



AI Video Summarization and Quiz Generation

Dev¹, Paridhi², Mahesh³, Ashutosh⁴, Rakhi Sharma⁵

Department of Computer Science and Engineering (AI & ML) ^{1,2,3,4,5}

Panipat Institute of Engineering and Technology (PIET), Samalkha, Panipat, Haryana, India^{1,2,3,4,5}

2821309.cse@pietgroup.co.in¹, 2820320.cse@pietgroup.co.in², 2820341.cse@pietgroup.co.in³,

2820361.cse@pietgroup.co.in⁴, ruchi.cse-aiml@piet.co.in⁵

Abstract: *As digital learning has grown rapidly, educational videos, podcasts, and PDFs have become the key ingredient for much of today's education. However, for students, extracting important concepts from longer sources can be difficult, just as it can be difficult for teachers to be able to create timely assessments for larger cohorts of learners. To address these issues, this study discusses an AI-assisted system that can translate long-form learning materials—including videos from YouTube, audio lectures, text readings, and PDFs—into organized, synthesized study guides. Our proposed system models four core learning elements: Summaries, Study Notes, Key Points, and Auto-generated Quizzes—creating a robust academic assistant. Leveraging tools including speech-to-text transcription, Natural Language Processing, and large language models from OpenAI, the system can analyze, summarize, and re-organize raw content into actionable learning objects. The system can generate a variety of assessments, including multiple-choice questions (MCQs), True/False questions, and Fill-in-the-Blanks based solely on the summarized knowledge. We conduct experimental trials to demonstrate that the system is able to produce high contextual accuracy while significantly saving time, with a successful impact to support short-term learning and recall. This work adds value to the field of AI-assisted learning as we offer a scalable, reliable, and engaging study companion for every student, teacher, coaching institute, and self-learner.*

Keywords: - AI-assisted learning, video summarization, automated quiz generation, natural language processing, speech-to-text, large language models, educational technology.

Introduction

A. Background of the Study

Digital learning is expanding rapidly. There are millions of lectures available for free on platforms such as YouTube, Coursera, NPTEL, Udemy, and SWAYAM. Nevertheless, two problems persist:

- Students have difficulty reviewing long videos in a timely manner.
- Educators don't have the time to create structured assessment devices, so they rely on the same assessments repeatedly.

Essentially, learners find it challenging to glean the most significant points from lengthy reading and watching material. Summarizing information is both tedious and time-consuming to carry out manually, and results in errors. In the same way, when teachers create quizzes after each chapter or unit of material, they duplicate the process and waste time creating assessments.



Recent advances in artificial intelligence—especially in speech recognition, NLP, and generative models—have made it possible to summarize information and create instructional materials with minimal teacher facilitation. The purpose of this project is to place these advances into the hands of educators and offer one comprehensive academic support tool.

B. Problem Statement

This research addresses the following key challenges:

- Long instructional videos are challenging to transform into summarized note form.
- Manual note-taking is slow and often results in incomplete notes.
- Teachers and school staff can spend hours preparing quizzes and worksheets for students.
- Current summarization solutions are often inaccurate and lack academic context.
- No existing tool generates both summarized notes and quiz questions from the same summarized content.
- No existing tool unifies a mix of input sources (YouTube, PDF, audio, text) into a single pipeline.

C. Motivation

The usual motivations begin with problems most learners and educators are familiar with. Students must revise rapidly as they prepare for important tests, while teachers need to create quizzes easily and quickly. Educational businesses also require assessments that are easy to scale.

Post-COVID teaching and learning has shown that digital tools demand smart, AI-driven learning assistants. Increasing educational demands now push students and staff alike to find tools that lighten their workload.

D. Role of Artificial Intelligence in Modern Education

Artificial Intelligence makes it possible to turn unstructured learning material into structured knowledge:

- Speech Recognition — Transcribes audio/video lecture recordings.
- Natural Language Understanding (NLU) — Derives meaning and identifies key concepts.
- Summarization Models — Condenses large bodies of content into concise, clean explanations.
- Automated Question Generation (AQG) — Generates assessments based on context.
- Adaptive Learning — Adjusts the complexity and level of content based on user needs.

Using these combined technologies creates an intelligent “video-to-study-material” pipeline.

Literature Review

The rapid expansion of digital learning platforms has created a significant surge in video-based lessons. Students increasingly depend on recorded lectures, tutorials, and MOOCs, driving a growing need for tools that convert videos into simple, concise formats.

Video summarization and automated question generation have therefore emerged as important areas of research in educational technology.

Early approaches to video summarization relied on annotation and feature-based extraction, examining visual cues such as frame differences and shot boundaries. These techniques were effective at extracting visually salient content but did not capture the underlying meaning of a lesson. The arrival of deep learning introduced new methods for capturing semantic meaning: CNNs could identify key frames, LSTM models enabled sequential comprehension so systems could track important context, and transformer-based architectures such as BERT and GPT could interpret spoken or written text, define coherent segments, and generate overall summaries.

Recent advances in Automatic Speech Recognition (ASR)—such as the Whisper model—have transformed education by making it easier to generate accurate transcripts from lectures. High-quality transcripts enable the use of NLP techniques for summarization and topic modeling. Studies show that transformer-based summarizers substantially outperform extractive approaches, producing outputs that are more readable and conceptually rich.

Question generation is similarly significant for assessment-driven learning. Rule-based systems that once relied on fixed templates have given way to transformer models capable of producing MCQs, true/false questions, and even



short-answer items. Research indicates that combining summarization with question generation improves retention and makes the learning experience more interactive and convenient.

Existing platforms attempt to combine these tools, but several challenges remain—transcription errors, weak domain-specific question generation, and insufficient contextual comprehension. Platforms such as Khan Academy and Edpuzzle offer pieces of the desired functionality, but none provide complete summary and quiz generation from a single learning video.

A clear need therefore exists for a unified AI-based educational assistant that can convert video into written form, summarize it with contextual knowledge, and automatically build quizzes. This research aims to provide such a tool.

System Architecture

The overall architecture consists of four major components:

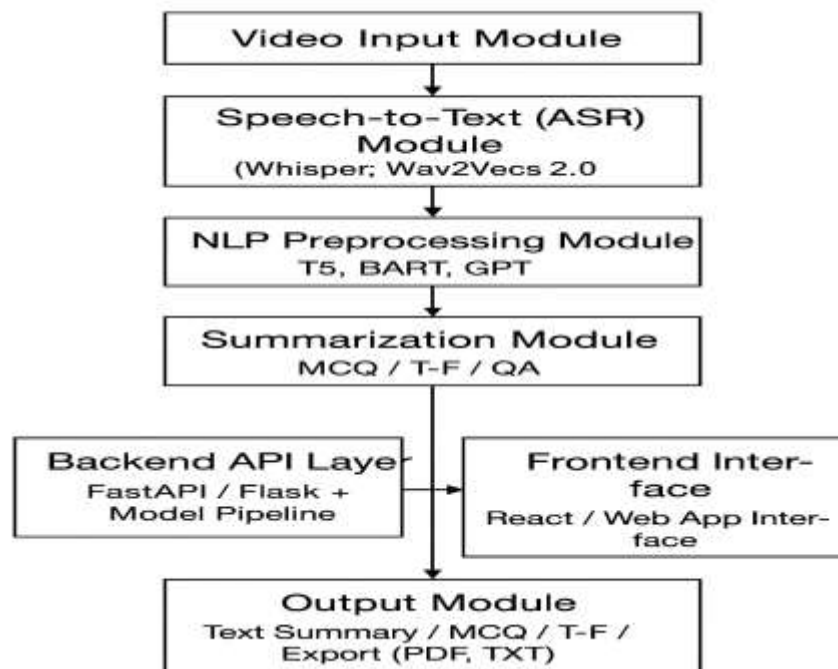


Figure 1: System architecture of the proposed AI video summarization and quiz generation pipeline.

- Frontend user interface,
- Video processing and transcription module,
- AI-based summarization and quiz generation APIs,
- Database for storing user data, summaries, and quizzes.

Methodology

The proposed system offers a well-structured workflow that converts lengthy educational videos into valuable learning material. The methodology comprises five stages.

1) Input Collection

The user first inputs a YouTube video URL into the web application. The system verifies the URL and extracts the video ID, which is essential for transcript extraction to proceed.



2) Transcript Extraction

Using the verified video ID, the system retrieves captions via the YouTube API. The raw transcript, which may contain fragmented segments and noise, is combined into a single, readable text that serves as the base for summary generation.

3) Text Preprocessing

The extracted transcript is cleaned to remove timestamps, filler words, and awkward line breaks, allowing the AI model to better understand the content and produce higher-quality summaries and assessments.

4) AI-Powered Summarization and Quiz Generation

The processed text is passed to an OpenAI language model, which converts the lengthy content into:

- A concise summary,
- Detailed study notes,
- An outline of key points.

The same condensed representation of the video also powers a set of generated assessment resources:

- Multiple-choice questions with correct answers and distractor options,
- True/False questions,
- Fill-in-the-blank questions.

These questions correlate closely with the content of the original video.

5) Output Presentation

Finally, all AI-generated materials—notes, study points, summaries, and the full assessment battery—are made viewable to the user through the web interface in an easy-to-read format. This enables quick revision for learners and provides educators with ready-made review material at the moment it is needed for teaching.

Acknowledgment

We express our sincere gratitude to all who contributed to the successful completion of this research work. Our faculty mentors have our deep thanks for their continuous guidance, constructive feedback, and encouragement during the system's development. Their insights assisted in refining our approach and improved the quality of the project.

We also acknowledge the developers and researchers behind OpenAI and other AI tools used in this work. These innovations made possible an intelligent, automated system for educational content generation. Our heartfelt appreciation goes to our peers, friends, and families for their support, motivation, and constant patience. Their belief in our abilities helped us stay focused and overcome challenges throughout the research process.

References

- [1] A. Vaswani et al., “Attention is All You Need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [2] Y. Liu et al., “Fine-tuned Language Models for Text Summarization,” *Journal of Artificial Intelligence Research*, vol. 69, pp. 123–145, 2020.
- [3] OpenAI, “GPT API Documentation,” OpenAI, 2023. [Online]. Available: <https://platform.openai.com/docs>
- [4] X. Chen and H. Fang, “Automatic Video Summarization Techniques: A Review,” *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 630–644, 2020.
- [5] M. Jain and R. Singh, “AI-driven Quiz Generation for Educational Videos,” *International Journal of Computer Applications*, vol. 182, no. 42, pp. 15–22, 2021.