Ryan McNeil

Assignment 3

1. Read through the code files under “./src” and describe what the following files are doing.
   1. Config.py : defining some variables & settings for our program. AI decoder models; embedding model (just the one); a temperature (something like “randomness”); Max Tokens for our decoder model; an app name (more of an intro statement). And directories for documents.
   2. app.py : creates the Streamlit web app. Loads the .env; imports the config; initializes the session state; sets the name; creates a sidebar in the app for selecting a model from the “Models” list, set temperature, & store those selections in the session state; runs upload & settings to handle file uploads; runs the chatbot to interact with the chatbot interface.
   3. app\_sections.py : defines functions for the app. run\_upload\_and\_settings() runs the upload and settings container. allows user to choose a file, copies it to the ‘raw’ folder & saves it to the session state. run\_chatbot() sees when the start button is clicked, builds a vector database from the uploaded PDF file, and allows for chatting; generates responses for user prompts.
   4. app\_utils.py : contains utility/helper functions for our app. Initializes API, defines our embedding function, defines utility functions. get\_embedding(): generates embedding vector from sentence. get\_retriever(): loads vector retriever from FAISS database. Initialize\_session\_state() initializes session state variables. Num\_tokens\_from\_string() calculates the number of tokens in a string. Generate\_response() generates responses to user prompts using the selected model & API. Create\_knowledge\_base() processes documents to compute embedding vectors & builds the FAISS database. Generate\_kb\_response() generates responses from the knowledge base based on prompts.