Job Searching AI Virtual Agent Documentation:

(Note – We pivoted from making a job-searching application for users, to a recruiter-based application to find candidates for the job they are looking for. Most of the architecture stayed the same. We did this for two reasons. First, we identified there are many AI-based job searchers already available. Secondly, we believe this modification is a unique idea that can be refined to implement in the industry.)

Textdata Link: https://textdata.org/submissions/6725088c22ea54f14be31ff0

Code Documentation:

1. *Data* – Utilized 25 resumes to create embeddings for this project, fabricated using ChatGPT model GPT4o. This data was formatted as JSON files to easily retrieve fields out from each resume. *In the future, there would be a way to add a function so that users can upload their resumes, and a function would convert it to JSON. Then it would be uploaded to the data file to be used for embeddings.*

2. Data Processing – Parses candidate resumes stored as JSON files into structured text to extract key details like Skills, Education, Experience, and Projects. Then, splits the structured text into smaller chunks using CharacterTextSplitter to enable embedding generation.

3. Vector Database Ingestion – Converts processed text chunks into vector embeddings using OpenAI's **text-embedding-ada-002** model. Next, stores these embeddings in a persistent ChromaDB directory for efficient retrieval during queries.

4. LLM – Creates a retriever from the ChromaDB vector store to find the most relevant resume chunks for a query. After, uses a RetrievalQA chain with OpenAI's **GPT-4** model to process retrieved data and generate accurate, context-aware responses to user queries.

5. Built-in UI – Implements a command-line interface for recruiters to interactively query the candidate database and receive AI-generated answers. Finally, builds a graphical user interface using ipywidgets. This features a text input box, a send button, and an output area to display conversations. *In the future, this would be a full-stack application with dedicated data storage.*