Project Daily Log

Date: 13/02/2025
Start time: [8 AM : 11:30 AM]
Location: [Loyalist college campus]

Task Summaries / Logs

1) Research on BERT TOPIC.

   Details : Reading up on Documentation on BERTopic for finding an alternative process to create chapter boundaries as there are some limitations on current LLM based approach.

   Duration : 8:30 AM - 9:45 AM

2) Meeting with ML OPS team.

   Details : Discussion with Yash from ML OPS team on the approach to create the testing pipeline for creating the initial Accuracy metrics for our current model.

   Duration : 10 AM - 10:30 AM

3) Python Scripting

Details : Coded a python script to retrieve youtube lecture videos from the given database / dataset and saving them locally along with converting the files to WAV format using our current library.

Duration : 10:30 AM - 11:15 AM

4) Code Cleanup

Details : Doing cleaning and examining for the chapter generation program to make sure everything works smoothly for the testing phase.

Duration : 11:15 AM - 12: PM

Notes / Observation

1) The current chapter generation program works but facing some issues in generating number of chapters with matching number of chapters in the dataset.

2) BERTopic can be used for creating chapter boundaries as an alternate to current LLM based approach.

Tasks for Next Session

1) Finish the chapter generation pipeline to create a database of the generated content.

2) Align the data from test dataset for comparing Scores.

3) Create program for ATON Scores and Cosine Similarity.

Signed by Yohan khan [500228403]