Question-Answering Agent

Problem statement

LLM-powered chatbots are powerful, but they can be even more powerful when combined with external sources of data. Your task is to build an LLM-powered chatbot that interacts with the user on open-ended questions, and then evaluate it for quality.

Your task

1. Build a simple LLM-based general-purpose conversational question-answering chatbot that makes use of tool calling (function calling). You can choose the underlying LLM, and you can choose the tools to integrate and the provider. You should have at least two tools, but you can choose to do more.

2. Design and conduct an evaluation of your chatbot. Choose a metric, obtain a small dataset, and write appropriate evaluation code.

3. Build a web-based dashboard to display and analyze the evaluation, as well as to interact with the chatbot live.

General instructions

- You can choose any tech stack for this task, as long as the project is runnable in a normal Linux machine. It is recommended to use either Python or JavaScript/TypeScript.

- Your submission should include all code (excluding publicly-available dependencies) and should include a brief README describing how to run, how the code is structured, and what it does.

- API keys for LLMs (OpenAI, Fireworks.AI, Anthropic) will be provided upon request. We can also provide API keys for 3rd party services if needed.

- Vibe-coding via Cursor or CoPilot is encouraged, especially for the web dashboard part, but you should review all code and be prepared to discuss it.