

SITS/IT/Projects/UG/2024-25/G10

A PROJECT REPORT ON

AUTOMATED VIDEO TRANSCRIPT SUMMARIZER

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY,
PUNE IN THE FULFILLMENT OF THE REQUIREMENTS FOR THE
AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING (INFORMATION TECHNOLOGY)

SUBMITTED BY

Agrawal Khushbu	Exam No.: B1905708501
Gajare Krushna	Exam No.: B1905708516
Pisal Pranav	Exam No.: B1905708538
Wagh Manasi	Exam No.: B1905708555

UNDER THE GUIDANCE OF

Prof. A. A. Yadav



Sinhgad Institutes

DEPARTMENT OF INFORMATION TECHNOLOGY
SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE
PUNE, 411041

SAVITRIBAI PHULE PUNE UNIVERSITY
Academic Year: 2024-25



C E R T I F I C A T E

This is to certify that *Ms Agrawal Khushbu Exam No B1905708501*, *Mr Gajare Krushna Exam No B1905708516*, *Mr. Pisal Pranav Exam No B1905708538*, *Ms Wagh Manasi Exam No B1905708555* have successfully completed the Project Stage-I entitled **AUTOMATED VIDEO TRANSCRIPT SUMMARIZER** under my supervision, in the partial fulfillment of Bachelor of Information Technology of Savitribai Phule Pune University.

Date :

Place :

Prof. A. A. Yadav
Guide

Prof. A. A. Yadav
Head of Department

External Examiner

-
-

Dr. S. D. Markande
Principal

ACKNOWLEDGEMENT

It has really been an exciting and a prolonged experience to work on this project. We are lucky to get invaluable contributions from all the people around us. We express our deepest and sincere gratitude to our project guides Prof. A. A. Yadav and Prof. S. R. Mane for her thorough guidance, constant availability and motivation to work harder. This project would not have been completed without his valuable insights and suggestions.

We are very much thankful to Prof A. A. Yadav , Head, Department of Information Technology, Dr. S. D. Markande, Principal and Dr.V. M. Rohkule , Vice principal, Sinhgad Institute of Technology and Science, Narhe for their help, support and co-operation during this project work.

We would also like to thank the Sinhgad Technical Educational Society for providing access to the institutional facilities for our project work.

Date:
Place: Pune

Agrawal Khushbu
Gajare Krushna
Pisal Pranav
Wagh Manasi

ABSTRACT

In today's digital age, the consumption of video content is omnipresent, with platforms like YouTube offering a vast repository of informative and entertaining videos. However, the time required to watch lengthy videos can often be a headache for busy individuals. There is a need for an efficient solution to extract concise and coherent summaries from the transcripts of lengthy video content, allowing users to quickly grasp the key information without the time investment of watching the entire video. This issue is particularly relevant for online educational content, lectures, and other informational videos. In the pursuit of an innovative solution, this paper proposes the system that takes the URL of a YouTube video as input and generates a concise and informative summary as output. The proposed solution aims to not only save time for viewers but also enhance accessibility, particularly for individuals with hearing disabilities who can benefit from text-based video summaries. It seeks to contribute to the ongoing discourse on making video content more accessible and time-efficient. To address this issue, this paper presents a survey and analysis of video transcript summarization using Natural Language Processing (NLP) techniques.

List of Abbreviations

NLP	Natural Language Processing
AI	Artificial Intelligence
API	Application Programming Interface
ASR	Automatic Speech recognition
JSON	JavaScript Object Notation
XML	Extensible Markup Language
RBAC	Role-based access controls
HTML	Hypertext Markup Language
CSS	Cascading Stylesheet
UI	User Interface
NLTK	Natural Language Toolkit
IDE	Integrated Development Environment
REST	Representational State Transfer
RAM	Random Access Memory
GPU	Graphics Processing Unit
ORM	Object-Relational Mapping
SEO	Search Engine Optimization
AVTS	Automated video transcript summarizer

List of Figures

3.1	Iterative Model	18
4.1	System Architecture	19
4.2	DFD level 0	20
4.3	DFD level 1	21
4.4	DFD level 2	22
4.5	Activity Diagram	23
4.6	Result 1	24
4.7	Result 2	25
5.1	Activity Sheet	28

Contents

Certificate	i
List of Abbreviations	iv
List of Figures	v
1 Introduction	1
1.1 Overview	1
1.2 Motivation	2
1.3 Problem Definition and Objectives	3
1.4 Project Scope	3
1.5 Methodologies of problem solving	4
2 Literature Survey	5
2.1 Literature Review	5
2.2 Summary of Literature Review	9
3 Software Requirement Specification	14
3.1 Assumptions and Dependencies	14
3.1.1 Assumptions	14
3.1.2 Dependencies	15
3.2 Functional Requirements	15
3.2.1 System Feature 1: Video Input Handling	15
3.2.2 System Feature 2: Summarization Algorithm	15
3.2.3 System Feature 3: User Interfaces	15
3.2.4 System Feature 4: API Interfaces	16

3.2.5	System Feature 5:Language Support	16
3.3	System Requirements	16
3.3.1	Hardware Requirements	16
3.3.2	Software Requirements	16
3.4	Analysis Models: SDLC model to be Applied	16
4	System Design	19
4.1	System Architecture	19
4.2	Data flow diagrams	20
4.3	Activity Diagram	23
4.4	System Implementation	24
5	Project Plan	26
5.1	Project Task Set	26
5.2	Risk Management	27
5.2.1	Risk Identification	27
5.2.2	Risk Analysis	27
5.2.3	Risks	28
5.3	Project Schedule	28
5.4	Team Organization	29
5.4.1	Team Structure	29
6	CONCLUSIONS	30
6.1	Conclusions	30
6.2	Future Work	31
6.3	Areas of Application	31
	Appendix A	32
	REFERENCES	33

Chapter 1

Introduction

This chapter discusses the background, relevance and the motivation behind the project.

1.1 Overview

The "AUTOMATED VIDEO TRANSCRIPT SUMMARIZER " project endeavors to develop an automated system that harnesses natural language processing (NLP) techniques to extract transcripts from video content and generate succinct summaries of the extracted text. Through the utilization of APIs and NLP libraries such as NLTK or spaCy, the system aims to process video transcripts efficiently, condensing them into concise summaries that capture essential information and main points. With potential integration of translation services for multi-language support and a user-friendly interface, the system seeks to enhance accessibility to video content and facilitate productivity in information consumption across various domains, including education, professional training, content creation, accessibility, research, legal, journalism, and media monitoring. Rigorous testing methodologies and user feedback evaluation will ensure the system's functionality, accuracy, and usability, ultimately culminating in a valuable tool that enables users to extract key insights and information from video content effortlessly. The "Automated Video Transcript Summarizer" project addresses a critical need in today's digital landscape, where video content has proliferated across various platforms. The sheer volume of videos

poses a challenge for users who seek relevant information but are deterred by the time-consuming nature of watching lengthy content. This project leverages advanced Natural Language Processing techniques as a transformative solution. By employing sophisticated algorithms, the system can automatically distill key insights from video transcripts, providing users with concise summaries. This not only streamlines the information extraction process but also empowers users to make efficient use of their time, unlocking the potential for more widespread engagement with valuable content.

The "Automated Video Transcript Summarizer" project addresses a critical need in today's digital landscape, where video content has proliferated across various platforms. The sheer volume of videos poses a challenge for users who seek relevant information but are deterred by the time-consuming nature of watching lengthy content. This project leverages advanced Natural Language Processing techniques as a transformative solution. By employing sophisticated algorithms, the system can automatically distill key insights from video transcripts, providing users with concise summaries. This not only streamlines the information extraction process but also empowers users to make efficient use of their time, unlocking the potential for more widespread engagement with valuable content.

1.2 Motivation

The motivation behind developing an Automated Transcript Summarizer using Natural Language Processing (NLP) stems from several challenges faced by users in the digital age. With the overwhelming amount of video content available online, extracting relevant information has become increasingly difficult, leading to information overload. Manual video consumption is time-consuming and often clashes with the busy schedules of users who need quick access to key insights. Traditional methods of watching videos and manually extracting information are inefficient, labor-intensive, and fail to provide timely summaries. The automated summarizer aims to solve these problems by streamlining the process, enabling users to focus on analyzing key takeaways instead of watching entire videos. By leveraging NLP for summarization, the system enhances productivity, making it a practical solution to the growing need for efficient and

effective information retrieval in the digital era.

1.3 Problem Definition and Objectives

In the modern digital era, the abundance of online video content has created a flood of information, making it difficult for users to efficiently extract insights from lengthy videos. Manual extraction is time-consuming and ineffective, prompting a demand for automated solutions. By utilizing Natural Language Processing (NLP), automated video transcript summarization offers a swift way for users to identify crucial information and patterns without watching the entire video.

The objectives are discussed below:

- To Implement advanced Natural Language Processing (NLP) techniques to automatically generate concise summaries from video transcripts.
- To Incorporate automated translation capabilities to facilitate summarization of video transcripts in different languages .
- To Design an intuitive and user-friendly interface that allows easy input of videos and retrieval of summarized content.

1.4 Project Scope

The project scope for the "AUTOMATED VIDEO TRANSCRIPT SUMMARIZER " involves the development of a comprehensive system capable of automating the extraction of transcripts from video content and generating concise summaries using natural language processing (NLP) techniques. The scope includes the implementation of functionalities such as video input processing, transcript extraction, text summarization, and optionally, translation into multiple languages. The system aims to cater to various user needs, including educational, professional, and research-oriented requirements, by providing an efficient and accessible tool for summarizing video content. The scope encompasses the integration of relevant APIs, NLP libraries, and potentially a user-friendly interface to facilitate ease of use and accessibility.

1.5 Methodologies of problem solving

Requirement Analysis:

Begin by thoroughly understanding the requirements and objectives of the project. This involves identifying the target audience, defining user needs, and understanding the key functionalities and features required in the system.

Research and Exploration :

Conduct comprehensive research on existing NLP algorithms, libraries, and APIs that can be utilized for video transcript summarization. Explore various approaches to transcript extraction, text summarization, and multi-language translation to determine the most suitable methodologies for the project.

Prototyping and Iterative Development:

Adopt an iterative development approach, starting with prototyping and gradually refining the system based on feedback and evaluation. Develop minimum viable products (MVPs) to test core functionalities and gather user feedback early in the development process.

Agile Methodology:

Implement Agile methodologies such as Scrum or Kanban to manage the project's development process effectively. Break down tasks into manageable chunks, prioritize features based on user needs, and regularly review and adapt the project plan to accommodate changes and feedback.

Collaboration and Communication:

Foster open communication and collaboration among team members, stakeholders, and potential users throughout the project lifecycle. Regular meetings, discussions, and feedback sessions facilitate idea sharing, problem-solving, and alignment of goals and expectations.

Chapter 2

Literature Survey

2.1 Literature Review

“YouTube Transcript Summarizer Using Flask and NLP.” *Journal of Positive School Psychology* 6, no. 8 (2022): 1204-1209 From [1], Author employed a Flask-based backend server powers a YouTube transcript summarization system with user-friendly features like translation, speech synthesis, downloads, and email sharing. It retrieves YouTube video subtitles, uses a default CNN model for summarization, and leverages NLP techniques. This practical project aligns with NLP and text summarization research, enhancing content consumption efficiency and sharing.

“AI-based Video Summarization using FFmpeg and NLP.” *International Journal of Innovative Science and Research Technology* Issue- 4, April – 2023. From [2], the author conducted a comprehensive research review on video summarization techniques. This survey covered both supervised and unsupervised methods, encompassing neural networks, attention mechanisms, and SVMs for video segmentation and summarization. FFmpeg emerged as a versatile tool for video processing, while Natural Language Processing (NLP) techniques proved crucial for audio transcript analysis. AssemblyAI’s ASR and NLP services were recognized for their role in precise transcription. The study concluded with an ablation study highlighting the importance of combining FFmpeg, NLP, and video segmentation for effective video summarization.

“SURVEY PAPER ON YOUTUBE TRANSCRIPT SUMMARIZER.” International Research Journal of Modernization in Engineering Technology and Science Issue:04/April 2023. From [3] author, it is evident that text summarization of YouTube videos benefits from NLP libraries like SpaCy and gensim. PyTube and Hugging Face’s transformers are employed for real-time audio extraction and transcript generation, while cosine similarity enhances text summarization. The proposed chrome extension offers users an efficient means to access video summaries and identify objectionable content, improving their viewing experience.

“Youtube Transcript Summarizer”, International Journal of Research in Engineering and Science (IJRES) Issue 5 — May 2023. From [4], the author emphasizes leveraging spoken text and context in video summarization, particularly for speeches. They propose an NLP-based model with API integration for YouTube transcript summarization, addressing coherence challenges. NLP’s role in enhancing computer understanding of human language and applications in customer service is highlighted. The project involves transcript retrieval, HuggingFace transformers for summarization, and Chrome extension development. Future directions include multi-lingual support and improved evaluation metrics.

“Video Summarization using NLP”, International Research Journal of Engineering and Technology (IRJET) Issue: 08 — Aug 2021. From [5], the author presents a video summarization system using NLP-based Latent Semantic Analysis (LSA) for YouTube content. Unlike complex alternatives, this system relies on YouTube’s transcript API and Pytube for efficient subtitle extraction and video retrieval. LSA identifies and ranks key sentences for summarization, which are then adjusted to meet user-set time constraints. The final video summary is generated using the MoviePy library. This innovative approach addresses the rising demand for streamlined video content management in the era of online video consumption

“ Survey on Abstractive Transcript Summarization of YouTube Videos”, International Journal of Advanced Research in Science, Communication and Tech-

nology (IJARSCT) Issue 2, April 2022. From [6], the author introduces extractive and abstractive methods for automatic text summarization. Extractive focuses on key sentences, while abstractive creates new ones. Techniques like tree-based, template-based, rule-based, and graph-based methods are used for extractive, and abstractive methods use semantic or graph-based models like PEGASUS. The text provides a comprehensive overview of these summarization strategies.

“Attentive and adversarial learning for video summarization.” In 2019 IEEE Winter Conference on applications of computer vision (WACV), pp. 1579-1587. IEEE, 2019 From [7], the author proposed an adversarial training framework for semi-supervised video summarization. Their approach utilizes an attention-based pointer network (PtrNet) as the generator, effectively addressing the issue of inconsistent lengths between training and test data. The discriminator, a 3D CNN classifier, determines the quality of summarization fragments. This framework operates as a generative adversarial network (GAN) and demonstrates state-of-the-art performance on benchmark datasets. Notably, the method leverages unpaired summarization fragments from the web, eliminating the need for additional original-versus summarized video pair

“Real time video summarization on mobile platform.” In 2017 IEEE International Conference on Multimedia and Expo (ICME), pp. 1045-1050. IEEE, 2017 From [8], the author emphasizes leveraging spoken text and context in video summarization, particularly for speeches. Author has used MoviePy library for transcript summarization and also used the google Transcribe for the purpose of transcript generation. They have used text rank algorithm in this process. Also, they have provided the average time of video and generated transcript for particular video representing the length of generated summary for short and long videos.

“Automated video summarization using speech transcripts.” In Storage and Retrieval for Media Databases 2002, vol. 4676, pp. 371-382. SPIE, 2001. From [10], the author presents a video summarization method utilizing automatic speech recognition transcripts. Segments are created based on pauses, with

scores derived from word and bigram frequencies. High-score-to-duration ratio segments are chosen for generating informative summaries that maximize coverage. A user study confirms the superiority of the

proposed algorithm over others. The paper also examines diverse video summarization techniques, encompassing playback speed, frame clustering, domain knowledge, and closed captions approaches. It discusses statistical and natural language processing-based methods for summarization.

“YOUTUBE TRANSCRIPT SUMMARIZER”, International Journal of Creative Research Thoughts. Volume 10 Issue 6 June 2022. From [10], the author focuses on building a chrome extension that is a small and easy to use tool for users who watch YouTube videos on browser. There’s no need to change the tab because summary of transcript can be generated in place with popup window. Abstractive text summarization is used for transcript summarization that improves the quality of the summary generated.

“YouTube Transcript Summarizer”, “International Journal of Science and Research (IJSR)”, Issue -2, February 2023 From [11] author, in the research survey reveals several key areas for future development in YouTube transcript summarizers. These include enhancing natural language processing techniques for improved text analysis, advancing summarization algorithms by incorporating contextual data, extending multilingual support, developing personalized summarization approaches, and integrating these tools with various platforms for more tailored user experiences. The landscape of YouTube transcript summarizers offers diverse opportunities for refinement and expansion, promising a dynamic and extensive scope for future research and innovation .

“YOUTUBE TRANSCRIPT SUMMARIZER”, INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS), Vol. 03, Issue 05, May 2023. From [12], the author presents a video transcript summarization tool impowered by Natural Language Processing techniques. Author has used the Hugging Face Library for the text summarization process that focuses on the Abstractive summarization model. The similarity of human on generating summary like using different words and

generating new sentences from existing sentences matches with the working of Abstractive summarization .

2.2 Summary of Literature Review

The reviewed papers cover diverse approaches to YouTube transcript and video summarization using Natural Language Processing (NLP) techniques. One study focuses on a Flask-based YouTube transcript summarization system that incorporates features like translation, speech synthesis, and sharing, enhancing content consumption efficiency. Another paper surveys AI-based video summarization, discussing supervised and unsupervised methods, and emphasizes the combination of FFmpeg for video processing and NLP for accurate transcript analysis. Several papers highlight the use of NLP libraries like SpaCy, gensim, and Hugging Face's transformers for real-time transcript generation and summarization, with one proposing a Chrome extension to access video summaries directly. Another research explores Latent Semantic Analysis (LSA) for video summarization, efficiently extracting subtitles and generating summaries within user-defined time constraints. Additionally, some papers discuss the benefits of both extractive and abstractive summarization methods, with the latter creating new sentences that better mimic human language. Adversarial learning frameworks, using attention mechanisms, and automatic speech recognition (ASR) are also explored for improving summarization quality. Other works emphasize mobile platform solutions and real-time summarization, while one paper examines the future development of these systems, focusing on multi-lingual support, personalized summarization, and the integration of contextual data. In conclusion, these papers collectively advance the field of video summarization by employing a range of innovative techniques and tools, showcasing the potential for improved user experiences in digital content consumption.

Sr. No.	Title	Strengths	Weakness
1	Flask-based YouTube Transcript Summarizer	<p>Integrates a user-friendly interface with multiple features like translation, speech synthesis, and email sharing, enhancing the system's accessibility.</p> <p>Utilizes a CNN model for summarization, providing a practical solution for summarizing YouTube content.</p>	Relies on a default CNN model for summarization, which may not be as efficient or accurate as more advanced models like transformers.
2	AI-based Video Summarization using FFmpeg and NLP.	Provides a comprehensive review of video summarization techniques, covering both supervised and unsupervised methods, and the application of FFmpeg and NLP.	The paper is a survey rather than an implementation, so it lacks hands-on experimentation or results.
3	Survey on YouTube Transcript Summarizer	Highlights the use of popular NLP libraries (SpaCy, gensim, Hugging Face) for real-time audio extraction and summarization. Proposes a practical solution via a Chrome extension for easy access to video summaries.	he focus on summarization using cosine similarity may not capture deeper semantic relationships in the text, limiting the quality of summaries.

Table 2.1: Literature Review

4	NLP-based YouTube Transcript Summarizer	<p>Focuses on enhancing the coherence of summarized content, which is crucial for maintaining the meaning of the video's message..</p> <p>Incorporates HuggingFace transformers for summarization, a state-of-the-art NLP technique.</p>	Limited discussion on how well the model generalizes to different video types or languages.
5	Video Summarization Using NLP	<p>The use of Latent Semantic Analysis (LSA) for summarization is an interesting, less complex approach compared to deep learning alternatives.</p> <p>Incorporates user-defined time constraints, providing a more tailored experience for end-users..</p>	The system relies heavily on external tools like Pytube and MoviePy, which could introduce dependency issues or limitations.
6	Abstractive and Extractive Summarization of YouTube Videos	<p>Provides a comprehensive overview of both extractive and abstractive summarization techniques, offering a well-rounded perspective.</p> <p>Highlights modern techniques such as PEGASUS for abstractive summarization.</p>	Lacks a clear implementation or application of these summarization techniques, which limits the practical value of the paper.

7	Attentive and Adversarial Learning for Video Summarization	Introduces a novel adversarial framework for semi-supervised video summarization, combining attention mechanisms and GANs, showing strong performance on benchmark datasets.	The adversarial approach may be complex and computationally expensive, making it less practical for real-time applications.
8	Real-Time Video Summarization on Mobile Platforms	Focuses on mobile platform optimization, making the summarization tool more accessible for users on the go. Uses MoviePy and Google Transcribe, offering an easy-to-understand pipeline for video summarization.	The text rank algorithm may not provide the most sophisticated summarization, potentially leading to less relevant summaries.
9	Automated Video Summarization Using Speech Transcripts	Offers an algorithm that selects video segments based on word frequencies, improving the quality of summaries by focusing on high-coverage segments. Includes a user study to validate the algorithm's effectiveness, which strengthens the credibility of the proposed method.	The reliance on word frequencies might miss important context or meaning, reducing the overall quality of the summary.

10	YouTube Transcript Summarizer Chrome Extension	<p>The Chrome extension offers a seamless, in-browser solution for summarizing YouTube transcripts, making it very user-friendly.</p> <p>Uses abstractive summarization to generate high-quality summaries.</p>	<p>The extension's effectiveness might be limited by the complexity of the video content, with abstractive summarization requiring advanced models.</p>
11	Future Development of YouTube Transcript Summarizers .	<p>Provides valuable insights into the future directions of YouTube transcript summarizers, including improvements in NLP techniques and multilingual support.</p> <p>Discusses the need for personalized summarization approaches, which could offer tailored user experiences.</p>	<p>The paper is more of a survey and does not provide concrete solutions or implementations.</p>
12	Abstractive YouTube Transcript Summarization	<p>Focuses on an abstractive summarization model that mimics human language, offering more natural and coherent summaries. Utilizes the Hugging Face library, ensuring the use of state-of-the-art NLP models.</p>	<p>Abstractive summarization models can be computationally intensive, potentially limiting real-time usage.</p>

Chapter 3

Software Requirement Specification

In this chapter we will be discussing the Software Requirement Specification required for our model.

In the further sub-sections, we would be discussing the System Requirements required for the project.

3.1 Assumptions and Dependencies

3.1.1 Assumptions

1. It is assumed that end users will have reliable internet connectivity to access the Automated Video Transcript Summarizer through a web interface.
2. The availability of NLP libraries and tools for the chosen programming language is assumed. The system will rely on these libraries for natural language processing capabilities.
3. It is assumed that users will adhere to standard security practices when using the system, such as protecting login credentials and ensuring secure transmission of data.

4. The accuracy of the summarization process relies on the quality and clarity of the input video transcripts. The system assumes that users provide well-formatted and comprehensible transcripts.

3.1.2 Dependencies

1. Natural Language Processing
2. Chrome/Firefox/Safari
3. API
4. Server
5. Security Protocols
6. Language Models

3.2 Functional Requirements

This section describes all the functionalities of the system. There are 5 functional requirements described below -

3.2.1 System Feature 1: Video Input Handling

The system shall expose APIs to allow programmatically submitting video transcripts, supporting standardized formats such as JSON or XML.

3.2.2 System Feature 2: Summarization Algorithm

The system shall implement an advanced summarization algorithm that condenses the content of video transcripts while maintaining the coherence and relevance of key information.

3.2.3 System Feature 3: User Interfaces

The system shall provide a user-friendly web interface accessible to end users for manual upload, configuration, and retrieval of summarized transcripts.

3.2.4 System Feature 4: API Interfaces

API documentation shall be provided to guide developers in the proper usage of the exposed interfaces

3.2.5 System Feature 5: Language Support

The system shall detect and support multiple languages in video transcripts, providing accurate analysis and summarization irrespective of the linguistic context.

3.3 System Requirements

3.3.1 Hardware Requirements

1. Multi-core processor (Intel Core i5+) and 8GB RAM.
2. Dedicated NVIDIA GPU (GeForce GTX/Tesla)
3. RAM: at least 16GB RAM, depending on traffic and task size.
4. Storage: Sufficient space for development tools and project code.

3.3.2 Software Requirements

1. Windows 10 OS (64 bit)
2. Visual Studio Code
3. Natural Language Toolkit
4. Django REST Framework

3.4 Analysis Models: SDLC model to be Applied

Agile Methodology: Agile stands out as an iterative and incremental approach to software development that prioritizes adaptability, collaboration, and the delivery of working software in short, manageable iterations.

Its characteristics include:

Flexibility: Agile embraces change and responds effectively to evolving requirements, allowing for adjustments at the end of each iteration.

Collaborative Approach: The methodology promotes frequent communication and collaboration among cross-functional team members, fostering a culture of shared responsibility and collective problem-solving.

Customer-Centric Philosophy: Agile places a strong emphasis on customer involvement throughout the development process. Continuous customer feedback loops are integral to ensuring that the product aligns with user expectations.

Incremental Delivery: Agile enables the delivery of a Minimum Viable Product (MVP) in the early stages, allowing for the quick deployment of essential functionalities. This iterative delivery approach facilitates the gathering of valuable user insights.

Agile methodology is highly suitable for projects where requirements are expected to evolve, and regular feedback is essential. In the context of the Automated Video Transcript Summarizer, Agile can seamlessly accommodate changes in summarization algorithms, user interface preferences, and other aspects identified through ongoing user interactions. The iterative nature of Agile aligns seamlessly with the dynamic aspects of natural language processing. It provides a framework for continuous improvements, ensuring that the project remains responsive to emerging needs throughout its development lifecycle. The flexibility inherent in Agile methodology positions the project to adapt to evolving requirements and prioritize user satisfaction.

The Iterative Model 3.1 is a software development process that breaks down the project into repeated cycles, called iterations, where each cycle involves planning, design, implementation, testing, and evaluation. Unlike traditional models, it delivers incremental versions of the software early and continuously, allowing stakeholders to provide feedback at each stage. This approach helps

refine requirements and address risks early, offering greater flexibility and enabling continuous improvements. However, it can be resource-intensive and challenging to manage, making it ideal for projects with evolving requirements or where early user feedback is critical.

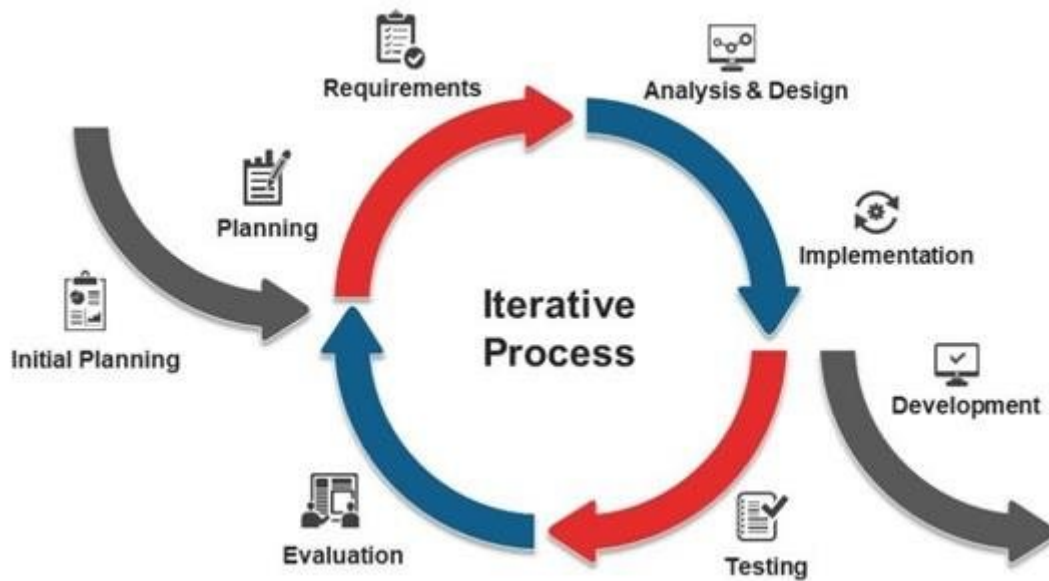


Figure 3.1: Iterative Model

Chapter 4

System Design

This chapter will discuss the system design of the project. It would provide an insight into the system architecture and also the various underlying architectures that the model would make use of.

4.1 System Architecture

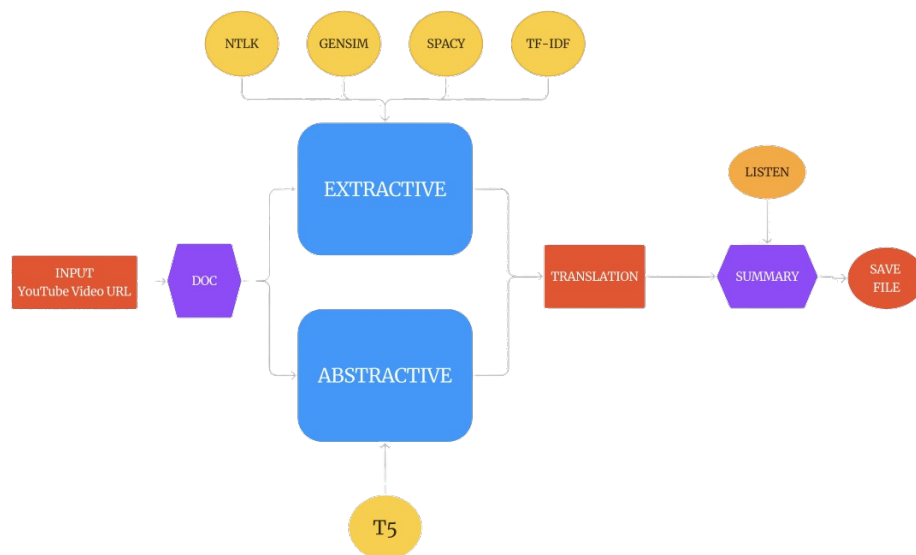


Figure 4.1: System Architecture

This process diagram outlines a method for summarizing YouTube videos by leveraging both extractive and abstractive summarization techniques. It starts with a YouTube video URL, which serves as the input. The video is first transcribed into text, creating a document for summarization. The summarization stage includes two main approaches. In extractive summarization, essential sentences or phrases are selected from the document to produce a concise summary, often using libraries like NLTK, Gensim, Spacy, and TF-IDF. For a more nuanced summary, abstractive summarization rephrases the content, capturing the document's core meaning, and models like T5 are commonly used for this. Optionally, the generated summary can be translated into other languages if desired. Additionally, there is a listening option, allowing users to hear the summary read aloud instead of reading it. Finally, the summary can be saved as a file, providing a convenient reference for future use. It is represented in the Fig.4.1.

4.2 Data flow diagrams

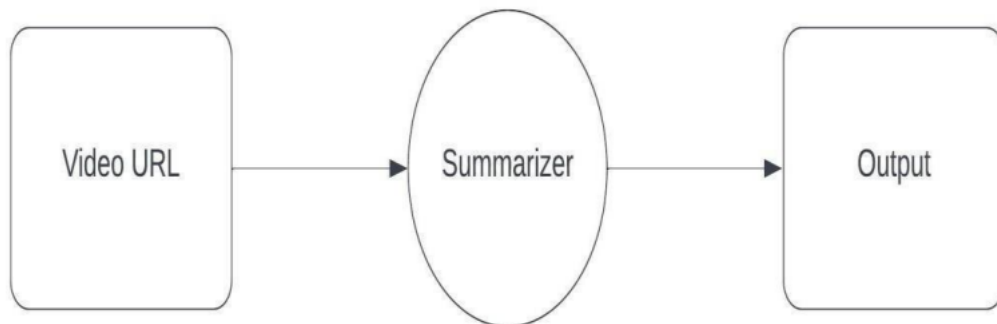


Figure 4.2: DFD level 0

The Level-0 DFD represents the entire Automated Video Transcript Summarization System (AVTS) as a single process, with input data in the form of either a video URL or a video file, and output as the generated video summary. The incoming data is processed, and the predicted output is the summary, which is provided to the user. This is illustrated in Figure 4.2.

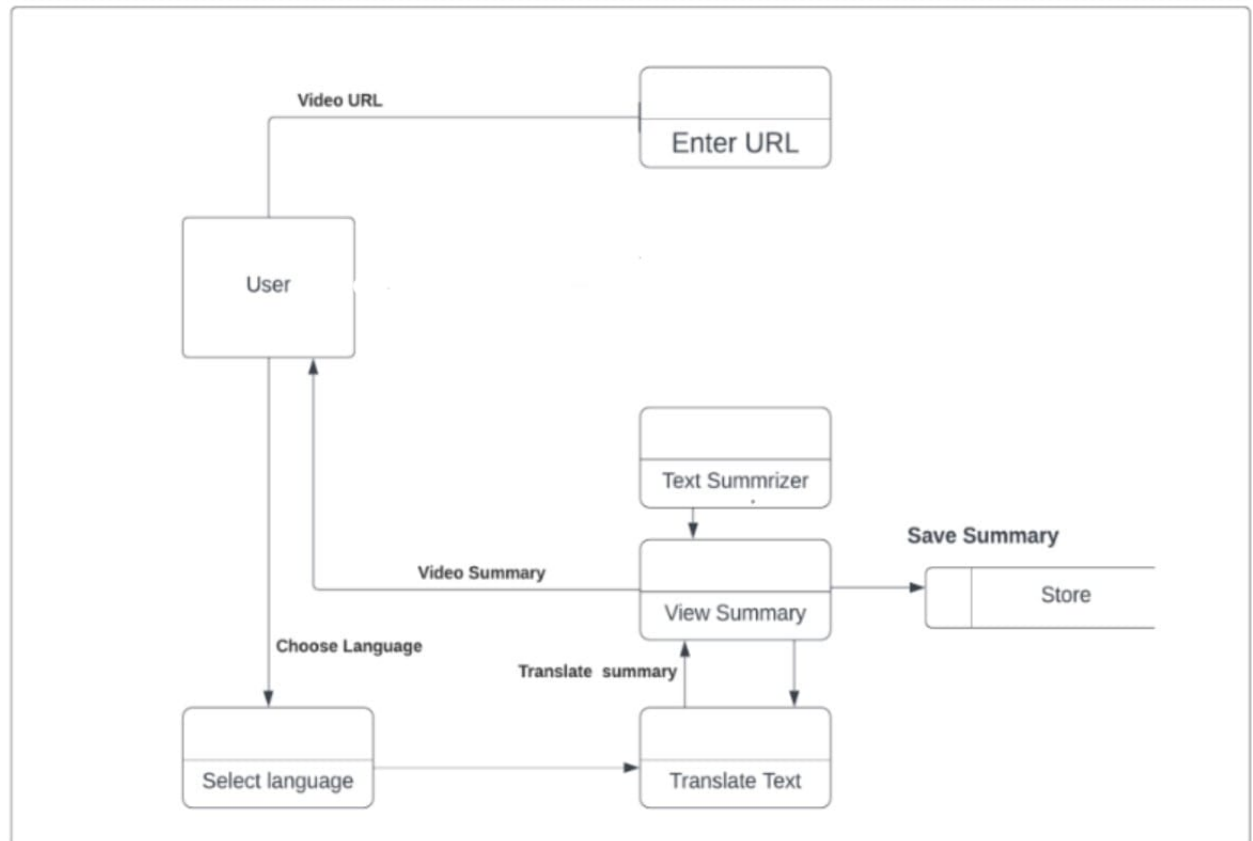


Figure 4.3: DFD level 1

In the Level-1 DFD (Figure 4.3), the system's functionality is broken down into smaller, more detailed components: The Input Module receives either a video URL or a video file. The Summarizer Module handles the extraction of transcripts and applies NLP libraries to generate a summary. The Output Module presents the generated summary to the user.

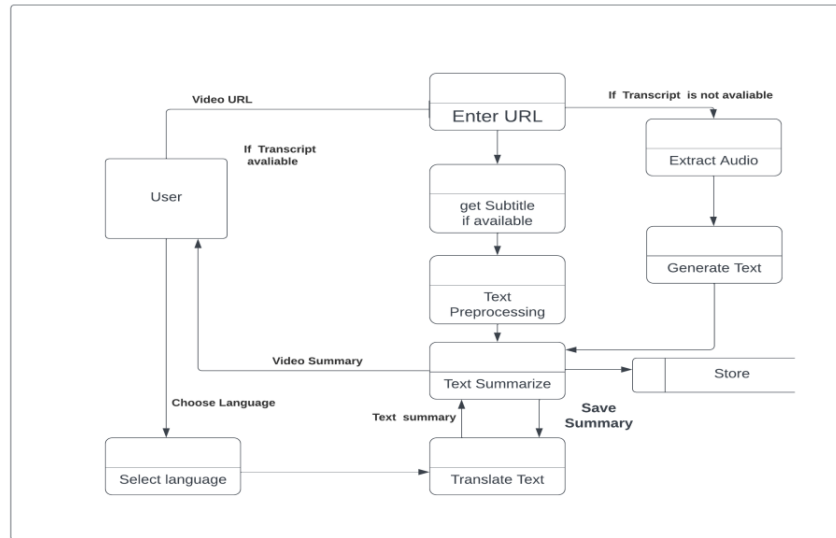


Figure 4.4: DFD level 2

The Level-2 DFD (Figure 4.4) further breaks down each module, showing detailed interactions between specific components. For example, the Summarizer Module may utilize the YouTube Transcript Summarizer API to fetch transcript data and then apply NLP techniques for summarization. This structured approach highlights the flow of data from input to output, detailing the processes involved in summarizing video content through transcript generation and NLP-based summarization.

4.3 Activity Diagram

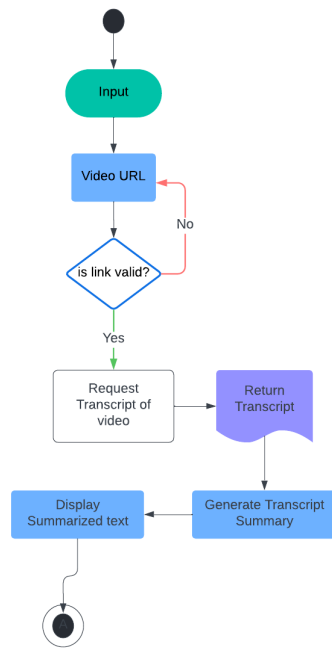


Figure 4.5: Activity Diagram

Activity Diagram outlines the steps and decision points involved in the automatic video transcript summarization process, from the user's initial request to the finalization of the summarized transcript.

4.4 System Implementation

The output of the "Automated Video Transcript Summarizer" project is a concise, easy-to-read summary that captures the key points of the original video content. After processing the video, the tool provides a structured overview that allows users to quickly understand the main ideas without watching the entire video.

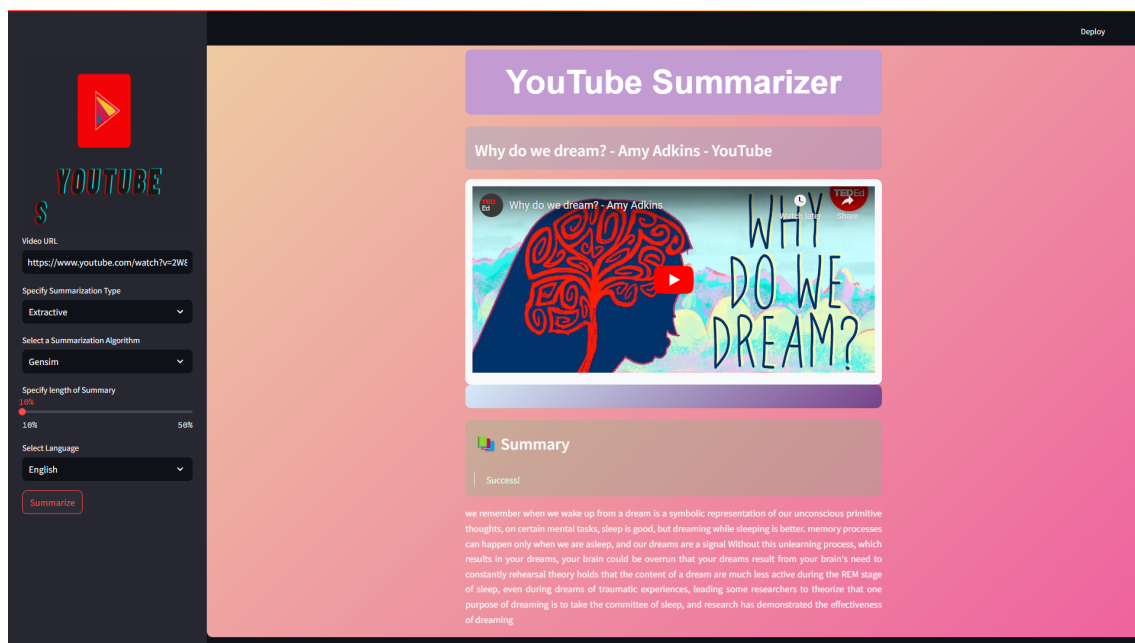


Figure 4.6: Result 1

The output of this project is a concise summary that captures the main points of a video, allowing users to understand its essential content quickly.

This summary can be saved for future reference, making it convenient for users to access essential information efficiently. The output enhances the user's ability to gain insights from video content in a fraction of the time, optimizing both comprehension and time management.

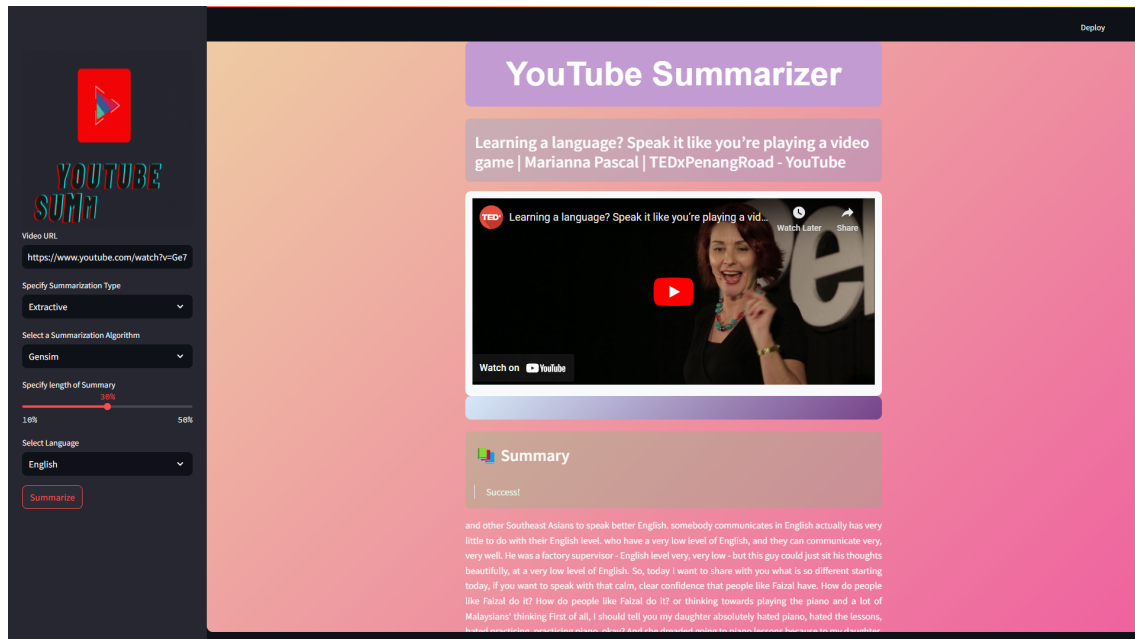


Figure 4.7: Result 2

The output of this project is a concise summary that captures the main points of a video, allowing users to understand its essential content quickly.

Chapter 5

Project Plan

This chapter tells the basic plan that have been considered in our project. It tells about various factors of the project including the overview of risk management and the schedule of the same.

5.1 Project Task Set

1. **Research NLP Models:** Explore text summarization algorithms and understand their pros and cons.
2. **Algorithm Development:** Create a custom NLP-based summarization model for video transcripts.
3. **Integration with Video Processing:** Link the summarization algorithm with video processing for transcript extraction.
4. **UI Design:** Design a user-friendly interface based on user feedback and requirements.
5. **Application Development:** Build frontend and backend for the summarizer app, enabling video uploads and content display.
6. **Testing:** Perform unit, integration, and user tests to identify and fix issues.

5.2 Risk Management

Project risk management is the process of identifying, analyzing and then responding to any risk that arises over the life cycle of a project to help the project remain on track and meet its goal. Risk management isn't reactive only; it should be part of the planning process to figure out risk that might happen in the project and how to control that risk if it in fact occurs.

5.2.1 Risk Identification

Minimal risks are involved in the project model owing to the learning done and avoidance of the previous models.

5.2.2 Risk Analysis

Risk analysis is handling the system response in the risk conditions which are identified. If the system response is incorrect then the developers work on it and try fixing out the identified risks. Risk analysis can affect on project objectives which can work incorrectly in the risk cases. The risk analysis assigns the severity to the risks and developers start working accordingly.

1. High severity: Catastrophic.
2. Medium severity: Critical.
3. Low severity: Marginal.

5.2.3 Risks

1. **Computer Malfunction: Drivers (Compatibility) Issues, Framework Issues.**
2. **Computational Uptime: Training and Testing time of CNN.**
3. **Real world scenarios: Certain real-world scenarios might be difficult to simulate (e.g. general traffic in India).**
4. **Low model accuracy: Using experimental CNN models may result increase in overall time consumption.**
5. **Unregulated Industry: Not enough solid facts to create a baseline for safety standards.**

5.3 Project Schedule

The time required for completion and the time coverage for various tasks are covered in Fig. 5.1.

Sr.No.	Activity	July	Aug	Sept	Oct	Nov
1	Project Synopsis					
2	Project Aproval					
3	Literature Survey					
4	Problem Statement					
5	Project Requirements					
6	Project Plan					
7	Project Architecture					
8	Project design Details					
9	Project implementation					
10	Abstract and Conclusion					
11	Report submission and PPT					

Figure 5.1: Activity Sheet

5.4 Team Organization

According to Savitribai Phule Pune University(SPPU) rules, the project team is supposed to be of 4 students from Final year of Information Technology. The team was formed in June 2024, comprising of- Agrawal Khushbu, Gajare Krushna, Pisal Pranav, Wagh Manasi.

5.4.1 Team Structure

Prof. A.A.Yadav - Project Guide

All the members contributed equally in the project and all the segments received appropriate attention from all the members consisting aspects of project designing, implementation and testing etc.

In this chapter, we looked at the project plan including Identification of risks. In the next chapter, we will conclude the report.

Chapter 6

CONCLUSIONS

In the previous chapter of the report the results of all the phases of the project implementation were explained. In this chapter, the final conclusion of the project, proposed and implemented model is given. Later on in this chapter the future scope and applications are also explained in brief.

6.1 Conclusions

The increase in video content made it near to impossible to watch entire video of one or two hours. Using Natural Language Processing Techniques Instead of manual summarization helps in saving precise time and improves the summary quality. According to video content, Extractive and Abstractive summarization models can be used for Summarization task. Using YouTube subtitles for generating summary may generate error for videos which does not have subtitles or have disabled subtitles. Instead of using subtitles for generating summary another method is more effective that uses video to audio conversion and in next step speech to text conversion using NLP APIs. Automated generated summaries save time and also improves the summary quality compared to manually generated summary.

6.2 Future Work

1. **Target Audience:** Visually impaired individuals and students consuming educational video content are key beneficiaries of the proposed solution.
2. **Transcript to Audio Conversion:** After generating a transcript summary, it can be converted to an audio format. This conversion aids visually impaired users by providing an audio summary of the video content for better understanding.
3. **Improved Accessibility:** The feature significantly improves content accessibility, making educational and video summaries more inclusive for visually impaired individuals.
4. **Potential Impact:** The personalized audio summaries enhance user engagement and provide a more familiar listening experience, increasing the usability of the solution.

6.3 Areas of Application

1. **Education and E-Learning:** Summarizes lecture videos and courses for quick understanding, aiding students' learning. Enhances e-learning platforms by providing summarized transcripts alongside video content.
2. **Professional Training:** Helps summarize training videos and webinars, boosting productivity in corporate environments. Speeds up onboarding for new employees by summarizing instructional materials.
3. **Content Creation and Marketing:** Assists content creators in generating concise summaries for promotional videos, boosting engagement. Supports market research by extracting key insights from summarized transcripts.
4. **Accessibility and Inclusion:** Provides alternative access to video content for individuals with hearing impairments or language barriers. Supports inclusivity in educational and organizational digital platforms.
5. **Research and Data Analysis:** Accelerates analysis of video data, helping researchers extract insights quickly. Facilitates identification of trends and patterns in large video datasets.

Appendix A

Problem Statement Feasibility and Completeness Analysis

For an automated video transcript summarizer, generating a concise and accurate summary from a long transcript requires complex natural language processing (NLP) techniques, which can be computationally intensive and may take non-linear time.

However, verifying a generated summary for accuracy and coherence can be done in linear or polynomial time.

Therefore, this problem could be classified as NP-Complete.

Alternate approaches could classify it as NP, depending on simplifications in summarization techniques or constraints on transcript length.

REFERENCES

- [1] Kumari, P. Vijaya, M. Chenna Keshava, C. Narendra, P. Akanksha, and K. Sravani. "YouTube Transcript Summarizer Using Flask and Nlp." Journal of Positive School Psychology 6, no. 8 (2022): 1204-1209.
- [2] Hansaraj Wankhede, R Bharathi Kumar, Sushant Kawade, Ashish Ramtekkar, Rachana Chawke. "AI-based Video Summarization using FFmpeg and NLP." International Journal of Innovative Science and Research Technology Issue- 4, April – 2023.
- [3] Inamdar, Eesha, Varada Kalaskar, and Vaidehi Zade. "SURVEY PAPER ON YOUTUBE TRANSCRIPT SUMMARIZER." International Research Journal of Modernization in Engineering Technology and Science Issue:04 April 2023.
- [4] Siddhartha, Prashu Pandey, Ansh Saxena GUIDEAnupam kumar sharma, "Youtube Transcript Summarizer", International Journal of Research in Engineering and Science (IJRES) Issue 5May 2023.
- [5] Sanjana R, Sai Gagana V, Vedhavathi K R, Kiran K N," Video Summarization using NLP", International Research Journal of Engineering and Technology (IRJET) Issue: 08 — Aug 2021.

- [6] S. Tharun, R. Kranthi Kumar, P. Sai Sravanth, G. Srujan Reddy, B. Akshay,” Survey on Abstractive Transcript Summarization of YouTube Videos”, International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Issue-2, April 2022.
- [7] Fu, Tsu-Jui, Shao-Heng Tai, and Hwann-Tzong Chen. ”Attentive and adversarial learning for video summarization.” In 2019 IEEE Winter Conference on applications of computer vision (WACV), pp. 1579-1587. IEEE, 2019.
- [8] Choudhary, Pradeep, Sowmya P. Munukutla, K. S. Rajesh, and Alok S. Shukla. ”Real time video summarization on mobile platform.” In 2017 IEEE International Conference on Multimedia and Expo (ICME), pp. 1045-1050. IEEE, 2017.
- [9] Taskiran, Cuneyt M., Arnon Amir, Dulce B. Ponceleon, and Edward J. Delp III. ”Automated video summarization using speech transcripts.” In Storage and Retrieval for Media Databases 2002, vol. 4676, pp. 371-382. SPIE, 2001.
- [10] Gousiya Begum, N. Musrat Sultana, Dharma Ashritha, “YOUTUBE TRANSCRIPT SUMMARIZER”, International Journal of Creative Research Thoughts. Volume 10 Issue 6 June 2022.
- [11] Sarthak Jain, Jahnavi Jain, Gurminder Kaur, Dr. Yatu Rani, “YouTube Transcript Summarizer”, “International Journal of Science and Research (IJSR)”, Issue -2, February 2023.
- [12] Tryphena Brighty, Pooja V, Jaya Varsini V, Vishnu Priya N, “YOUTUBE TRANSCRIPT SUMMARIZER”, INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT

AND SCIENCE (IJPREAMS), Vol. 03, Issue 05, May 2023.

- [13] Prof. S. H. Chaflekar, Achal Bahadure, Hosanna Bramhapurikar, Ruchika Satpute, Rutuja Jumde, Sakshi A. Bakhare, Shivani Bhirange, “YouTube Transcript Summarizer using Natural Language Processing”, International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 2, Issue 1, April 2022.

- [14] Pradeepika Verma and Anshul Verma, “A Review on Text Summarization Techniques”, Journal of Scientific Research, Volume 64, Issue 1, 2020