

Report for Agentic AI Project

Names: M. Shehryar , Shahbaz Hussain

Team: Rusty

1. Overview:

Our agentic AI project is all about efficiently calling different tools based on user's query. There are 4 different types of tools which are connected with agent.

1. Calculator
2. Web Search
3. Document Retrieval
4. Math Solver

Our agent is based on LLM which is like the brain of this project. It does decision making routing and returning the output. Our LLM is GPT-5-mini which is the latest version of Open AI models. It is a little bit cut down version of GPT-5.

2. Tools Used:

We mostly used tools of Open AI. Overall we have tried many tools like first we use **duck duck go** for web search, but it was not giving correct answers. Then we moved to GPT-4o-mini for web search.

Also like that we use several different models for solving math problems. First we use **Gemini-1.5-flash** but it did not work well. Then we moved to **Qwen 3b**, but it also did not performs well. Then finally we moved to **GPT-5-mini**.

3. Challenges:

First challenge was the correct routing and decision making of agent for different tools.

Second was gsm8k was giving 0 percent or 3-4 percent accuracy of many different models. It means the models we used were totally collapsing against it.

Third challenges was LAMA test which test our patient. Because we have to manage two things for it.

Manage the controller prompt in such a way that the agent would mostly answer the LAMA questions by itself not from web. But also along with that it should use web search if I gave it a certain query. So these were two opposite things to happen at once.

4. Results:

The results are quite fine. We have achieved 96 percent accuracy on gsm8k and 93 percent accuracy on LAMA dataset.

We have ran just 100 examples of both due to lack of time and api tokens.

5. Limitations:

It is obvious that there is certain gap of improvement in our project. In several things like agent. Web search tools etc.

6. Team Contribution:

We both work together all along the project. My teammate Shahbaz do the searching work and I do the implementation work.