

Building a Chat Application

Tools leveraged

Chainlit – Python Library Building Conversational UI

LangChain – Framework to streamline the development of applications with LLMs

Vector Store – FAISS or Chroma DB

Conversational Chain

Differences among various retrieval_qa in langchain - <https://github.com/langchain-ai/langchain/issues/7845>

- [retrieval_qa](#)
- [question_answering](#)
- [qa_with_sources](#)
- [conversational_retrieval](#)
- [chat_vector_db](#)
- [question answering](#)

Chainlit = DocumentQA (is based on LangChain = conversational_retrieval)

Pending Items

- ~~Can explore LazyLoad for Website~~
- Reranker of results with Cohere
- Make the chat listen to Audio and respond in Audio
- Support switch to process audio input and output
- Integrate with CrewAI for Agentic framework

Build and Deploy using docker images into GCP

```
gcloud builds submit --tag gcr.io/PROJECT_ID/image_name .
```

```
gcloud run deploy SERVICE --image gcr.io/PROJECT_ID/image_name
```

Few other items

- Other Chat frameworks – Rasa, Botpress
- <https://docs.livekit.io/agents/quickstarts/s2s/>
- LLM Ranking - <https://botpress.com/llm-ranking>
- Explore Realtime API for faster communication

Deploying on GCloud

<https://medium.com/@tahreemrasul/deploying-langchain-apps-with-langserve-on-gcp-1908476cdc49>

<https://pseudohvr.medium.com/deploying-chainlit-on-gcp-72231ba6b77f>

<https://www.googlecloudcommunity.com/gc/Developer-Tools/Error-to-deploy-Google-App-Engine-with-gcloud-build-submit/m-p/820892>

References

<https://docs.chainlit.io/examples/qa>

<https://www.zinyando.com/building-a-conversational-chatbot-with-crewai-groq-chromadb-and-mem0/>

<https://nayakpplaban.medium.com/building-an-llm-application-for-document-q-a-using-chainlit-qdrant-and-zephyr-7efca1965baa>