**Experienced Developer Needed for Medical Terminology Chatbot**

*∆ Overview*  
Develop chatbot that can answer for free based on spreadsheets that have great information about common medical terminology.  
  
*∆ Technologies*  
Front-End: **React**  
Back-End: **Python, FastAPI**  
LLM model: **GPT4ALL-J**DataBase: **SQLite**  
Vector Database: **Chroma**  
Other: **Langchain**  
  
*∆ Step*1. Chatbot AI System Development: 2 days  
2. Chat Bot UI development: 1 day  
3. Backend For Blog Management: 1 day  
4. Frontend For Blog Page in User Side: 1 day  
5. Frontend For Blog Page in Admin Side: 1 day  
6. Test and Deploy: 1 day

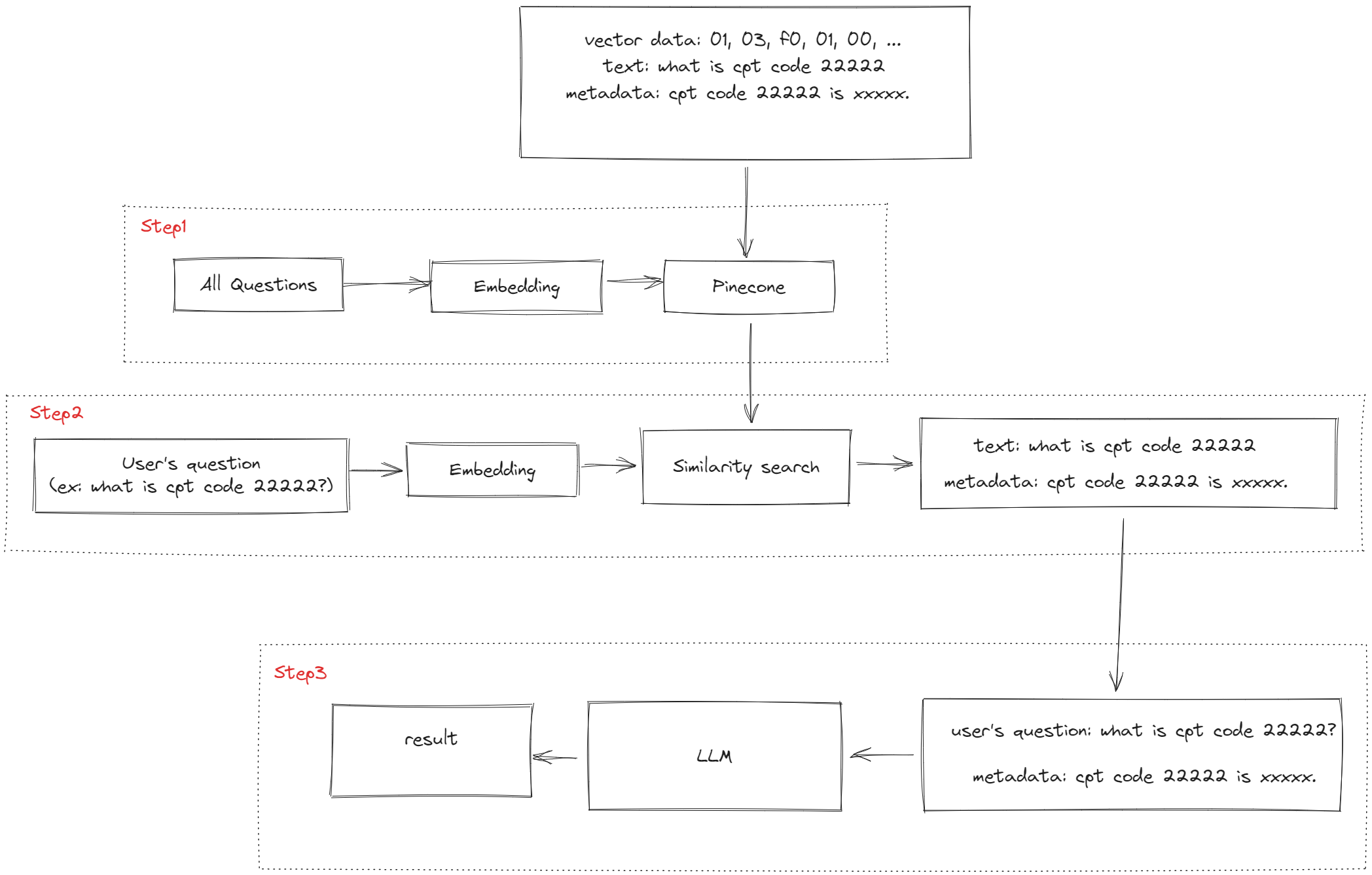
*∆ How to train*  
- Uses **LangChain** tools to parse the document and create embeddings locally using **HuggingFaceEmbeddings**. It then stores the result in a local vector database using **Chroma** vector store.  
  
- Uses a local LLM based on GPT4All-J to understand questions and create answers.  
  
- Data from Excel is converted into sentences and included in a vector database.

For example, "code: AAA, description: BBB" will be converted to something like "If code is AAA, description is BBB"

AI performs semantic search with the user's queries and gets answers.

The chatbot answers the spreadsheet, and if the chatbot doesn't find the right data, it replies "I don't know."  
  
*∆ How to record queries for data gathering*- Implement with Session or Cookie. Each time a user sends a POST request to the chatbot server, the application checks if the user already has an assigned id. If they don't, a new universally unique identifier (UUID) is created and sent back to the user as a cookie. The user's browser will then send the cookie with each subsequent request.

If it's the first time this user (identified by their id) sends a query, the query is stored in the database.

*∆ Architecture*- workflow  
**

A chatbot consists of three steps:

Step 1

- All questions are embedded with openai and stored in Pinecone.

- Vector data has the following format.

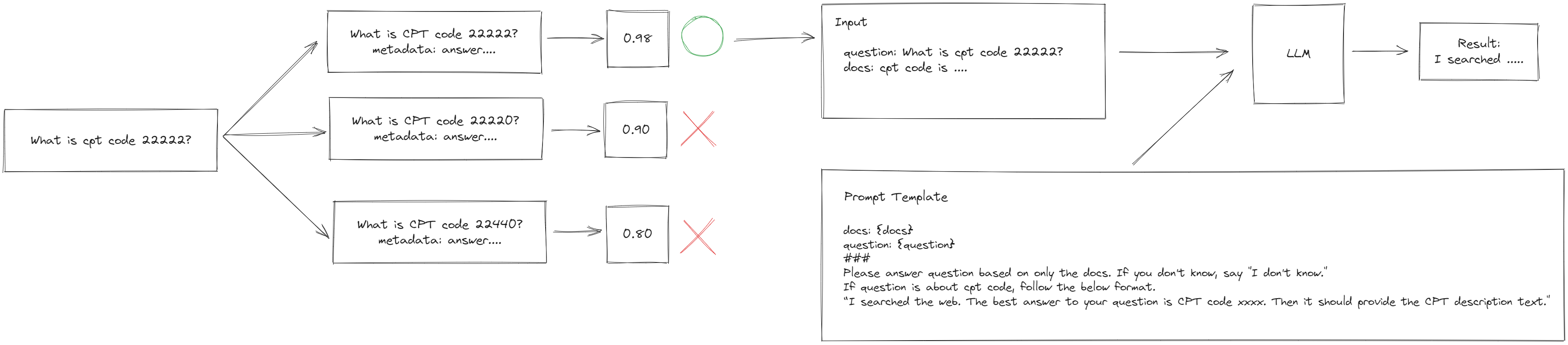
*value: 0.0253637824, 0.0104482248, 0.0157223437, …*

*text: what is cpt code 22222?*

*metadata: answer: The CPT code for osteotomy of spine, posterior or posterolateral approach, 1 vertebral segment; lumbar is 22214. That's my best pick.*

Step 2, Step 3

- Insert the user's question into the Pinecone vector database and perform a similarity search.



*∆ How to run on local*  
  
- Frontend   
\* install node.js 18.0.0  
\* npm install  
\* npm start  
  
- Backend  
\* install python 3.10

\* pip install virtualenv

\* virtualenv venv

\*run venv/Scripts/activate.bat (venv/bin/activate in Ubuntu)  
\*pip install -r requirements.txt  
\*python main.py