Week 7

Overview

- Our team concentrated on coding the beta version of the entire pipeline and its implementation. This work built upon the skeleton code developed during the previous week.

Key Activities

1. **Beta Implementation**
   - Followed guidelines specified in the project report to develop the code.
   - Researched best practices (via LLMs, Google, etc.) to ensure efficient implementation.
   - Completed an initial version of the final product (see GitHub repo), which, although simplified from the midterm report plans, is fully functional for demonstration purposes.
2. **Testing & Quality Checks**
   - Performed initial testing on the accuracy of document and audio file transcriptions.
   - Manually reviewed a random sample of 40 files (7 audio) to confirm correct transcription. All tested files passed.
   - Experimented with different chunk sizes for embedding generation and with varying numbers of top documents (k) returned by the search to determine optimal performance.
     - Concluded that a chunk size of 500–750 tokens is optimal (leaning toward 750).
     - Found that k=10 provides the most informative answers but may slow down the process compared to k=5.
3. **UI Layout Experimentation**
   - Experimented with various user interface layouts to determine the best approach for displaying chat conversation history and search results.
   - Kept the current UI design simple (as shown in the presentation slides and GitHub repo) but plan to consult with the client and TA for final layout decisions.
4. **Next Steps**
   - Will meet with the project manager (TA) to demonstrate the current version and discuss any final changes needed before scaling up to a final version for our end-of-class demonstration and presentation.

Looking Ahead

- We plan to refine the beta implementation based on feedback from our TA and client.
- We will continue optimizing data processing, improving the UI, and finalizing all project components in preparation for the final demonstration.