

Kalpit Pvt Ltd, UK - AI Intern Hiring : Assignment 1

Phase 1 - Core Skills Evaluation - Building A Functional Prototype

The Task

You are required to build a simple command-line Q&A system. The system will ingest the text from a provided short speech by Dr. B.R. Ambedkar and answer questions based *solely* on that content.

Your system must perform the following steps:

1. Load the provided text file (speech.txt).
2. Split the text into manageable chunks.
3. Create Embeddings and store them in a local vector store.
4. Retrieve relevant chunks based on a user's question.
5. Generate an answer by feeding the retrieved context and the question to an LLM.

Technical Requirements & Constraints

1. Programming Language: Python 3.8+
2. Core Framework: You must use the LangChain framework to orchestrate the RAG pipeline.
3. Vector Database: use ChromaDB as your vector store. It is open-source, free, and can run entirely locally.
4. Embeddings: : HuggingFaceEmbeddings with sentence-transformers/all-MiniLM-L6-v2 - NO API keys, NO accounts, NO costs.
5. LLM: Ollama with Mistral 7B - 100% free - no accounts, no API keys, no limits
6. Deliverables: You must provide a public GitHub repository containing:
 - o Your well-commented Python code (e.g., main.py).
 - o A requirements.txt file with all dependencies.
 - o A detailed README.md file explaining how to set up and run your code.
 - o The provided speech.txt file.

Provided Data

Create a file named speech.txt and paste the following text (a short excerpt from the "Annihilation of Caste" for demonstration purposes):

text

Text for speech.txt

The real remedy is to destroy the belief in the sanctity of the shastras. How do you expect to succeed if you allow the shastras to continue to be held as sacred and infallible? You must take a stand against the scriptures. Either you must stop the practice of caste or you must stop believing in the shastras. You cannot have both. The problem of caste is not a problem of social reform. It is a problem of overthrowing the authority of the shastras. So long as people believe in the sanctity of the shastras, they will never be able to get rid of caste. The work of social reform is like the work of a gardener who is constantly pruning the leaves and branches of a tree without ever attacking the roots. The real enemy is the belief in the shastras.

Submission Instructions

1. Create a public GitHub repository named AmbedkarGPT-Intern-Task.
2. Push all your deliverables (code, requirements.txt, README.md, speech.txt) to this repository.
3. Share the repository URL with us via the portal/email as specified in the initial hiring communication.
4. You have 7 days to complete this assignment from the date you receive this brief.

Getting Started Hints

- Set up a virtual environment: Use venv or conda to manage your dependencies.
- Install Ollama First: Download and setup Ollama with Mistral 7B before starting coding.
- Start Simple: Begin by testing Ollama LLM alone, then add embeddings and Chroma step-by-step.

- Test Incrementally: Verify each component works before integrating into the full RAG pipeline.

Key LangChain Imports you will likely need:

python

- *from langchain.document_loaders import TextLoader*
- *from langchain.text_splitters import CharacterTextSplitter*
- *from langchain.embeddings import HuggingFaceEmbeddings*
- *from langchain.vectorstores import Chroma*
- *from langchain.llms import Ollama*
- *from langchain.chains import RetrievalQA*

Quick Ollama Setup:

```
curl -fsSL https://ollama.ai/install.sh | sh  
ollama pull mistral
```

We are not looking for a production-ready system, but for a clean, functional prototype that demonstrates your understanding of the fundamental building blocks.

Good luck! We look forward to seeing your implementation.

Hiring Manager

Kalpit Pvt Ltd, Uk

kalpiksingh2005@gmail.com