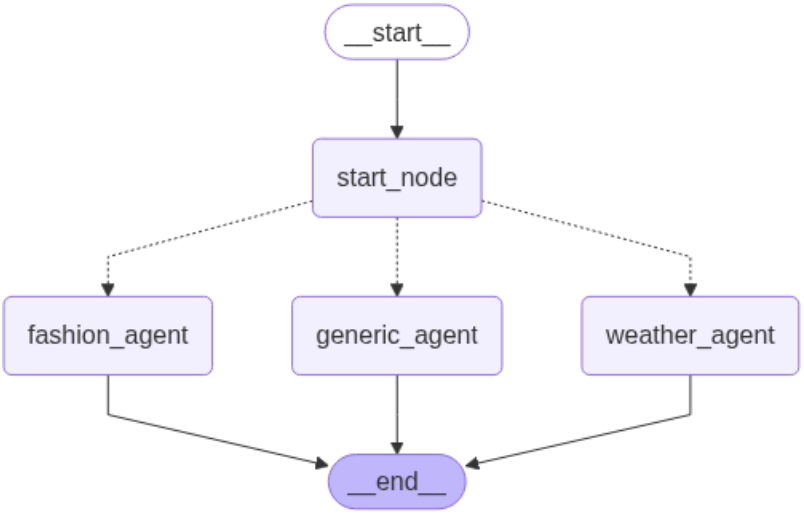
Project Report: Task 1

1. Introduction

The notebook showcases the design of a modular AI assistant framework that intelligently routes user queries to specialized agents, such as a weather agent, or falls back to a generic conversational agent when no specific context is detected.

NOTE: More details of each approach and experimentations are written in the Jupyter notebook.

2. Approach and Architecture



2.1 Base (Generic) Agent

- Acts as a fallback when no specific topic is identified.

- Maintains context and ensures a smooth conversation flow.

2.2 Weather Agent

- Tailored to respond to weather-related queries.

- Can use the API to give real time responses.

- Uses system prompts like "You are a weather assistant..." to guide responses.

2.3 Fashion Agent

- Answers fashion related queries

- Can use the API to give real time responses.

2.4 Router

- Evaluates the human prompt and directs it to the appropriate agent

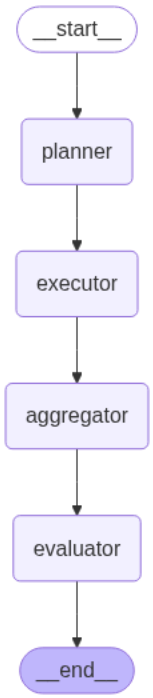
2.5 State Management

- It retains all the prior queries and responses as additional context for generating the next response and maintaining a conversational flow in the session.

- Limitations: If the history becomes too big, the input token limit of the LLM gets exceeded and leads to errors.

- Remedy: Maintain a summary of all the responses instead of the actual responses. Create vector embedding of the responses.

3. Experimentations



3.1 Attempt at Level 4(present in weather_mind_multi_agent.ipynb)

-Planner agent: Analyzes the human input and identifies the specialist agent(s) required to respond. Makes decisions on how to dynamically orchestrate multiple agents and tools to provide the best possible output.

-Executor: It could be one or more agent(s) collaborating to synthesize the output.

-Aggregator: Generates a cohesive response by compiling responses from all the relevant agents.

-Evaluator: Evaluates the response with respect to the prompt and can trigger a re-run of the workflow.

NOTE: An attempt was made, however, I was not able to fully refine this.

4. Results & Insights

The system can reliably redirect to the appropriate agent based on the user prompt.

It is able to have contextual conversations, i.e. it is able to remember previous messages and give relevant responses

Appendix

- Full code available in ` weather_mind.ipynb` notebook.

- Experimental code in` weather_mind_multi_agent.ipynb`