Week 6

Overview

- This week, the team focused on preparing the midterm presentation (slides and report) and began developing the skeleton code to establish data flow and initial implementation.

Key Activities

1. **Midterm Presentation and Report**
   - Developed the midterm report using information from the previously completed project proposal.
   - Prepared presentation slides to accompany the report.
2. **Scope Adjustment**
   - Realized that the original dataset exceeds 100 GB, which is too large to handle within the scope of this course and Decided (in consultation with the professor) to reduce the dataset scope for the class portion of the project. This subset will be used for demonstration and testing.
3. **Data Flow Finalization**
   - Finalized the project's data flow after internal discussions and consultations with the TA.
   - Documented the flow in the report for clarity.
4. **Tool Justification & Research**
   - Conducted additional research to justify the choice of tools and frameworks used in the project.
   - Compared alternatives and included this analysis in the midterm report.
5. **Initial Data Cleaning**
   - Began preliminary data cleaning and sorting processes.
   - Ensured that we remain on track to start data pre-processing and embedding generation next week (according to the project timeline).