CSC490 Assignment 1

Team Members
Minh Le (1007894432), Prashanth Shyamala (1008819021),
Shaan Purewal (1008332939), York Ng (1009153577)

1 Interest Statements

Our team wants to work on the problem of semantic search and retrieval of academic lecture videos, with the goal of helping students efficiently locate the specific moments where a concept is defined, explained or applied. Many recorded lectures are long and often labeled only by date or other non-descriptive metadata, which makes it difficult to find targeted content during studying. As a result, students frequently spend substantial time manually scrubbing through videos to revisit a single topic such as "*the determinant of a matrix*". We care about this problem because we encounter it regularly in our own coursework and we believe that improving navigability can make recorded lectures more effective as a learning resource.

Our proposed system supports a query-based workflow in which a user searches across a set of lecture videos, the system identifies relevant segments using semantic signals from audio and visual information and it then composes a concise mini-lecture by stitching together the most relevant clips. This approach aims to surface key components of an explanation, including definitions, notation and worked examples, even when these occur in differing parts of one lecture or across multiple lectures.

Name	**Contribution Comments**
Minh Le	*I'm interested in implementing research ideas in VideoRAG and/or Vision-Language Models to process videos at scale, and I am also interested in learning more about ML engineering practices.*
Prashanth Shyamala	*I am interested in learning about the infrastructure behind scalable ML systems, especially in the LLM and Vision space. I plan on contributing specifically in this area in the project - analyze trade-offs and improve efficiency of the final design.*
Shaan Purewal	*I'm particularly interested in contributing to the infrastructure aspects of this project, with a focus on learning and applying infrastructure-as-code practices using Terraform to create easily reproducible environments, as well as exploring system design considerations at scale.*
York Ng	*I hope to implement the multimodal topic segmentation pipeline while exploring ways to support large-scale inference, thereby balancing accuracy (of topic representations) and efficiency.*

Table 1: Contribution Comments from the Team

2 Landscape Analysis

Item	Type	Description	Commentary
Panopto (Smart Search)	Competitor (Company)	Lecture and video platform with "search inside video" using automatic speech recognition and OCR	Strong lecture-native indexing (speech + on-screen text). Gaps: It primarily returns timestamped matches, so students still inspect results manually instead of pre-evaluated relevance.
NoteGPT	Competitor (Company)	Tool for video summaries, chapter labeling, and note generation	Reference for UX design. Gaps: not integrated into Quercus, users need to manually upload videos to be processed.
TwelveLabs (Video Search API)	Competitor (Feature)	Multimodal video search that finds specific moments using natural language queries	Strong baseline for semantic moment retrieval beyond keywords. Gaps: It is only a retrieval infrastructure; does not assemble retrieved moments into a coherent instructional sequence.
Opencast API	Infra	Open-source video management system used by UofT	Interface to fetch video assets (MP4) and existing metadata securely from Quercus.
ColPali	Research	Vision-Language Retrieval model that treats document pages (slides) as images rather than OCR-ing text	Important for slide-heavy and technical courses where OCR fails on complex diagrams or equations. It also matches queries directly to visual slide content.
Video-RAG	Research	Methodologies for retrieving relevant video clips to answer complex user queries	Provides the theory behind how to chunk long videos (1hr+) effectively without losing context.
Whisper	Model	Automatic Speech Recognition (ASR)	Standard for obtaining high-quality transcripts, although might be very expensive.
QwenVL	Model	Open-source Vision Language Model	SOTA VLM for obtaining visual understanding of images and videos, although will require a lot of GPU compute to run queries.
LlamaIndex	Framework	Open-source data framework for connecting data to LLMs, specializes in multimodal RAG	Framework to build chunking and retrieval pipeline to store video nodes with timestamps.
Elasticsearch	Database + Search	Supports semantic search of unstructured multimodal data	Well-documented vector database and search engine, suitable for vectorized video frames

Table 2: Landscape Analysis of Relevant Companies and Tools in Video Understanding

Overall, given that most lecture recordings consist of static visual elements (slides) rather than complex motion, a major opportunity for our project is an approach specialized for lecture content. By using Vision-Language Models and Video-RAG, we can achieve multimodal alignment, enabling the system to resolve deictic references such as linking the spoken phrase "this equation" to the specific formula displayed on the slide. Furthermore, another opportunity is data privacy. Since we are offering an integration within Quercus at UofT, external APIs might be restricted for student data.

3 Project Outline

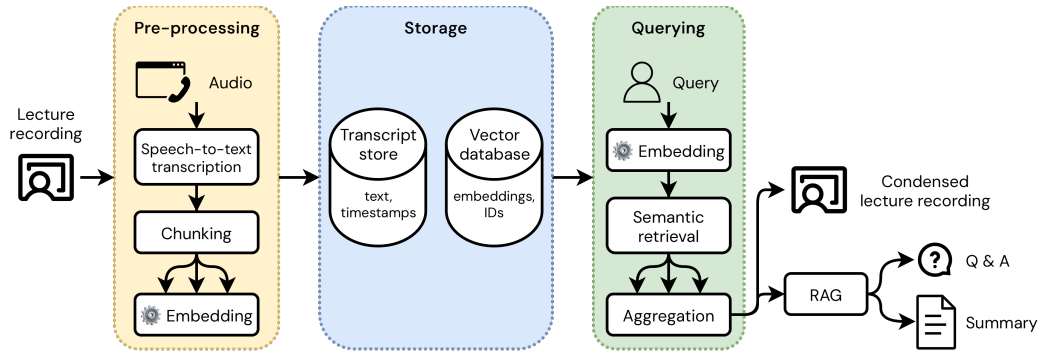
Problem Statement. Students devote excessive time watching lecture recordings. This is because recordings are long and unstructured, making it inefficient to locate explanations of specific concepts. Moreover, existing platforms provide limited support for topic-based search or targeted review, forcing students to manually scrub through recordings. Combined, these factors increase time spent searching rather than learning. This project addresses the need for a system that automatically indexes and summarizes lecture recordings to enable fast and accurate retrieval of content specific to user queries.

Proposed Solution. We propose **LectureClip**, a system that transforms long, unstructured lecture recordings into a searchable and reviewable learning resource. The system enables students to retrieve timestamped video clips relevant to a given concept or question and provides concise summaries to support efficient review. By automatically indexing lecture content and supporting semantic search, LectureClip reduces time spent navigating recordings and improves access to specific explanations.

Technical Approach. LectureClip uses a (1) pre-processing and a (2) query-time retrieval pipeline.

During pre-processing, lecture recordings are transcribed into timestamped text segments. Each segment is converted into semantic text embeddings, while sampled video frames are converted into visual embeddings, both stored alongside timestamps in a vector database. Topic segmentation is applied to group adjacent segments discussing similar concepts.

At query time, a user's input (a topic) is embedded and matched with transcript embeddings using semantic similarity search. Retrieved text and visual segments are temporally aggregated to form coherent video clips, which are returned to the user with precise timestamps. These retrieved transcripts and visual context are then provided to an LLM to generate concise summaries and support question answering using VideoRAG.



Milestones

1. **Data Ingestion and Preprocessing**
   Automated lecture video ingestion, speech-to-text transcription with timestamped segmentation, and structured storage of transcripts and metadata.
2. **Multimodal Embedding and Indexing**
   Generate text and visual embeddings for transcript segments and video frames, and index them in a vector database to support multimodal similarity search.
3. **Query-Time Retrieval Service**
   Convert queries into embeddings, retrieve relevant transcript segments via vector search, aggregate adjacent segments into coherent clips, and return results via an API.
4. **Summarization and Q & A**
   Integrate a retrieval-augmented generation (RAG) pipeline to produce concise summaries and support grounded question answering.

Unknowns to Investigate

- *Segmentation and retrieval*: How to meaningfully segment long lectures and retrieve relevant content using text-only versus multimodal (text + visual) embeddings for known and novel queries.
- *Context aggregation and evaluation*: How much context to include when forming clips, how to promote coherence between clips, and how to evaluate usefulness and faithfulness of retrieved content and LLM-generated summaries (e.g. LLM-as-a-Judge).

4 Press Release

Introducing LectureClip: Find the Exact Moment You Need from Any Lecture Recording

Toronto, ON — January 2026 — Today, we are excited to announce **LectureClip**, a new application designed to help students quickly find and understand the most relevant parts of their lecture recordings. Instead of scrubbing through hours of audio or video, students can simply upload their lecture recordings, enter a topic or question, and instantly receive the most relevant clips along with concise summaries.

University students increasingly rely on recorded lectures to keep up with coursework, but finding specific explanations after the fact is time-consuming and frustrating. LectureClip solves this problem by turning long and unstructured lecture recordings into a searchable learning resource.

> "I used to spend 30–40 minutes rewatching lectures just to find one explanation," said *Denys*, a second-year engineering student. "With LectureClip, I type in the concept I'm stuck on and get the exact clip in seconds along with an explanation. It's like having a smart dictionary for every lecture."

How It Works

After uploading a lecture recording, students can enter a topic, keyword, or question, such as *'back-propagation,' 'Fourier transform,'* or *'midterm information.'* LectureClip analyzes the recording and returns:

- **Relevant timestamped clips** where the topic is discussed
- **Short summaries** that explain the key ideas covered in each clip
- **A clean interface** that lets students directly view the relevant clip they need

This workflow allows students to review material more efficiently, reinforce understanding before exams, and revisit complex topics without cognitive overload.

Designed for Real Student Needs

LectureClip was built with accessibility and practicality in mind. It supports long lecture recordings, works across different course styles - PowerPoint recordings, in-person demonstrations, etc. - and focuses on clarity rather than technical complexity.

> "Our goal was to reduce the friction between recorded lectures and actual learning," said *Prashanth Shyamala*, Developer. "Students shouldn't have to rewatch entire lectures to answer one question. LectureClip helps them learn faster and with less frustration."

Faculty and academic support staff also see value in the platform.

> "Tools like LectureClip empower students to take control of their learning," said *Dr. Emily Chen*, a fictional faculty advisor. "By making lectures searchable and summarized, it supports diverse learning styles and encourages deeper engagement with course material."

Key Benefits

- **Time savings** by instantly locating relevant lecture segments
- **Improved comprehension** through focused summaries and context-aware clips
- **Reduced stress** during exam preparation and assignment deadlines
- **Better accessibility** for students reviewing content at their own pace

Looking Ahead

LectureClip represents a step toward more intelligent and student-centered learning tools. By transforming passive lecture recordings into interactive study resources, the application aims to help students spend less time searching and more time understanding.

4.1 Earlier Iteration of the Press Release

Introducing LectureClip: Find the Exact Moment You Need from Any Lecture Recording

Toronto, ON — January 2026 — Today, we are proud to announce LectureClip, a new application that helps students quickly find the most relevant segments of their lecture recordings. With recorded lectures now central to university learning, locating specific explanations after the fact is often slow and frustrating. LectureClip streamlines this process by converting lengthy, unstructured recordings into a searchable study tool.

> "I used to spend 30–40 minutes rewatching lectures just to find one explanation," said *Denys*, a second-year engineering student. "With LectureClip, I type in the concept I'm stuck on and get the exact clip in seconds along with an explanation. It's like having a smart dictionary for every lecture."

How It Works

Students can upload a recording and search for a topic/keyword. LectureClip analyzes the recording and returns:

- **Relevant timestamped clips** where the topic is discussed
- **Short summaries** that explain the key ideas covered in each clip
- **A clean interface** that lets students directly view the relevant clip they need

This workflow allows students to review material more efficiently, reinforce understanding before exams, and revisit complex topics without cognitive overload.

Designed for Real Student Needs

LectureClip was built with accessibility and practicality in mind. It supports long lecture recordings, works across different course styles - PowerPoint recordings, in-person demonstrations, etc. - and focuses on clarity rather than technical complexity.

> "Our goal was to reduce the friction between recorded lectures and actual learning," said *Prashanth Shyamala*, Developer. "Students shouldn't have to rewatch entire lectures to answer one question. LectureClip helps them learn faster and with less frustration."

Key Benefits

- **Time savings** by instantly locating relevant lecture segments
- **Improved comprehension** through focused summaries and context-aware clips
- **Reduced stress** during exam preparation and assignment deadlines
- **Better accessibility** for students reviewing content at their own pace