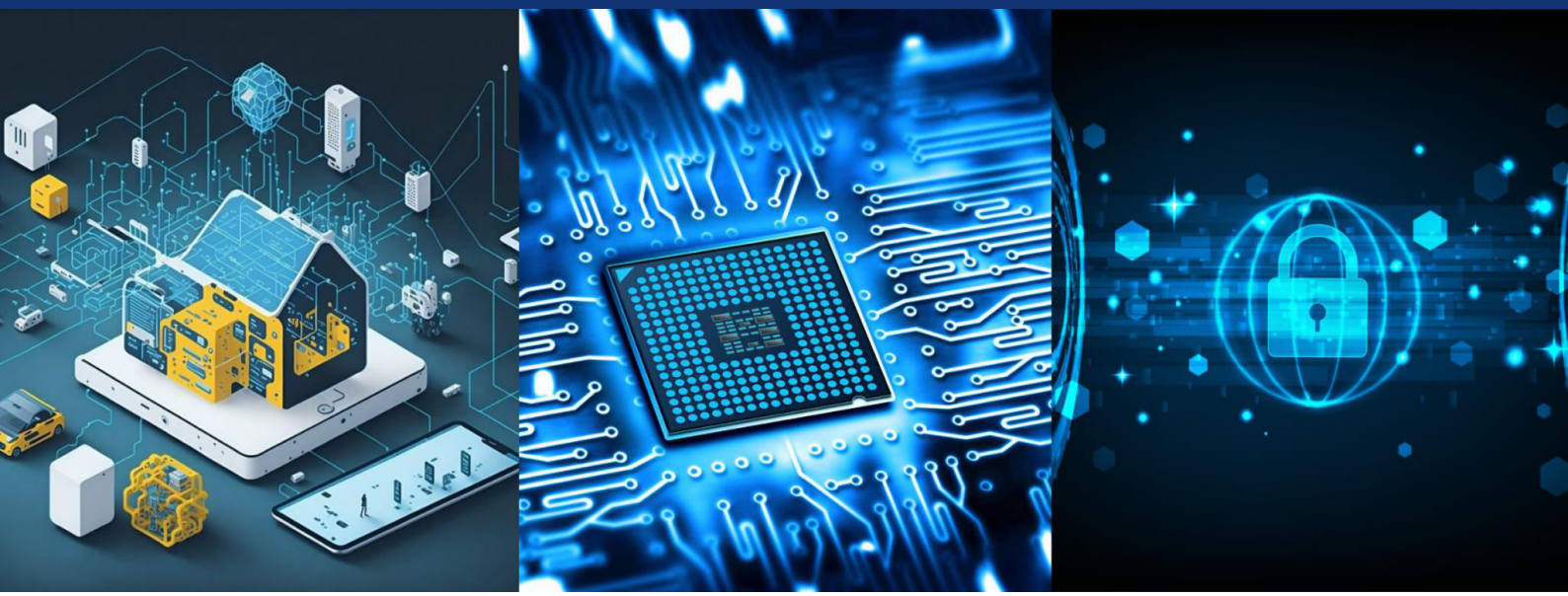
IJIRCCE

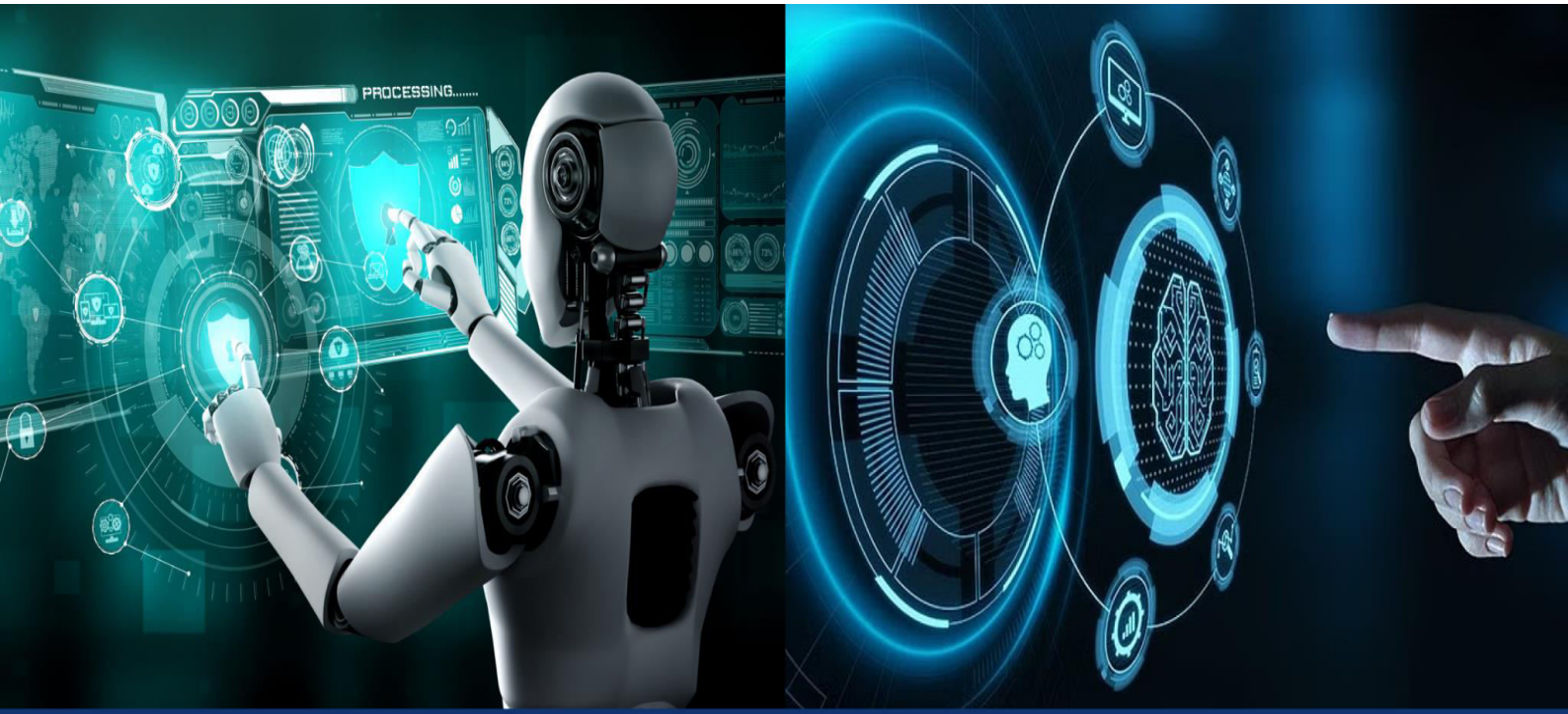
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304052

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Video Summarizer using Transformers

G. Janani Shree[1], S. Shenbaha[2]

Student, Department of Computer Science with Data Analysis, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science with Data Analysis, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India [2]

ABSTRACT: The Video Summarizer Using Transformers project focuses on developing an AI-powered system that can automatically generate concise summaries of video content. This project leverages advanced machine learning techniques, particularly transformer-based models, to simplify and automate the summarization process. The generated summaries provide users with a quick and accurate understanding of the video's content, reducing the need to watch entire videos. transformer models have a token limit, the extracted transcript is divided into segments of 1000 characters each before being fed into the summarization model.

KEYWORDS: Information quick, accurate, meaningful, easy to understand, simplify, time saving.

I. INTRODUCTION

Video Summarizer Using Transformer project develops an AI-powered system that generates concise summaries of video content automatically. With the massive growth in online video data, manually extracting key information from lengthy videos proves inefficient. This project addresses the challenge by utilizing advanced transformer-based machine learning models to automate the summarization process effectively. The system begins by extracting subtitles from YouTube videos using the youtube_transcript_api. These subtitles serve as input data and are divided into manageable 1000-character segments, allowing the model to operate within token limits. Each segment is processed individually using a state-of-the-art pre-trained transformer model from Hugging Face, ensuring meaningful and accurate summaries. This approach allows the system to condense lengthy transcripts into easily digestible content while preserving essential points.

The automated summarization improves accessibility and usability, enabling users to quickly grasp critical information without watching entire videos. This system is particularly useful in fields such as education, research, journalism, and media analysis, where efficient information retrieval is crucial. By eliminating redundant content and structuring summaries for readability, the system provides significant value to diverse user groups, including students, professionals, and researchers. Furthermore, the project includes automatic transcript extraction, allowing users to retrieve video transcripts effortlessly in various languages. The user-friendly interface makes the tool accessible even to individuals with minimal technical expertise. By integrating automation, the system saves time, ensures accuracy, and highlights the transformative role of AI in content consumption. This project showcases the potential of AI in processing large volumes of multimedia data and demonstrates its applications in creating video highlights, summarizing lengthy video content, and improving overall efficiency in media consumption. The "Video Summarizer Using Transformer" establishes itself as a cutting-edge contribution to the growing field of automatic video summarization

II. LITERATURE SURVEY

Youtube Transcript Summarizer, Gousiya Begum , N. Musrat Sultana , Dharma Ashritha, 2022, IJCRT, Volume 10, Issue 6 June 2022, ISSN: 2320-2882, this paper highlights the growing need for efficient video summarization tools, emphasizing AI-powered Chrome extensions using HuggingFace transformers and WebAPI for user-friendly, accurate, and accessible content summarization. Multimodal Video Summarization using Attention based Transformers (MVSAT) Kushagra Singh, Pranav R, Pavan Kumar Nuthi, Nikhil Raju Mohite, Mamatha H R, 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), DOI: 10.1109/IATMSI60426.2024.10503010, IEEE Xplore: 24 April 2024, this research emphasizes advanced video

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304052

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

summarization using attention-based transformers. By utilizing multi-modality approaches, including visual features and captions, it achieves improved F1 scores on SumMe and TVSum datasets, showcasing enhanced efficiency. Learning to Summarize YouTube Videos with Transformers: A Multi-Task Approach, R. Sudhan, D.R. Vedhaviyassh, G. Saranya, 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), DOI: 10.1109/ACCAI58221.2023.10201219, IEEE Xplore: 04 August 2023, this paper explores video summarization using NLP and machine learning, addressing limitations of YouTube API. It employs speech-to-text, Hugging Face models, and transformers, achieving efficient summarization with promising evaluation metrics. Users face inefficiencies summarizing YouTube videos, relying on manual efforts or basic tools with accuracy and accessibility limitations. Current methods lack features like personalization, quick processing, and handling caption-less videos. Many tools are impractical for diverse content, offering plain outputs without interactivity or insights, hindering user experience significantly.

III. PROPOSED METHODOLOGY

3.1 System Overview

The system addresses the challenges users face when summarizing lengthy YouTube video content. It leverages machine learning and natural language processing (NLP) to automate the extraction and summarization of video transcripts, significantly reducing the effort and time required for understanding video content. By breaking down the process into three core stages—transcript extraction, text preprocessing, and AI-driven summarization—the system ensures efficiency while retaining essential information. The end goal is to simplify content consumption by providing users with concise summaries of video content.

3.2 System Design

The system operates using modular architecture that divides tasks into specific components:

1. User Interface Module: Designed to offer a user-friendly interaction, this module allows users to paste a YouTube video URL into the interface, submit the URL for processing, and view results.

2. Transcript Extraction Module: Utilizes the YouTube Transcript API to fetch video transcripts. This ensures that even multilingual videos can be summarized if subtitles are available.

3. Preprocessing Module: Handles cleaning and structuring raw transcripts. Tasks include removing timestamps, eliminating special characters, and dividing text into manageable segments for efficient processing by transformer models.

4. Summarization Module: Implements state-of-the-art transformer-based NLP models, such as BART or T5, for generating high-quality summaries from text.

5. Output Module: Displays the final summarized content in a readable format and optionally allows users to download it as a .txt file.

3.3 Operational Workflow of the Summarization System

The summarization system functions as the core of the system, processing user-provided YouTube URLs by first extracting the video ID (e.g., "A4OmtyaBHFE" from a URL). It then retrieves transcripts using the YouTube Transcript API or, if unavailable, applies speech-to-text conversion via automatic speech recognition models. The retrieved transcripts are cleaned to remove irrelevant elements like timestamps and divided into smaller chunks, typically 1000 characters, for efficient processing. These chunks are fed into pre-trained transformer models, such as BART or T5, which generate concise, context-aware summaries. Finally, the summarized chunks are combined into a coherent and readable format, which is displayed to the user through the frontend interface.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304052

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

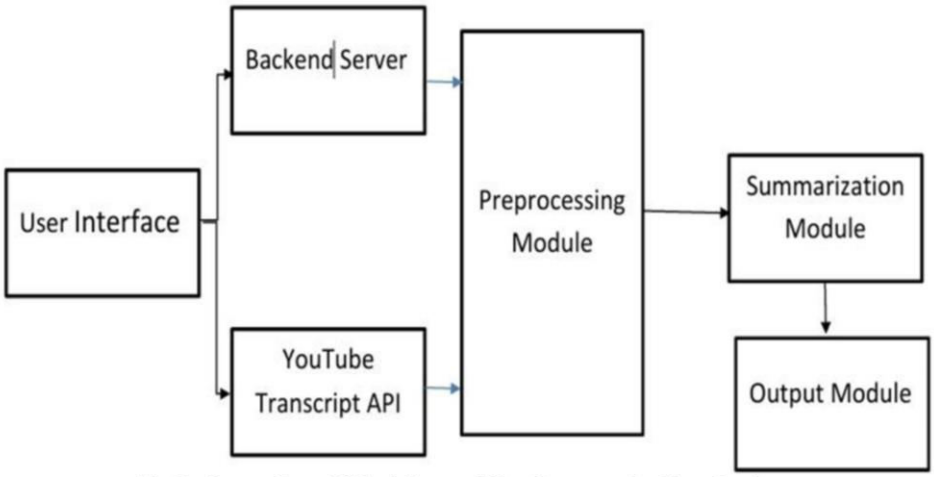


Fig.1. Operational Workflow of the Summarization System

3.4 User Interaction Flow and Interface Design

The system is designed with a strong focus on usability and simplicity, utilizing modern web technologies to ensure an intuitive user experience. It includes a URL input field where users can paste YouTube video links, along with a Submit button to initiate backend processing and transcript retrieval. During processing, visual feedback such as a spinner or progress bar keeps users informed about the ongoing operations. Once the process is complete, the final summary is displayed in a clear and readable format. Additionally, a download option allows users to save the summary locally as a text file for future use.

3.5 Architecture

The Video Summarization System is designed with a modular, client-server architecture to ensure scalability, efficiency, and seamless integration of components. This architecture divides the system into two primary components—the frontend (client) and backend (server)—which interact seamlessly to deliver automated summarization of video transcripts.

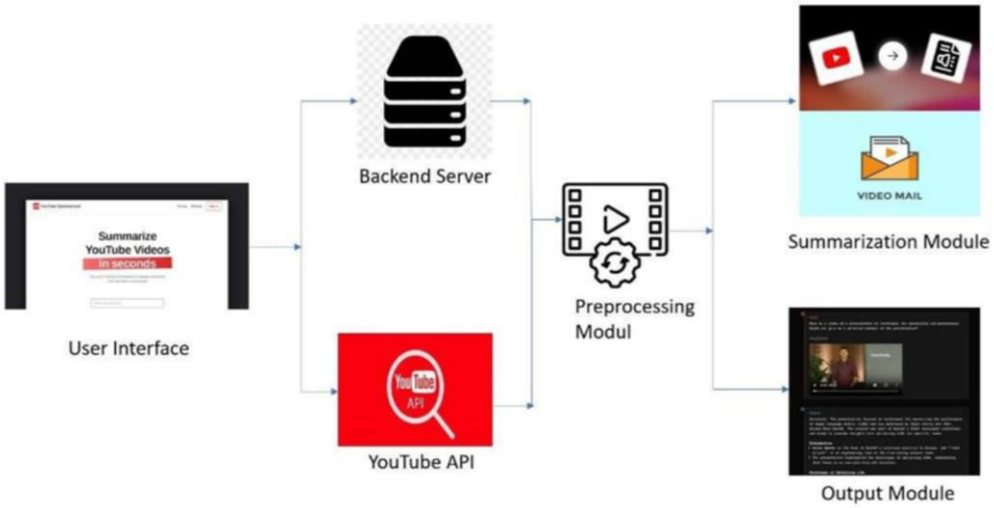


Fig. 2. Architecture

The frontend serves as the user-facing portion of the system, prioritizing accessibility, simplicity, and responsiveness. It allows users to interact with the system by providing input and viewing the output in an easy-to-read format. Users paste the YouTube video URL into a designated input field. Once the "Submit" button is clicked, the frontend sends the video URL to the backend for processing. During backend operations, visual elements such as loading spinners or progress bars provide real-time feedback, keeping users informed about ongoing processes. The finalized summary of the video

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304052

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

transcript is displayed clearly in the frontend, ensuring readability. A downloadable summary in `.txt` format is offered for users seeking offline access to the condensed information. Developed using modern web technologies, the frontend ensures responsiveness across devices and provides an intuitive interface for diverse user groups, including students, professionals, and researchers.

The backend is the operational core of the system, handling data retrieval, processing, and summarization through machine learning and natural language processing (NLP). It is designed to perform complex tasks such as transcript extraction, text cleaning, and summarization efficiently. The backend extracts the unique video ID from the user-provided YouTube URL. For instance, in the URL `https://www.youtube.com/watch?v=A4OmtyaBHFE`, the video ID is identified as `A4OmtyaBHFE`. Using the YouTube Transcript API, the backend retrieves video transcripts when available. For videos lacking transcripts, speech-to-text conversion techniques are employed using automatic speech recognition models to generate textual data. The backend processes the retrieved transcript to remove irrelevant details such as timestamps, special characters, and non-essential text. It divides the cleaned text into manageable chunks (e.g., 1000-character segments) for smoother handling by transformer-based models. Each text chunk is passed through pre-trained AI models, such as BART or T5, which analyze the context and condense the information into concise summaries. These models ensure that critical points from the transcript are preserved while removing redundant information. The individual summaries generated for each chunk are merged to form a coherent and readable final summary, ready to be displayed to the user.

Users interact with the frontend by providing a YouTube video URL in the input field and submitting it for processing. The backend extracts the video ID from the URL, retrieves and preprocesses transcripts, and uses AI models to generate concise summaries. Summarized text chunks are combined into a coherent summary, which is then displayed in the frontend. Users may also download the summary for offline access.
This modular architecture ensures efficient interaction between the frontend and backend, delivering accurate and user-friendly summarization results. By leveraging cutting-edge AI technologies, the system transforms lengthy video transcripts into concise, easy-to-read summaries, greatly enhancing content accessibility.

IV. RESULTS AND DISCUSSION

Setting up the development environment ensures efficient functionality for the YouTube Video Summarization System. Users install Python 3.x and libraries such as `transformers`, `youtube_transcript_api`, and Flask to enable transcript extraction, text summarization, and web app deployment. A virtual environment manages dependencies for compatibility. The backend extracts transcripts using the YouTube Transcript API, preprocesses the text by cleaning and chunking it, and applies Hugging Face transformer models like BART or T5 to generate accurate summaries. The frontend offers a user-friendly interface with features to paste YouTube URLs, trigger summarization, and view results interactively.

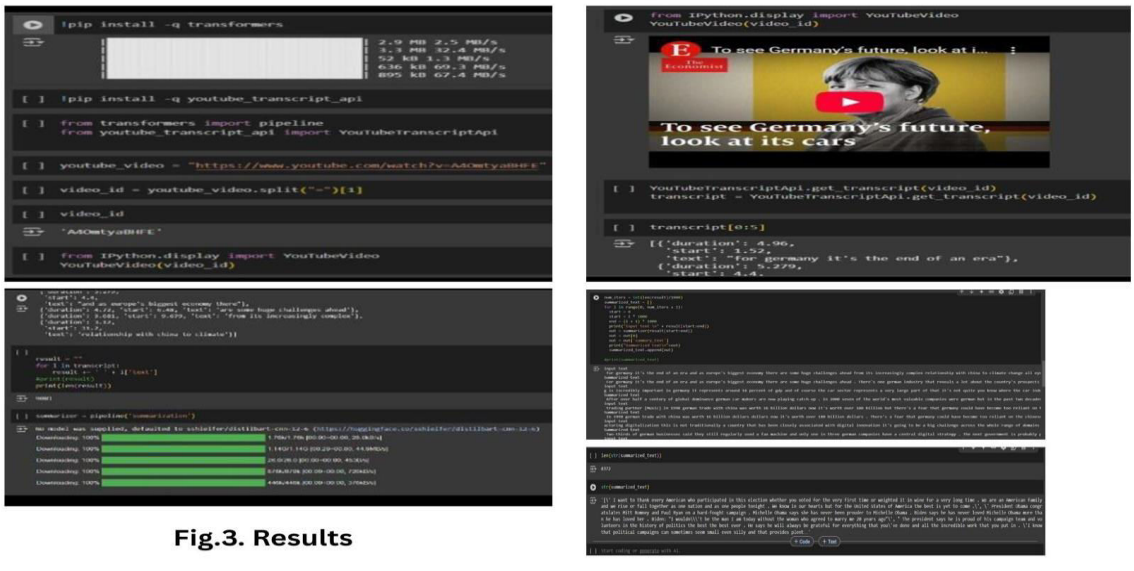


Fig.3. Results

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Testing ensures reliability through unit, integration, and performance tests, while deployment on cloud services such as AWS or Google Cloud ensures scalability, accessibility, and monitoring for real-world use.
Future enhancements include multi-language support with translation APIs, speech-to-text integration using ASR models, and advanced AI models like GPT-based summarizers for improved accuracy and usability.

V. CONCLUSION

The YouTube Video Summarization System stands as an innovative solution for transforming lengthy video transcripts into concise and meaningful summaries. By integrating advanced technologies such as the YouTube Transcript API, Natural Language Processing (NLP), and AI-based models like BART and T5, the system enhances the accessibility and efficiency of video content consumption. It handles large transcripts effectively, ensuring users obtain key information quickly without manually watching entire videos—a process that is often impractical and time-consuming. Comprehensive testing validates the system's accuracy and performance. Unit tests confirm the correctness of individual components, while integration and functional testing ensure seamless module interaction and user satisfaction. Performance evaluations highlight its processing speed, and user acceptance testing (UAT) reinforces its ease of use and utility.

The structured implementation phase ensures the reliability of backend APIs, AI models, and the frontend user interface. Deployment on cloud platforms guarantees scalability, accessibility, and real-world optimization. Future enhancements, such as multi-language support, voice-based interaction, and real-time summarization, expand its versatility and impact. The system's ability to handle missing transcripts, complex video structures, and diverse scenarios ensures consistent accuracy and efficiency. It fulfills user expectations by delivering clean, structured summaries, representing a significant achievement in AI-driven content summarization and accessibility.

REFERENCES

1. Potapov, D., Douze, M., Harchaoui, Z., & Schmid, C. (2014). Category-specific video summarization. In Springer (Ed.), European Conference on Computer Vision (pp. 540-555). [S.l.]Google Scholar
2. Ghauri, J. A., Hakimov, S., & Ewerth, R. (2021). Supervised video summarization via multiple feature sets with parallel attention. In IEEE (Ed.), 2021 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1–6s). [S.l.]: IEEE. Crossref Google Scholar
3. Song, Y., Vallmitjana, J., Stent, A., & Jaimes, A. (2015). Tvsum: Summarizing web videos using titles. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5179-5187). Google Scholar
4. Mubarak, A. A., Cao, H., & Ahmed, S. A. (2021). Predictive learning analytics using deep learning model in MOOCs' courses videos. Education and Information Technologies, 26(1), 371-392. Digital Library Google Scholar
5. Ghauri, J. A., Hakimov, S., & Ewerth, R. (2020). Classification of important segments in educational videos using multimodal features. arXiv preprint arXiv:2010.13626. Google Scholar
6. Oliveira, L. M. R., Busson, A. J. G., Salles, S. N. Carlos de, Santos, G. N. dos, & Colcher, S. (2021). Automatic generation of learning objects using text summarizer based on deep learning models. In SBC (Eds.), Anais do XXXII Simpósio Brasileiro de Informática na Educação (pp. 728-736). [S.l.]. Crossref Google Scholar
7. Alrumiah, S. S., & Al-Shargabi, A. A. (2022). Educational videos subtitles' summarization using latent dirichlet allocation and length enhancement. CMC-Computers Materials & Continua, 70(3), 6205–6221.Crossref Google Scholar
8. Abhilash, R. K., Anurag, C., Avinash, V., & Uma, D. (2021). Lecture video summarization using subtitles. In EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing (pp. 83-92). Springer. Crossref Google Scholar
9. Moraes, L., Marcacini, R. M., & Goularte, R. (2022, November). Video summarization using text subjectivity classification. In Proceedings of the Brazilian Symposium on Multimedia and the Web (pp. 133-141). Digital Library Google Scholar
10. de Souza Barbieri, T. T., & Goularte, R. (2020, November). Investigating Subjectivity Criterion for Multi-video Summarization. In Proceedings of the Brazilian Symposium on Multimedia and the Web (pp. 137-144). Digital

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Library Google Scholar

11. Potapov, D., Douze, M., Harchaoui, Z., & Schmid, C. (2014). Category-specific video summarization. In Springer (Ed.), European Conference on Computer Vision (pp. 540-555). [S.l.]. Google Scholar
12. Ghauri, J. A., Hakimov, S., & Ewerth, R. (2021). Supervised video summarization via multiple feature sets with parallel attention. In IEEE (Ed.), 2021 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1–6s). [S.l.]: IEEE. Crossref Google Scholar
13. Song, Y., Vallmitjana, J., Stent, A., & Jaimes, A. (2015). Tvsum: Summarizing web videos using titles. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5179-5187). Google Scholar
14. Youtube Transcript Summarizer, Gousiya Begum , N. Musrat Sultana , Dharma Ashritha, 2022, IJCRT | Volume 10, Issue 6 June 2022 | ISSN: 2320-2882
15. Multimodal Video Summarization using Attention based Transformers (MVSAT) Kushagra Singh, Pranav R, Pavan Kumar Nuthi, Nikhil Raju Mohite, Mamatha H R, 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), DOI: 10.1109/IATMSI60426.2024.10503010, IEEE Xplore: 24 April 2024.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details