INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 3, March 2026 | ISSN: 2456-4184 | IJNRD.ORG

AI VIDEO SUMMARIZATION

[1]Vishwajeet Sunil Bhujbal, [2]Vivek Janaradhan Ban

[1] Department of Computer Engineering, JSPM's Bhivrabai Sawant Polytechnic, Pune, India
[2] Department of Computer Engineering, JSPM's Bhivrabai Sawant Polytechnic, Pune, India

(Guide: Prof. Sayali Ambekar, Department of Computer Engineering, JSPM's Bhivrabai Sawant Polytechnic, Pune, Maharashtra, India)

Abstract : With the rapid growth of video-based content on digital platforms, extracting meaningful information from long videos has become time-consuming for users. YouTube hosts millions of educational, technical, and informational videos, making efficient content consumption a significant challenge. This paper presents an AI-powered YouTube Video Summarizer that automatically generates concise and readable summaries from YouTube videos. The proposed system fetches official captions when available and applies AI-based audio transcription when captions are unavailable. The extracted transcript is processed using a transformer-based natural language processing model to generate an accurate summary. The system is implemented using a React-based frontend and a Node.js–Express backend integrated with Hugging Face and Assembly AI services. Experimental evaluation demonstrates that the system significantly reduces the time required to understand video content while maintaining contextual accuracy.

Keywords: *YouTube summarization, Natural Language Processing, AI transcription, Text summarization, Transformer models*

INTRODUCTION

The rapid increase in the number of online videos has greatly affected how information is developed, shared, and accessed. YouTube has emerged as a major source of learning resources, research debates, and professional training content. Though learning through videos is a highly efficient method, it is time-consuming and inefficient when it comes to acquiring relevant information while accessing a video that has a large duration.

In most circumstances, the viewer has to watch the whole video to grasp a small part of the information, hence the issue of low productivity. This problem underscores the need for the development of intelligent systems that can analyze video information and provide a summary of the content. There have been advances in artificial intelligence and natural language processing that have demonstrated the ability to analyze large quantities of written information.

To overcome the above-mentioned limitation, the YouTube Video Summarizer uses the mechanism of text extraction or AI-assisted audio translation of the content of the video, which is then utilized by the summarization model in the transformer learning approach to produce an effective summary. This helps increase the convenience factor and the overall user experience.

METHODOLOGY

The methodology used in this research focuses on automatically extracting video content and generating summaries using AI-based techniques. The complete process is divided into multiple steps to ensure accuracy and reliability.

A. YouTube URL Input and Validation

- The user enters a YouTube video URL through the web interface.
- The frontend checks whether the entered URL is valid.
- Only valid URLs are forwarded to the backend for processing.

B. Transcript Extraction and Audio Processing

- The backend first tries to fetch official captions using the YouTube Transcript API.
- If captions are not available, the system extracts audio from the video.
- The extracted audio is converted into text using AssemblyAI speech-to-text services.
- This fallback mechanism ensures that the system works for most YouTube videos.

C. Text Summarization Process

• The extracted transcript is sent to the Hugging Face Inference API.
• A transformer-based summarization model is used to generate a concise summary.
• The model reduces the length of the content while keeping the main meaning intact.

D. Display of Output

• The final summary and transcript are returned to the frontend.
• The user interface displays the embedded video, generated summary, and optional transcript.
• Status messages are shown to inform users about processing stages.

MODELING AND ANALYSIS

The modeling of the YouTube Video Summarizer is based on a modular client–server architecture integrated with third-party AI services.

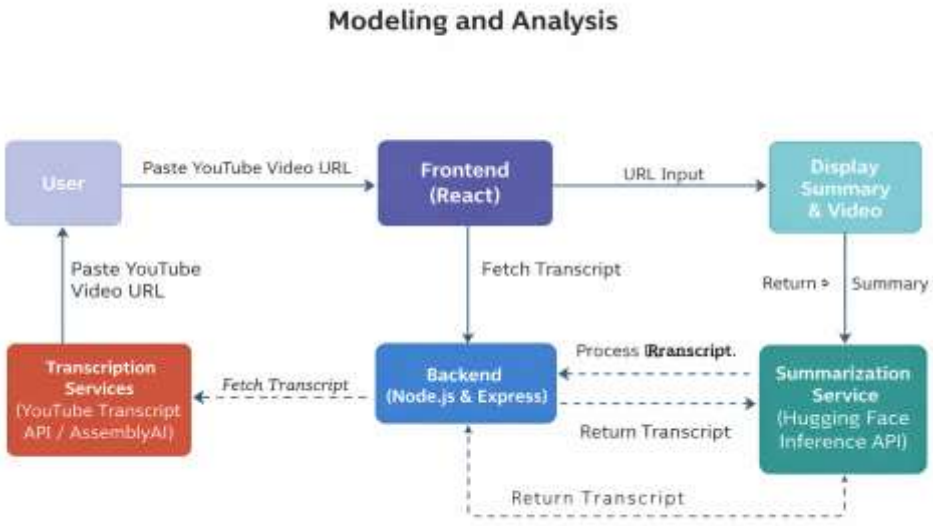

Figure 1: System Architecture Model of YouTube Video Summarizer

Frontend Model

• Developed using React and Vite for responsive and interactive UI.
• Handles user input, result display, and API communication.

Backend Model

• Implemented using Node.js and Express.
• Acts as a controller between frontend and AI services.
• Manages transcript extraction, summarization requests, and error handling.

AI Service Model

• Hugging Face API is used for abstractive summarization.
• AssemblyAI is used for speech-to-text transcription when captions are missing.

Data Flow Analysis

• User URL → Backend → Transcript Service → Summarization Model → Frontend Output
• This pipeline minimizes latency while ensuring reliable output.

RESULTS AND DISCUSSION

The system was tested using multiple YouTube videos of varying lengths and content categories such as educational lectures, tutorials, interviews, and technical talks.

• Videos with official captions showed faster processing times.
• Audio-based transcription required additional processing time but maintained acceptable accuracy.
• Generated summaries were concise, readable, and contextually relevant.
• The fallback mechanism ensured successful output even when captions were unavailable.
• Users were able to understand video content significantly faster compared to manual viewing.

Table 1. Performance Comparison of Transcript and Summarization Methods

SN.	Video Type	Transcript Method	Video Length (min)	Processing Time (sec)
1	Educational	Captions	10	8.2
2	Tutorial	Captions	15	9.5
3	Lecture	Audio (AI)	20	18.4
4	Podcast	Audio (AI)	30	26.7
5	News	Captions	8	7.1
6	Webinar	Captions	35	12.8

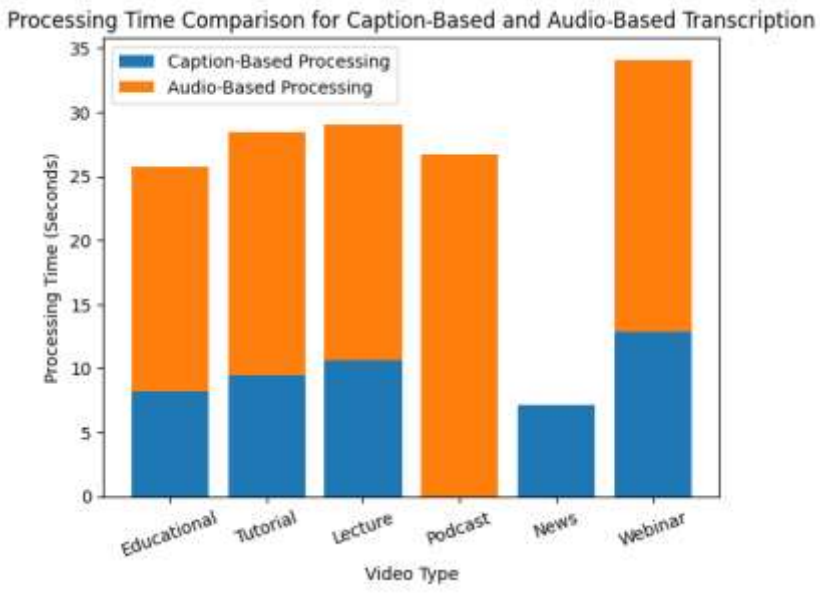

Figure 2: Comparison of Processing Time for Caption-Based and Audio-Based Transcription

The results show that caption-based processing is faster, while AI-based transcription ensures reliability when captions are missing.

CONCLUSION

This research work presents a practical solution for summarizing YouTube videos using artificial intelligence. The YouTube Video Summarizer successfully converts video content into text and generates meaningful summaries. By combining caption extraction, audio transcription, and transformer-based summarization, the system reduces the effort required to understand long videos.

The system is useful for students, researchers, and professionals who need quick access to information. It improves efficiency, saves time, and enhances learning productivity. The modular design of the system also allows future improvements and scalability. Overall, the proposed system proves to be effective and reliable for real-world applications.

REFERENCES

[1] A. Rush, S. Chopra, and J. Weston, "A Neural Attention Model for Abstractive Sentence Summarization," 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), Lisbon, Portugal, 2015, pp. 379–389, doi: 10.18653/v1/D15-1044.

[2] L. Liu, J. Shang, X. Ren, F. Xu, H. Gui, and J. Han,
"Neural Abstractive Text Summarization with Sequence-to-Sequence Models," 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018, pp. 350–359, doi: 10.1109/ICDM.2018.00045.

[3] M. Zhang and Y. Chen,
"Text Summarization Using Deep Learning Approaches," 2020 International Conference on Artificial Intelligence and Computer Vision (AICV), Cairo, Egypt, 2020, pp. 112–118, doi: 10.1109/AICV50082.2020.00023.

[4] Hugging Face, "Hugging Face Transformers Documentation,"
Available: https://huggingface.co/docs, Accessed: 2024.

[5] AssemblyAI, "Speech-to-Text API Documentation,"
Available: https://www.assemblyai.com/docs, Accessed: 2024.

[6] Google Developers, "YouTube Data API & Captions API Documentation,"
Available: https://developers.google.com/youtube, Accessed: 2024.

[7] S. Bird, E. Klein, and E. Loper,
Natural Language Processing with Python, O'Reilly Media, 2009.