Progress Report

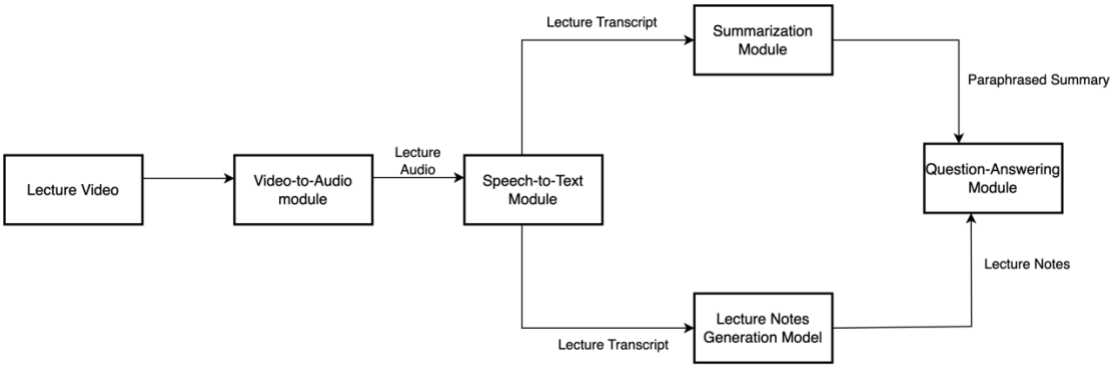
Study Buddy

Lecture Summarization and Question-Answering System using Video-to-Speech-to-Text

-By Akshay Chavan, Gaurav Tejwani, Ananya Asthana

All Project Objectives: The core objectives of our project are:

1. Video-to-audio: Using python moviepy to convert lecture videos to speech or audio
2. Speech-to-text: Using OpenAI's Whisper model to generate audio transcription
3. Summary and paraphrase generation: The transcribed text is processed to produce paraphrased sentences which are then converted into a summary using nltk library and T5 Tokenizer. Text scoring using Cosine Similarity is implemented to weigh more relevant parts of the lecture.
4. Lecture Notes Generation: The transcribed text is also processed to generate well-structured lecture notes using OpenAI's GPT-4o mini model and prompt engineering.
5. Question-Answer Module: The Summary, Paraphrased sentences and Lecture Notes generated in the previous steps are now fed into our Q&A model, which uses FLAN-T5-Large and vector-based semantic search with Instructor XL embeddings to answer questions from the user based on the topic of the lecture.
6. A responsive, user-friendly UI that allows the user to upload the lecture video, view the transcript, view the summary, paraphrased sentences and lecture notes, and also interact with the Question & Answer module in chat form. This is built using Next.js, React.js, Tailwind and Python Flask.



Current Progress:

We have completed points 1, 2, 3, 4, and 6 from the project objectives listed above. We are now able to upload a video, see its transcript generated, get a summary created from paraphrased sentences after text scoring and also view lecture notes. We have a complete frontend UI for the user to perform these tasks on.

Here is a demo video showing our current tool progress:
https://www.youtube.com/watch?v=ti9medwwD8k

Next Steps:

Point 5, which is our Question-Answer Module, is our next goal to complete. We will feed in our summary, which was generated from paraphrased sentences and used sentence scoring, as well as our lecture notes

which were generated using OpenAI's GPT-4o mini model into Q&A module and allow users to ask questions based on the lecture content, for which they should receive customized answers based on the lecture content. Currently the UI for this segment is complete. On the backend we are currently experimenting with different models that are available on HuggingFace to see if we can find a better alternative to FLAN-T5-Large and vector-based semantic search with Instructor XL embeddings, else we will complete it using this model itself.