Intoduction

- **What is the Emcure Chatbot?**
  - Emcure Pharmaceuticals, a leading Indian pharmaceutical company, is owned by Namita Thapar, who is also a prominent investor on "Shark Tank India.
  - An Agentic chatbot designed to assist users with medical and pharmaceutical inquiries.
  - Link : https://www.emcure.com/

- **Purpose**:
  - Provide accurate and timely information about Emcure products and services.
  - Enhance customer experience through text-based interactions.

- **Key Features**:
  - Text input and responses.
  - Context-aware responses using RAG.
  - Integration with a knowledge base.

Technology Used

- **Python**: A versatile programming language for backend development and data processing.

- **LangGraph**: A tool for representing and interacting with complex data structures and graphs.

- **LangChain**: A framework for building applications using large language models (LLMs) with advanced natural language processing.

- **Flask**: A lightweight web framework for building web applications, providing routing, templating, and handling HTTP requests.

- **HTML**: Standard markup language for creating structured web pages and web applications.

- **CSS**: Style sheet language used for describing the presentation of a web page, including layout, colors, and fonts.

- **JavaScript**: A dynamic scripting language for creating interactive and responsive web pages.

- **MySQL:** SQL Database use to store all user and bot conversation in Table

- **VectorDatabase :** I used Chroma vector database to store website infromation in numerical form.

Why This Technology Used

LangChain

- **Purpose:** Agentic chatbot, workflows, LangGraph
- **Pros:** Easy complex workflows, LangGraph integration, rich ecosystem
- **Cons:** Config for advanced use, steep learning

Google Generative AI Embeddings

- **Purpose:** Fast, open-source embeddings
- **Pros:** High-speed, open-source, efficient
- **Cons:** Less customization, limited support

Gemini LLM ("gemini-1.5-flash")

- **Purpose:** Rapid chatbot responses
- **Pros:** Fast, open-source, real-time
- **Cons:** Nuanced queries, limited docs

Note : I Want to use Ollama (llama3.2.8b) model but it was take too much time to give response so I use gemini model but this model is not so good in Rag Context Generation. I also wan to use Groq model but it Has token limit so decide to use this model.
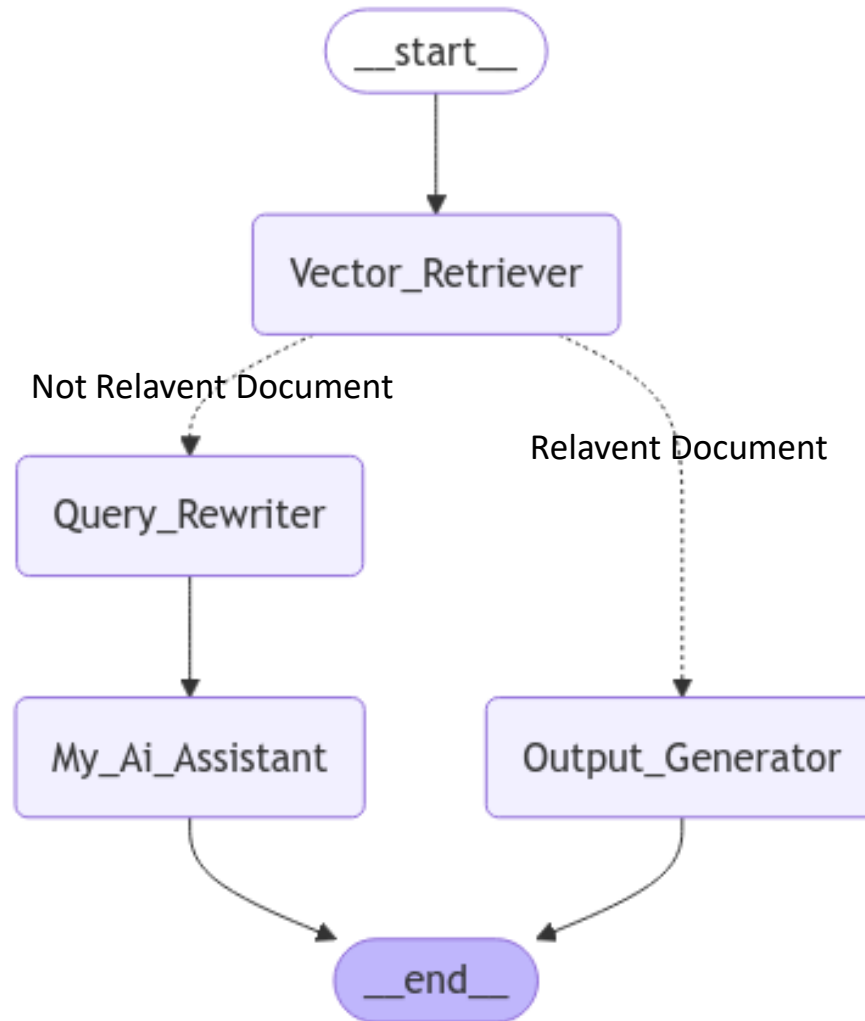
Chatbot Interface
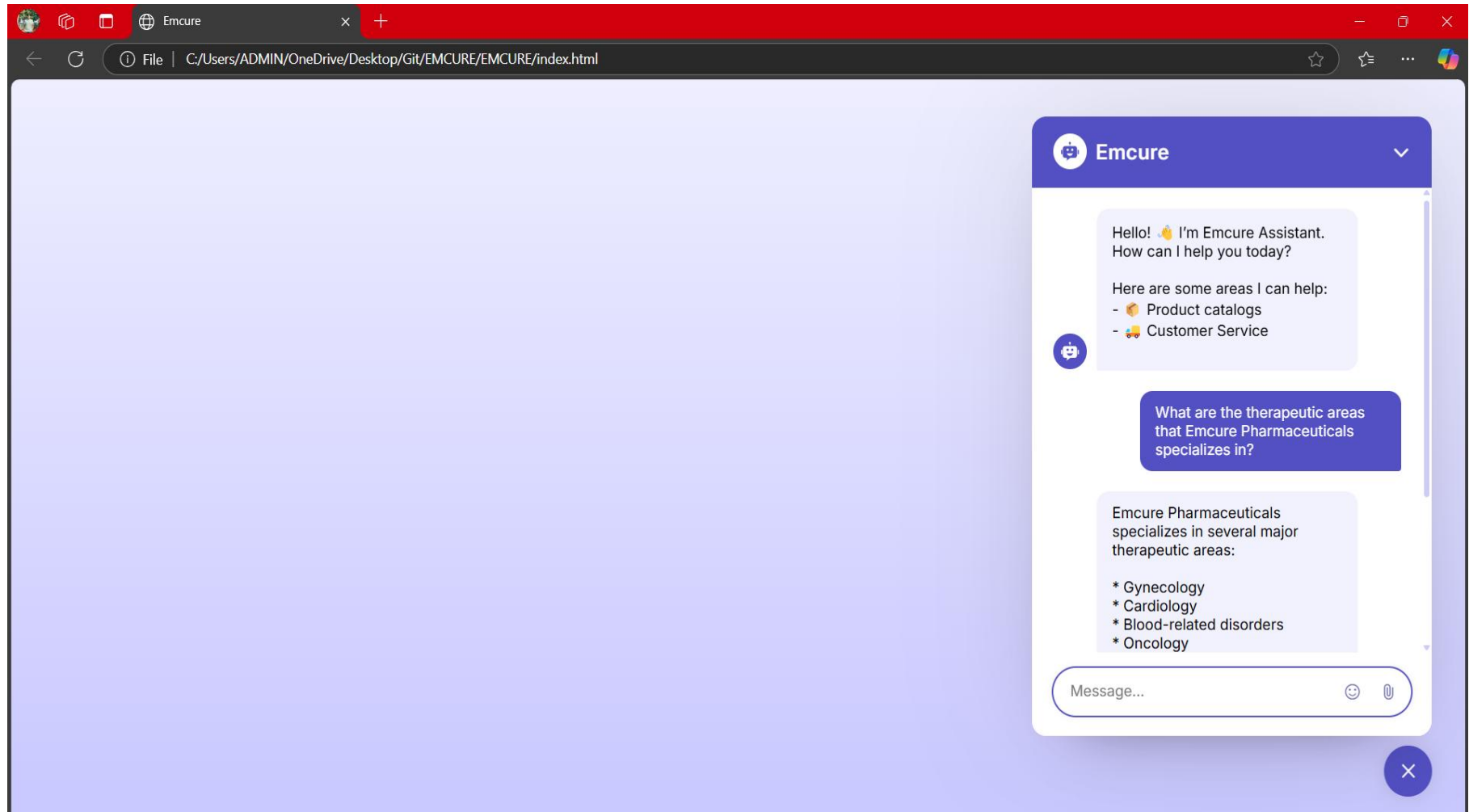
Database Design

COLUMN NAME	DATA TYPE	DESCRIPTION
id	INT	Primary key, auto-incremented.
User_id	VARCHAR(255)	Unique identifier for the user
Role	VARCHAR(50)	Role of the message sender (User or assistant)
Content	TEXT	The Actual message Content
Timestamp	DATETIME	Timestamp of the message was saved

Data Flow Diagram
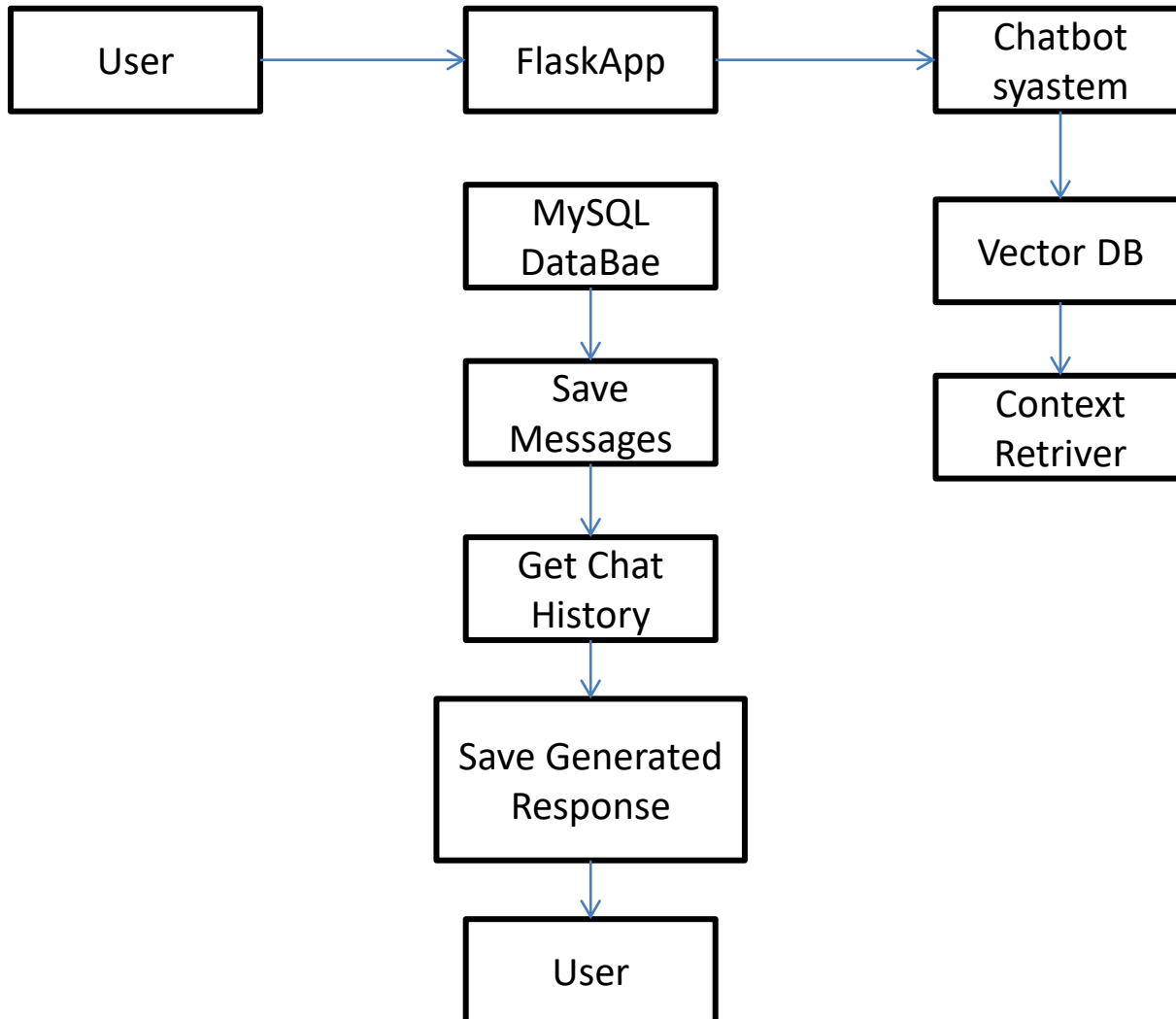
Agentic Chatbot Workflow

Agentic RAG Workflow: From Input to Output

Workflow Initialization:

- A StateGraph is set up with an initial state to manage the process flow.

Key Nodes & Their Roles:

- **Vector_Retriever:**
  Starts the process by retrieving relevant documents.

- **Conditional Branch:**
  Grades the retrieved documents to decide the next step:

  - **Output_Generator:**
    If documents are sufficient, it generates the final answer.

  - **Query_Rewriter:**
    If more refinement is needed, it rewrites the query for further processing.

- **My_Ai_Assistant:**
  Processes the refined query and produces a tailored response.

- **Flow Overview:**

  - **START → Vector_Retriever**

  - From **Vector_Retriever**, a conditional decision is made:

    - If graded well → **Output_Generator → END**

    - Otherwise → **Query_Rewriter → My_Ai_Assistant → END**
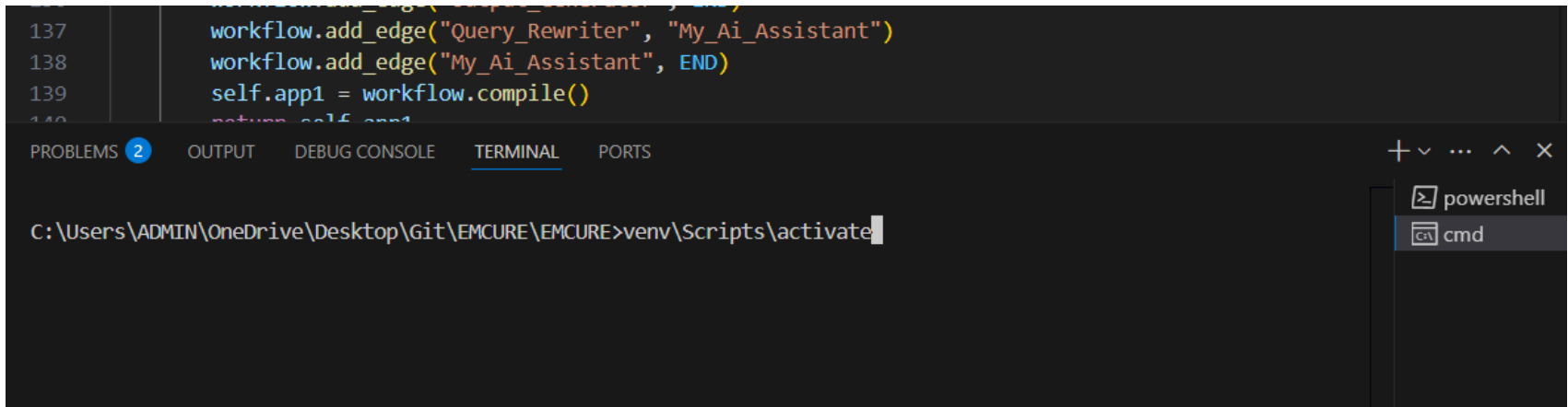
- **Summary:**
  This agentic workflow dynamically routes the query: initially retrieving documents, then conditionally choosing between generating an output or refining the query. This iterative approach ensures optimal response accuracy and adaptability.

Future Enhancements

- **Enhanced Accuracy:**
  Integrate **Llama 3.3** and **DeepSeek** models to boost chatbot response precision.

- **Feedback Loop:**
  Add a **Feedback Agent Node** in the workflow:

- **Positive Feedback:** Final output is delivered to the user.

- **Negative Feedback:** Output is reprocessed for regeneration.

- **Additional Improvements:**
  Explore integration of a **React Agent** and a **Self-Asking Agent** for even more accurate responses, acknowledging that these options may require longer generation times.

HOW TO RUN

Open folder and open terminal and add environ ment run this "venv\Scripts\activate"
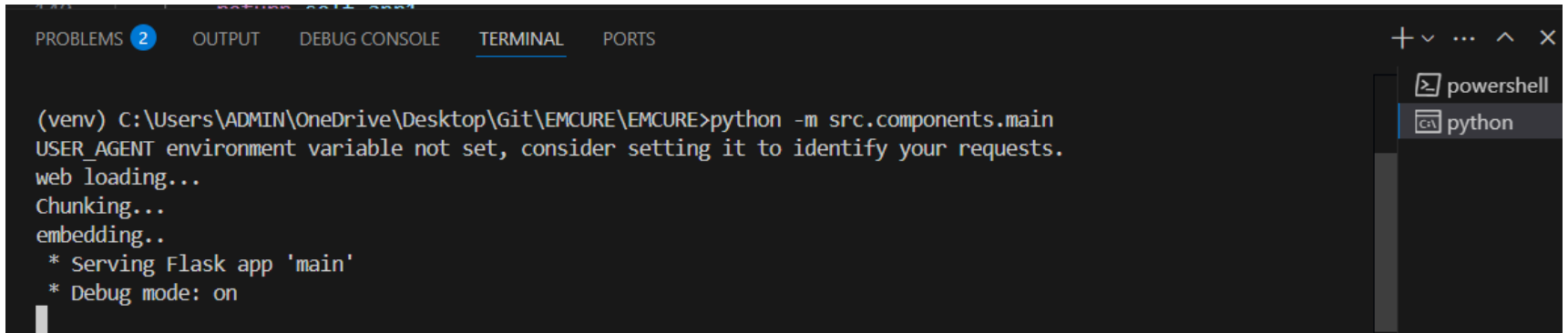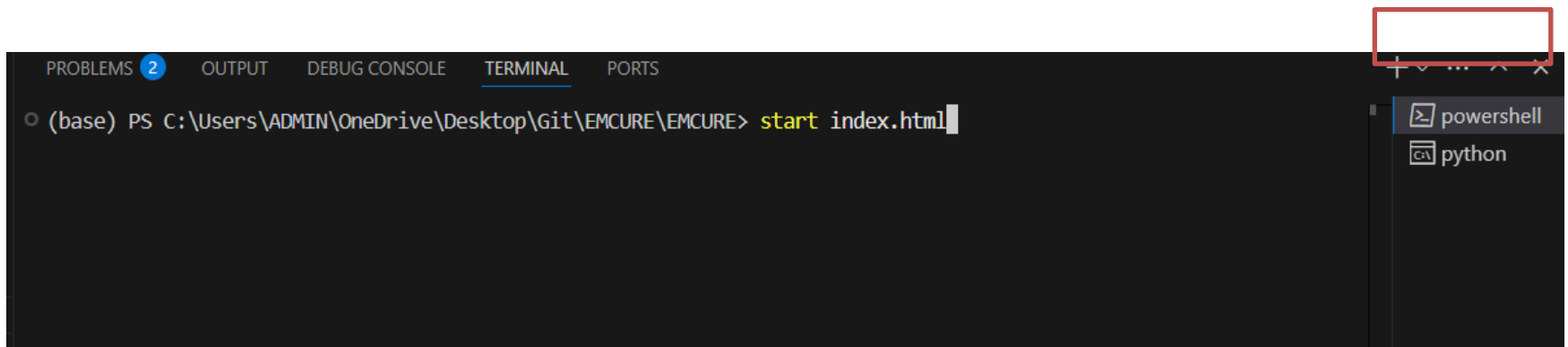
Run : "python -m src.components.main" in cmd it will take time because of embedding from website
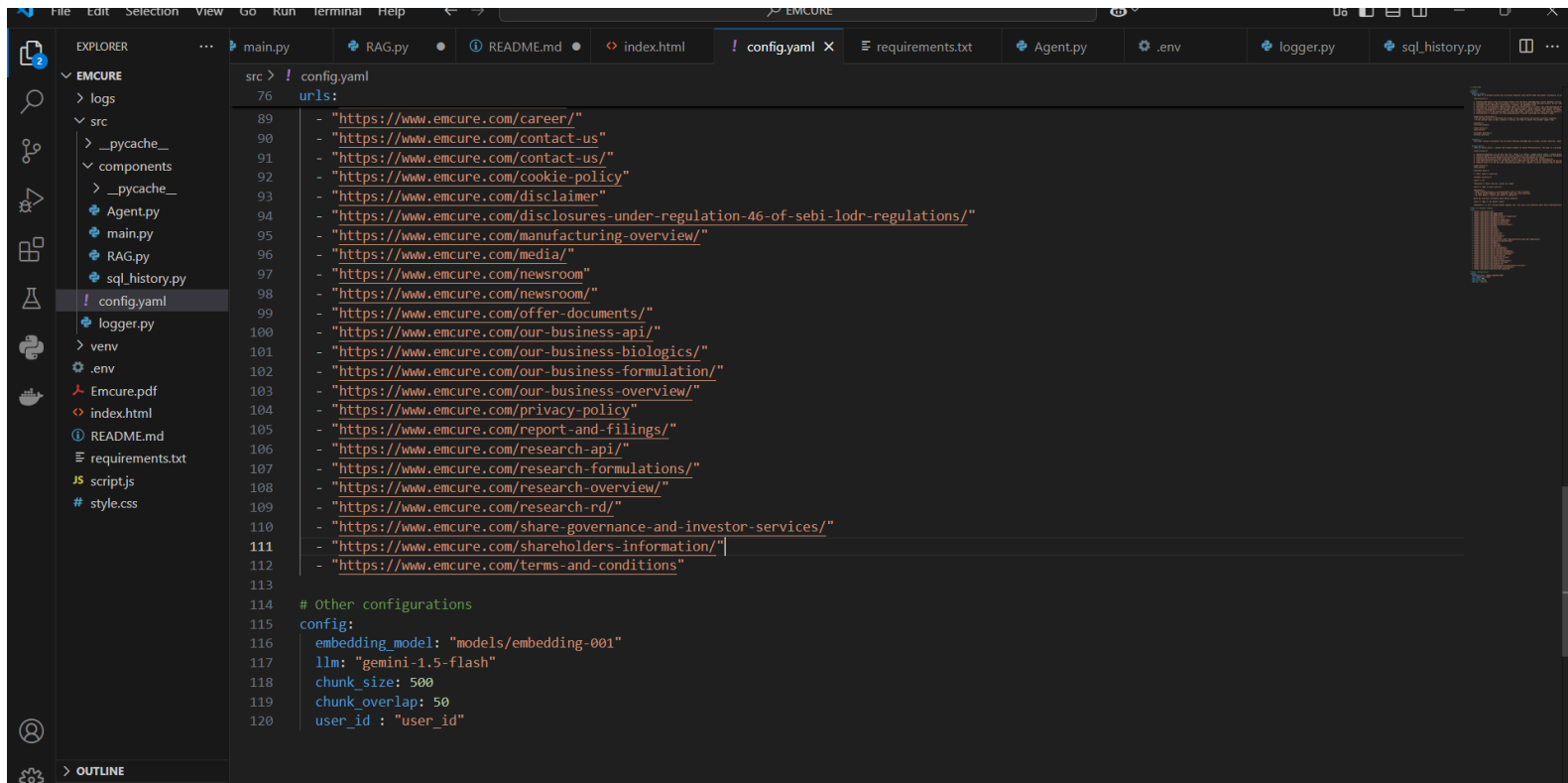
Wait until it run properly

Run : "start index.html"

it will open chatbot interface where you can talk to agentic chatbot

All prompts and webpages link, embeddingmodel , llm model and user_id stored in config.yaml file

Chatbot Testing Questions

1. What are the therapeutic areas that Emcure Pharmaceuticals specializes in?
2. What is Emcure's mission statement?
3. Does Emcure have any information about its research and development initiatives?
4. Where can I find Emcure's latest news or press releases?
5. Does Emcure provide information about its commitment to corporate social responsibility (CSR)?
6. What are Emcure's contact details for customer support or medical inquiries
7. Are there any career opportunities listed on the Emcure website?

Test Memory Of Chatbot

1. Hi my name is jay
2. What is my name

Thank you