya Bhargav Akella
 - Nithyakamal Ilamurugu (Kamal)
 - Saeed Misaghian (Microsoft, External)
 - Neenu Vincent (Microsoft, External)
-

1. Project Progress Updates

Pablo:

- Implemented streaming functionality via SSE for real-time backend-to-frontend updates.
- Updated frontend to support streaming and added a new health endpoint for diagnostics.
- Demonstrated UI improvements including event logging and health checks.
- Discussed Azure deployment strategy — Azure Container Apps for backend; React app for frontend.
- Received feedback to modularize event handler for easier agent integration.
- Assigned investigation into secure user authentication and login (e.g., passkeys) and data storage (Cosmos DB suggested).

Action:

Modularize event handler and implement authentication and logging

Explore Cosmos DB integration

Investigate fallback messages in UI for failed requests

2. Agent Evaluation and Prompt Engineering

Firose:

- Presented methodology for agent evaluation using metrics: relevance, coherence, similarity, F1 score.
- Showcased dramatic improvements post-prompt optimization (F1 ↑ to 0.92).
- Plan to apply same framework to evaluate Kamal's and Aditya's agents.
- Tasked with automating prompt engineering + evaluation pipeline for future agents.
- Assigned exploration of agent observability and feedback systems.

Action:

Work on automated evaluation pipeline with prompt engineering loop

Investigate tracking and observability tools (e.g., Azure AI Foundry, LLM-as-a-Judge)

3. Backend and Agent Development

Kamal:

- Fully migrated multi-agent setup to Azure OpenAI using `autogen.AssistantAgent`.
- Implemented Python tool executor with local execution.
- Coordinated with Pablo (UI) and Firose (evaluation) for tighter integration.
- Assigned exploration of sandboxed code execution with virtual environments and Docker.
- Open question: how to isolate/refresh environments between sessions to prevent library conflicts.

Action:

Refactor backend to support modular execution environments

Coordinate agent integration with UI and evaluation

Investigate secure and resettable execution environments for code tools

4. Performance Optimization and Framework Assessment

Aditya:

- Policy agent operational using web-scraped data.
- Diagnosed performance bottlenecks (latency due to long conversations, tool invocation, async contention).
- Proposed database caching strategy to replace live API queries.
- Investigating switch from GroupChat to more performant alternatives (e.g., ConcurrentAgentChat).
- Planning benchmarking for different agent orchestration models and tool integration setups.

Action:

Benchmark autogen configurations (latency, reliability)

Prepare caching/data retrieval layer for agents

Evaluate trade-offs between agentchat and agentboard designs

5. System Architecture and Authentication

General Discussion:

- Agreed on the importance of integrating role-based access control (e.g., hide health diagnostics for non-admins).
- Passkey authentication discussed as a secure alternative.
- Cosmos DB favored for fast, scalable storage of conversation logs and user metadata.
- Need to define schema for user info, logs, and telemetry tracking.

Action:

Pablo to research and implement secure login system

Define DB schema for user data, sessions, and logs

6. Next Steps

- Pablo to modularize event handling and implement authentication
- Firose to build automated evaluation loop and observability dashboard
- Kamal to integrate agents and investigate secure sandboxing
- Aditya to conduct benchmarking and optimize orchestration
- Shared goals: improve latency, agent consistency, secure access, and structured data storage

Next Meeting:

To be scheduled (Aditya and Kamal to sync with Saeed for technical follow-up)

Meeting Adjourned

Prepared by:

Pablo Periañez Cabrero

2 July 2025