STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 3, 2024
September–December
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

Transcript Summarizer of YouTube Videos Using Deep Learning

Jayesh Chavan[1], Smita Palnitkar[2,*], Ajinkya Kaje[3], Chaitanya Bhopnikar[4]

Abstract

Transcriber Sum is a deep learning-based YouTube transcript summarization tool. It employs advanced machine learning techniques to automatically generate concise summaries of YouTube video transcripts, enabling users to quickly grasp the key content and insights of videos without the need to watch or read the entire transcript. This Transcriber Sum addresses the challenge of providing users with concise and informative summaries of video content by harnessing the power of deep learning techniques. The objectives encompass the creation of advanced models for natural language processing and video content analysis, building a robust framework for processing and summarizing video transcripts, collecting and annotating a diverse dataset of YouTube video transcripts, implementing data preprocessing and augmentation techniques, training and evaluating the model's performance, developing a user- friendly application for real-time transcript summarization, and continuously refining the model based on user feedback. The methodology involves developing deep learning models for natural language understanding, utilizing advanced video content analysis techniques, and seamlessly integrating these components into an application for effortless YouTube transcript summarization. Performance evaluation is conducted through a combination of simulation using synthetic data and real-world testing with a wide range of YouTube video transcripts. Transcriber Sum aims to empower users with a time-efficient and informative tool for quickly summarizing the wealth of content available on YouTube, ultimately enhancing the overall viewing experience and accessibility of valuable information in the digital age.

Keywords: Transcript summarization, deep learning, machine learning techniques, natural language processing, real-time, AI/deep learning

INTRODUCTION

The "Transcriber Sum" project addresses the challenge of navigating the vast volume of content available on YouTube by proposing an automated transcript summarization system. Leveraging cutting-edge technologies, such as deep learning and advanced machine learning algorithms, the project aims to streamline content consumption and knowledge acquisition in today's digital age, where video content consumption is ubiquitous. By recognizing evolving user preferences and behaviors, particularly the demand for instant access to information and dwindling attention spans, the project seeks to bridge the gap between abundant content and meaningful insight extraction. Through the development of sophisticated models for natural language processing and video content analysis as well as the creation of a robust framework for transcript summarization, the project endeavors to empower users across diverse demographics and

***Author for Correspondence**
Smita Palnitkar
E-mail: smita.palnitkar_skncoe@sinhgad.edu

[1,3,4]Student, Department of Electronics and Telecommunication, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India
[2]Assistant professor, Department of Electronics and Telecommunication, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

Received Date: July 03, 2024
Accepted Date: August 06, 2024
Published Date: September 11, 2024

Citation: Jayesh Chavan, Smita Palnitkar, Ajinkya Kaje, Chaitanya Bhopnikar. Transcript Summarizer of YouTube Videos Using Deep Learning. Journal of Artificial Intelligence Research & Advances. 2024; 11(3): 119–126p.

interests. Additionally, by democratizing access to valuable insights within YouTube transcripts, the project contributes to knowledge dissemination and collaboration among researchers, educators, and content creators, enriching the digital landscape and fostering a culture of innovation worldwide.

Natural Language Processing

Natural Language Processing (NLP) is essential for extracting and summarizing textual data from YouTube videos. NLP algorithms enable a system to understand and process human language and transform raw transcripts into concise summaries. These algorithms analyze texts to identify key points, themes, and significant insights. NLP techniques preprocess the transcript by cleaning and standardizing the text and removing extraneous elements such as timestamps and speaker labels. This ensured that the data was ready for further analysis and summarization.

The Bidirectional and Auto-Regressive Transformers (BART) summarization model, which is a state-of-the-art NLP architecture, excels in understanding contextual nuances and generating informative summaries. By segmenting the transcript and feeding it into the BART model, the system produces coherent summaries that capture the essence of the original content, thereby making it easier for users to grasp the main points without watching the entire video.

Transcript Extractor

The transcript extractor is a vital component of the system and is responsible for retrieving and preparing raw transcript data from YouTube videos. It starts by parsing the YouTube link to extract the unique video ID, which is essential for accessing video data. Using the YouTube Transcript API, the extractor requests the raw transcript, which often contains non-textual elements, such as timestamps and speaker labels, that require removal. During the extraction, the transcript was cleaned and standardized to retain only the relevant textual content. This refined transcript forms the foundation for subsequent segmentation and summarization, ensuring that the system can effectively process and summarize information for the user.

Transcript Summarizer

Transcript summarizers are key components that are designed to condense extensive video transcripts into concise and coherent summaries. After the transcript has been cleaned and segmented, the summarizer processes each segment using the BART model, a cutting-edge neural network architecture known for its ability to understand contextual nuances.

BART excels in generating informative summaries by distilling key points and essential concepts from each transcript segment. Once all the segments have been summarized, the individual summaries are aggregated to create a comprehensive overview of the entire video. This process ensures that users can quickly grasp the main points and obtain significant insights from the video without needing to watch the entire content, making it an efficient tool for extracting essential information.

LITERATURE SURVEY

Ilampiray et al. [1] proposed an NLP-based solution using bidirectional encoder representations from transformers (BERT) Summarization to extract concise summaries from online videos, thereby emphasizing their significance in education and entertainment. This study surveys techniques, including linear discriminant analysis (LDA) and neural extractive summarization models. It introduced a methodology involving URL processing, video-to-text extraction, NLP, and bidirectional encoder representations from transformers (BERT) Summarization. By focusing on clean and accurate YouTube video summaries, this study aims to save users' time and enhance information retrieval by utilizing Python libraries, YouTube Transcript API, Google Translate API, and the Flask framework. Visual aids and discussions on future enhancements complement the study, addressing the need for efficient video content consumption.

Dharmapuri et al. [2] introduced a Chrome Extension, the YouTube Video Summarizer, which utilizes NLP techniques to generate quick video summaries from English-language transcripts. This tool aims to simplify content evaluation, enabling users to assess relevance without requiring extensive viewing time. The methodology highlights Python's role, machine learning (ML), and NLP in transcript summarization, with empirical results showing model accuracy. The framework offers a solution to efficiently summarize YouTube content for quick insight, benefiting users seeking to save time while consuming educational content.

Albeer et al. [3] from the University of Kerbala, Iraq, explored automated text summarization for YouTube videos. Utilizing the term frequency-inverse document frequency (TF-IDF) method, lengthy videos were condensed into concise summaries, aiding individuals with limited time. This study surveys various summarization techniques, proposing a system involving video transcription extraction, text preprocessing, keyword extraction using TF-IDF, and final summary generation. The evaluation of the CNN-daily mail-master dataset demonstrated the effectiveness of the method, emphasizing its utility in accessing relevant information quickly within lengthy videos.

Devi et al. [4] proposed a user interface to generate concise summaries of YouTube videos using NLP and ML. Addressing the challenge of finding relevant content amidst the vast daily uploads on YouTube, this study introduces an abstractive summarization model. The existing methods highlighted include TF-IDF, LSA, and the hugging-face transformer, emphasizing the advantages of summarization in saving user time. The proposed system offers features such as text cleaning, summarization, grammar checks, translation, and email sharing, with future directions focusing on language translation expansion and its applicability in other streaming services.

It summarizes techniques tailored for YouTube videos, addressing the challenge of time-consuming content consumption. It emphasizes YouTube's role in education and information dissemination, detailing a methodology using a HuggFace transformer coupled with Python APIs for subtitle retrieval and summarization. This study reviews various summarization approaches, including extractive and abstractive methods, highlighting the utility of hugging-face transformers in NLP tasks. Evaluation metrics such as recall-oriented understudy for gisting evaluation (ROUGE) demonstrate the effectiveness of the proposed method in reducing text size while preserving essential information, thereby offering a streamlined solution for YouTube video summarization. Previous studies have investigated various factors and methodologies to enhance the efficiency of YouTube transcript summarization. These endeavors have explored the integration of diverse NLP techniques, leveraging different Python libraries and machine learning algorithms to improve summarization accuracy and reduce false positives. Moreover, they have implemented robust alarm systems as a pivotal component, ensuring timely alerts to users and facilitating appropriate actions based on the summarized content.

SYSTEM ARCHITECTURE

The proposed model comprises two main components, Transcript Extractor and Summarizer.

Transcript Extraction

The YouTube_transcript_API library simplifies the process of retrieving transcripts from YouTube videos by abstracting the complexity of API requests and responses [5–7]. This allows developers to easily integrate YouTube transcript retrieval into Python applications, offering features such as retrieving full transcripts, captions in specific languages, and graceful error handling, as shown in Figure 1.

In addition, it provides methods for cleaning and preprocessing transcript data, making it suitable for further analysis or processing. This specialized tool is designed for efficient interaction with the YouTube API, enabling seamless extraction and refinement of video transcripts.

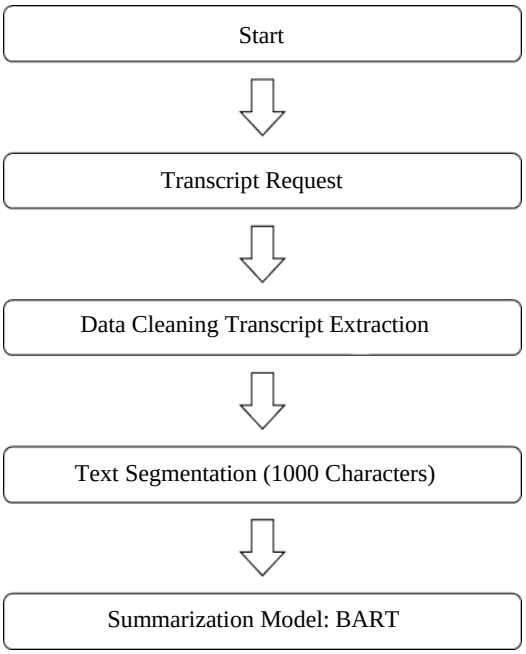

Figure 1. Transcript extraction.

Text Summarization (BART)

BART is a state-of-the-art model developed by Facebook AI and leveraged by hugging-face transformers for precise and efficient text summarization. This model combines bidirectional and auto-regressive techniques, capturing contextual dependencies and semantic relationships with its encoder, similar to BERT, and generating coherent output sequences with its decoder, akin to GPT [8]. BART's training involves a denoising autoencoder objective, enabling it to reconstruct the original text from corrupted versions, making it highly robust in handling noisy input data, as shown in Figure 2.

The pre-training and auto-regressive training phases help BART capture general language patterns and generate contextually relevant sequences. With approximately 140 million parameters, BART surpasses models such as bidirectional encoder representations from transformers (BERT) and GPT-1 in performance, effectively tackling tasks such as text summarization, machine translation, and question answering. Its versatility extends to applications such as text generation, document classification, and the creation of domain-specific models through fine-tuning, making it a valuable tool in diverse NLP domains [9, 10].

METHODOLOGY

- *User:* A user finds a relevant YouTube video and initiates transcript extraction on a web browser extension, as shown in Figure 3.
- *ID extraction:* The system extracts a unique video ID from a YouTube link.
- *YouTube_Transcript_API:* The system uses the video ID to request raw transcripts from YouTube API.
- *Transcript extraction:* The raw transcript is cleaned by removing timestamps, speaker labels, and non-textual elements.
- *Segmentation:* The cleaned transcript is divided into smaller segments for easier summarization.
- *BART summarizer:* Each segment is summarized using the BART neural network model, which distills the key points and essential concepts.
- *Video transcript summary:* The system aggregates individual segment summaries into a comprehensive overview of the video content.
- *User review:* The synthesized summary is presented to the user, who efficiently integrates the extracted insights into his research.

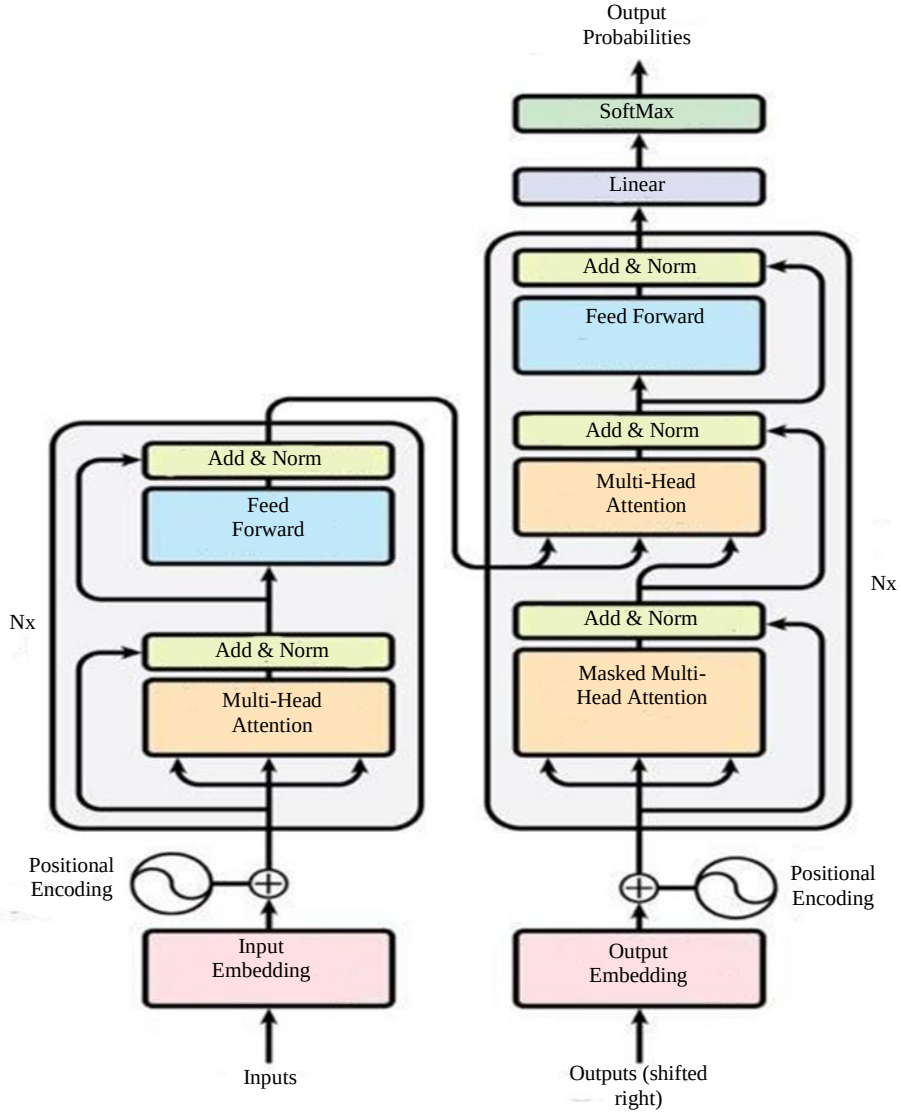

Figure 2. Text summarization (BART).

IMPLEMENTATION AND RESULT

The project aimed to condense lengthy YouTube video transcripts into concise summaries using NLP techniques. By leveraging the transformer library and YouTube _transcript_api, the code effectively processed the transcripts from the provided video link. Upon successful execution, the code generated summarized text segments based on a 1000-character limit per iteration.

The summarization pipeline employed a progressive approach to creating coherent and condensed representations of video content. The subsequent sections present the summarized text segments derived from the transcript, providing an accessible overview of the video's key points and discussion. This functionality exhibited promising results, effectively summarizing the content while maintaining relevance and coherence.

As shown in Figure 4, the "YouTube Summarizer" extension is pinned in the browser's extension panel. Clicking on its logo opens a window titled "YouTube Transcript Summarizer" with a "Summarize" Summarize' button. To use it, open a YouTube video and click "Summarize." The extension retrieves the video's transcript, processes it, and displays a concise summary, allowing for a quick understanding of the video's main points without watching the entire content.

The summarization process is initiated by clicking the button, as shown in Figure 5, which varies in duration based on the transcript length and size. It is necessary to stay on the same window because changing it will cause the popup window to disappear.

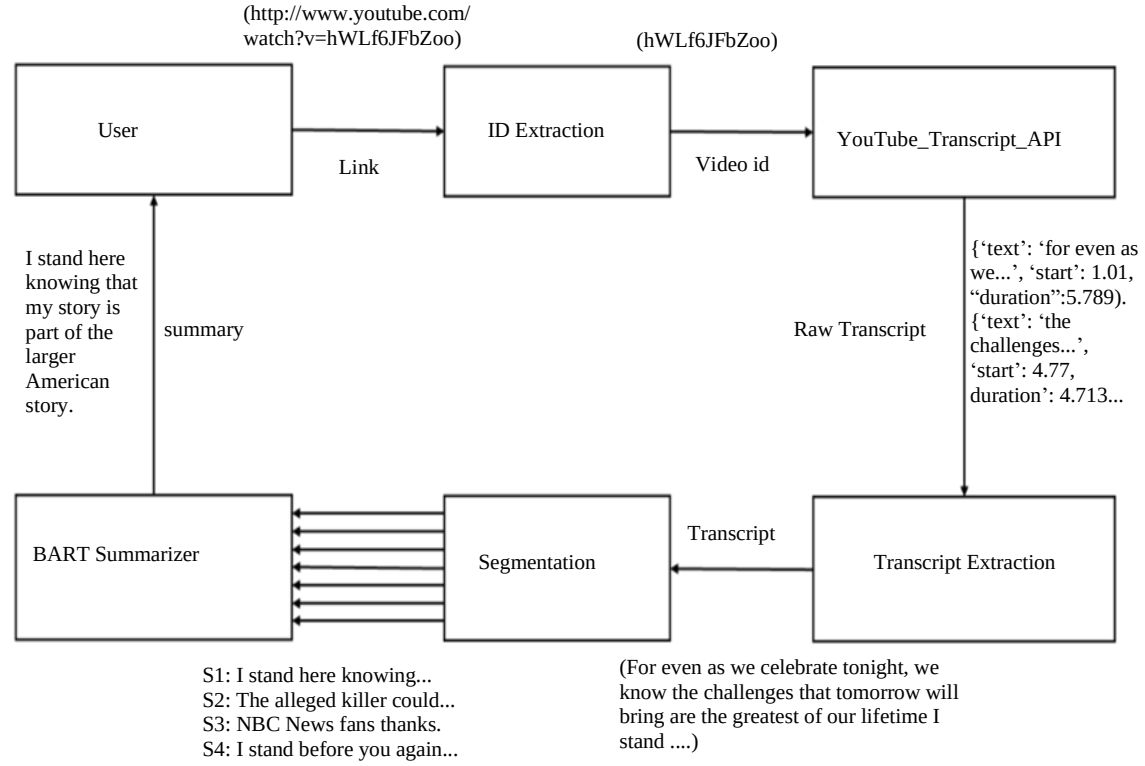

Figure 3. Block diagram of the proposed methodology.

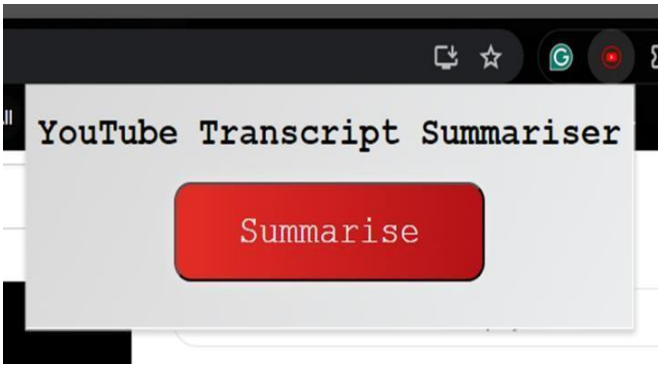

Figure 4. Extension interface button.

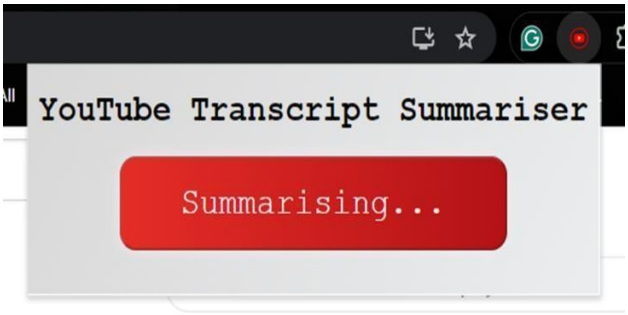

Figure 5. Extension interface button running.

The final summary of the video is displayed below the "Summarize" button. Users can read the summary shown in Figure 6.

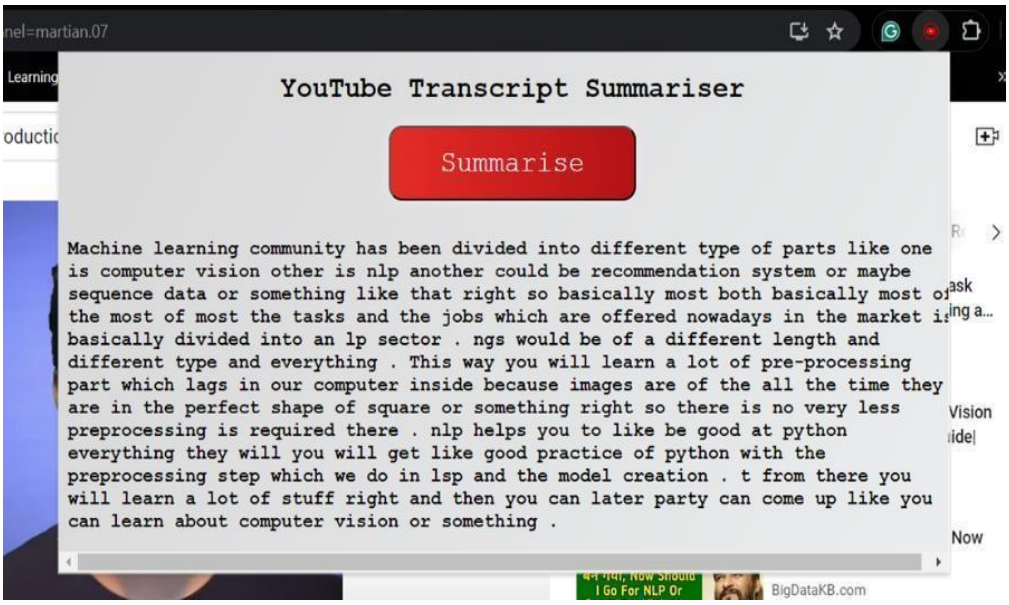

Figure 6. Extension interface button running.

Table 1. Word comparison after before and after summarization.

S.N.	Video title	Duration	Transcript length (number of words)	Summary length (number of words)
1	President Barack Obama's Greatest Speeches | NBC News	3	424	189
2	To see Germany's future, look at its cars00	11	1638	578
3	Steve Jobs' 2005 Stanford Commencement Address	15	2340	829
4	Adolf Hitler: Speech at Krupp Factory in Germany (1935)	1	54	48
5	The HU - Yuve Yuve Yu (Official Music Video)	5	226	99
6	Sadhguru TOP 7 Teachings and Wisdom on Life | Sadhguru Best Speech	18	2524	796
7	Barack Obama's Victory Speech Full - Election 2012	20	2192	805
8	Geography Now! GEOLANDIA	5	1089	418
9	Demon Slayer Explained in ONLY 10 Minutes	10	2291	914
10	The Dark Side of Engineering	7	2106	745
11	2am - Based on a True Incident - Horror Short Film	32	706	225
12	Why Genghis Khan Refused to Invade India	22	3394	1031
13	Exploring Deep History with Raj Vedam: Volcanic Eruptions, Meteor Strikes, Galaxies and More | TRS 372	73	13225	4302
14	Geography Now! UNITED STATES OF AMERICA	65	13364	4417
15	Romeo and Juliet | Full Movie | Classic Romance Drama | Complete Mini Series	201	15309	4779

RESULT

Table 1 shows a word comparison before and after summarization.

- The shortest video had the smallest transcript and summary lengths with 54 and 48 words, respectively.
- The longest video had the largest transcript and summary lengths with 15,309 and 4,779 words, respectively.
- Longer videos generally have longer transcripts and summaries.
- Summary length does not always correlate directly with video duration or transcript length.
- Summary length relative to transcript length varies, indicating differences in detail retention.
- Complex topics tend to have longer transcripts and summaries.
- Summary length reflects content complexity and information density.
- Some short videos have long transcripts, requiring concise summarization techniques.

CONCLUSION

This study demonstrates the effectiveness of NLP models in summarizing extensive video transcripts into coherent summaries using advanced transformer techniques and the YouTube Transcript API. It successfully condensed a 2301-word transcript to 1047 words, highlighting the efficiency of the model. Future improvements could enhance contextual understanding and summarization accuracy with potential applications extending beyond YouTube to journalism, research, and content curation. This underscores the transformative potential of NLP and machine learning in simplifying and extracting essential insights from vast textual data, thereby contributing to the development of efficient information-processing tools.

REFERENCES

1. Ilampiray P, Thilagavathy A, Nithin AS, Raj I. Video transcript summarizer. E3S Web Conf. 2023;399:04015. DOI: 10.1051/e3sconf/202339904015. Published online 2023 Jul 12.
2. Dharmapuri S, Desu S, Alladi K, Gummadi H, Gupta H, Shareef SNM. An automated framework for summarizing YouTube videos using NLP. E3S Web Conf. EDP Sciences. 2023;430. DOI: 10.1051/e3sconf/202343001056.
3. Albeer RA, Al-Shahad HF, Aleqabie HJ, Al-shakarchy ND. Automatic summarization of YouTube video transcription text using term frequency-inverse document frequency. Indones J Electr Eng Comput Sci. 2022;26:1512–9. DOI: 10.11591/ijeecs.v26.i3.pp1512-1519.
4. Devi S, Nadar R, Nichat T, Lucas A. Abstractive summarizer for YouTube videos. In: International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022). Atlantis Press; 2023. p. 431–8.
5. Ilampiray P, Thilagavathy A, Nithin AS, Raj I. Video transcript summarizer. E3S Web Conf. 2023;399:04015. DOI: 10.1051/e3sconf/202339904015. Published online 2023 Jul 12.
6. Li H, Zhu J, Ma C, Zhang J, Zong C. Read, watch, listen, and summarize: Multi-modal summarization for asynchronous text, image, audio and video. IEEE Trans Knowl Data Eng. 2018;31:996–1009. DOI: 10.1109/TKDE.2018.2848260.
7. Emad A, Bassel F, Refaat M, Abdelhamed M, Shorim N, AbdelRaouf A. Automatic video summarization with timestamps using natural language processing text fusion. In: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC). IEEE Publications; 2021. p. 0060–6.
8. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; 2020. p. 38–45. DOI: 10.18653/v1/2020.emnlp-demos.6.
9. Jain A, Kulkarni G, Shah V. Natural language processing. Int J Comput Sci Eng. 2018;6:161–7. DOI: 10.26438/ijcse/v6i1.161167.
10. Joseph SR, Hlomani H, Letsholo K, Kaniwa F, Sedimo K. Natural language processing: A review. Int J Res Eng Appl Sci. 2016;6:207–10.

