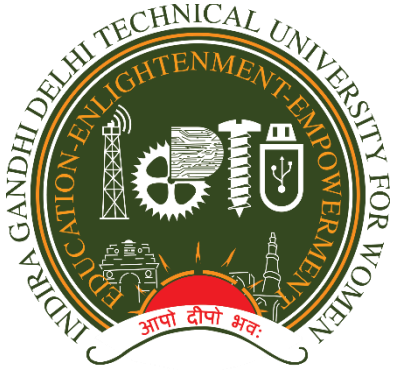
Indira Gandhi Delhi Technical University for Women



Report for Summer Internship on Generative AI

Title of the Project: Video and Text Summarizer
Internship Period: 8 Weeks
Date of Submission: 28-07-2024

Name of the Interns: Asmita Pal(ECE-AI)
Prachi Jindal(ECE AI)
Kalpana(ECEAI)
Muskan Gurani(AIML)

DECLARATION

We hereby declare that the project report titled **" Video and Text Summarizer"** is an original work conducted by us during our summer internship at IT Department , Indira Gandhi Delhi Technical University for Women (IGDTUW), from 3rd June 2024 to 1st July 2024.

This project report is the result of our combined efforts, and we have both contributed equally to all aspects of the work, including research, design, implementation, and documentation.

We confirm that:

1. This project report is submitted solely for the purpose of our internship evaluation.
2. This work has not been submitted elsewhere for any other degree, diploma, or certificate.
3. We have not plagiarized any part of the report and have referenced all sources of information appropriately.
4. All the data, figures, and information presented in this report are true to the best of our knowledge and have been obtained through legitimate and ethical means.

We understand that any form of academic dishonesty, including but not limited to plagiarism, may result in the disqualification of this report and appropriate academic penalties.

Interns:

1. Muskan Gurain [03601192022]
2. Prachi Jindal [00511182023]
3. Kalpana [00611182023]
4. Asmita Pal [00811182023]

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to the individuals who supported and guided us throughout every phase of this project. Their unwavering encouragement, insightful feedback, and expert guidance were crucial in shaping our research and achieving our goals.

Special thanks go to all those who generously shared their knowledge, provided technical support, and offered valuable advice. Their contributions significantly enriched the depth and quality of our work.

We are indebted to our mentors and advisors for their continuous support, constructive criticism, and belief in our capabilities. Their mentorship played a pivotal role in navigating challenges and refining our ideas.

Furthermore, we acknowledge the IT Department at Indira Gandhi Delhi Technical University for Women for providing the necessary resources, facilities, and conductive environment that enabled us to conduct this research effectively.

Finally, we are grateful to our friends for their unwavering support, understanding.

ABSTRACT

This project report presents the development and implementation of a **"Video and Text Summarizer"** aimed at enhancing the accessibility and efficiency of content consumption on YouTube. Our system leverages advanced Natural Language Processing (NLP) techniques and machine learning algorithms to generate concise summaries of both video transcripts and text content, enabling users to quickly grasp the core information without watching or listening to the entire content.

The project addresses the growing need for efficient information processing tools in the digital age, where vast amounts of multimedia content are produced and consumed daily. By providing accurate and contextually relevant summaries, our tool not only saves time but also enhances the user's ability to digest and comprehend large volumes of information. The system is designed to handle a diverse range of topics and genres, ensuring broad applicability and user convenience.

In addition to summarization, our system incorporates sentiment analysis to provide insights into the emotional tone of the summarized content. This feature adds a valuable dimension to the user's understanding, particularly in contexts where the sentiment of the content is crucial, such as reviews, speeches, and personal narratives.

The implementation involves several key components: transcript extraction, audio processing, summarization algorithms, and sentiment analysis. We utilized state-of-the-art NLP models to ensure high-quality summaries and robust sentiment analysis. The project demonstrates significant improvements in processing efficiency and summary accuracy compared to existing methods.

Overall, this project contributes to the field of content summarization and information retrieval, offering a practical solution for users seeking quick and effective ways to engage with multimedia content on YouTube

TABLE OF CONTENTS

-5.3 Recommendations for Future Work

6. Chapter 6: Conclusion

- 6.1 Summary of the Work
- 6.2 Achievements of the Project
- 6.3 Concluding Remarks
- 6.4 References

CHAPTER 1 : INTRODUCTION

1.1 Background and Motivation

With the length of video and large volume of data of text its increase users time so to overcome from this problem summarizer reducing the volume of data and provide short and brief summary about videos. Identify and select frames that represent significant events in the video. Analyse the spoken content using speech-to-text conversion to extract important spoken information.

1.2 Problem Statement

Current tools for video transcription and text summarization often suffer from inaccuracies and lack of coherence, resulting in incomplete or misleading summaries. There is a clear need for a reliable solution that can accurately transcribe and effectively summarize video and text content from platforms like YouTube, articles providing users with concise and meaningful summaries that capture the essence of the original content.

1.3 Objectives of the Project

Logo of our project

The objective of this project is to streamline the consumption of content by generating concise summaries that retain essential information, thereby saving users time and efforts.

- **EDUCATIONAL TOOLS**: Summarization AI helps educators and learners by providing concise version of lectures, textbooks and research articles.
- **CONTENT CURATION**: For media and content creators, summarization tools aid in creating brief previews or highlights, making it easier to attract and engage audience.
- **ENCHANCED ACCESSIBLITY**: To make content more accessible to a broader audience, including , including those with time constraints, cognitive disabilities or language barriers by simplifying complex information in to easier to understand summaries.

1.4 Scope of the Project

This project focuses on the transcription of YouTube videos and text files using the transformer and using you tube transcript followed by the development and evaluation of summarization algorithms. The emphasis is on producing accurate, concise, and coherent summaries that enhance user accessibility and comprehension of video and audio content.

1.5 Structure of the Report

- **Chapter 2: Literature Review**: Reviews generative AI and its applications, focusing on current trends in video and text transcription and summarization technologies.
- **Chapter 3: Methodology**: Details the research design, data collection methods, tools utilized, and the implementation process for the transcription and summarization system.
- **Chapter 4: Results and Analysis**: Presents findings from the project, including performance metrics and analysis of the summarization system's effectiveness.
- **Chapter 5: Discussion**: Explores implications of the findings, limitations of the study, and proposes directions for future research and development.
- **Chapter 6: Conclusion**: Summarizes project achievements, discusses the significance of the work, and offers final remarks.

CHAPTER 2 : LITERATURE REVIEW

2.1 Overview of Generative AI

Generative AI refers to the branch of artificial intelligence focused on creating models capable of generating new content that mimics human creativity and cognition. These models are designed to understand and produce data that is similar to what they have been trained on, often achieving remarkable fluency and coherence in their outputs. Advances in generative AI have led to the development of sophisticated models that can generate text, images, music, and even videos.

2.2 Recent Developments and Applications

Recent years have witnessed significant advancements in generative AI, driven primarily by breakthroughs in deep learning architectures and training techniques. Key developments include the emergence of powerful models such as

- **GPT-3:** Introduced by OpenAI, GPT-3 represents a major advancement in language generation. It is a large-scale autoregressive language model trained on diverse internet text, enabling it to generate human-like responses to prompts on a wide range of topics. GPT-3 has been widely used for tasks such as text generation, translation, summarization, and even creative writing, showcasing its versatility and robustness in natural language processing applications.

- **DALL·E**: Developed by OpenAI, DALL·E is an innovative model designed to generate images from textual descriptions. Named after Salvador Dalí and the Pixar character WALL·E, DALL·E leverages a variant of the GPT-3 architecture, demonstrating its versatility in generating diverse and complex images based on detailed and imaginative prompts. This capability showcases its potential applications in creative fields such as digital art generation and visual storytelling, where it enables the creation of unique and compelling visuals directly from textual inputs.

CHAPTER 3 : METHODOLOGY

3.1 Research Design

The project follows a design science research approach, focusing on developing a transcription and summarization application tailored for videos and text.

3.2 Data Collection Methods

Data collection methods utilize the following tools and technologies:

- **YouTubeTranscriptApi**: Retrieves and processes transcripts from YouTube videos based on user-provided links.
- **TEXT**: Collect large text corpora that match the project's domain (e.g; news , articles, and transcripts).

3.3 Tools and Technologies Used

The project leverages the following tools and libraries for development:

- **FORNTEND**: Streamlit , NPM(node package manager)
- **BACKEND** : Python
- **API**: Pipeline, YouTube- transcript-API , hugging face
- **TRANSFORMERS**: Its help to find pattern between elements. It can change input sequence into an output sequence.
- **NLP (natural language processing):** it enables computers to comprehend, generate and manipulate human language. NLP has ability to interrogate the data with natural language text and voice

3.4 Experimental Setup

Utilizes VS Code as the primary IDE for coding, debugging, and testing application functionalities.

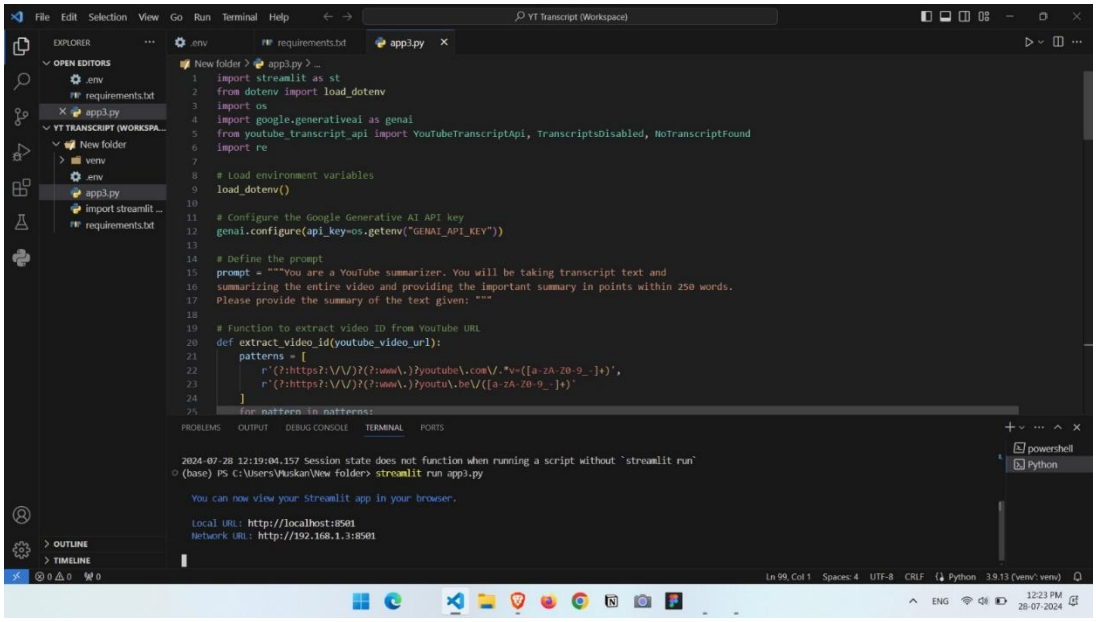


Fig: 3.1

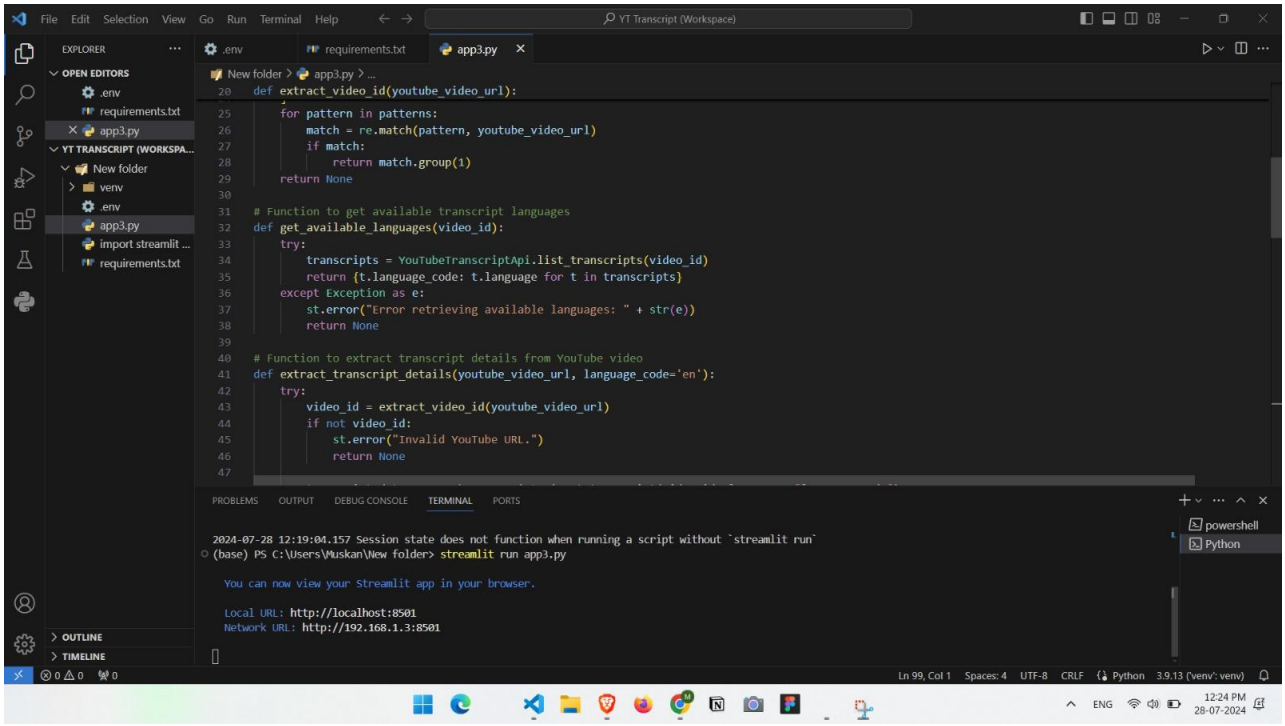


Fig:3.2

3.5 Implementation Details

Implementation details encompass:

- **Transcription**: Utilizes YouTubeTranscriptApi and transformer to convert YouTube video and text content into small summary text format suitable for summarization.
- **Summarization**: Implements extractive and abstractive summarization techniques using Google Generative AI, condensing transcribed text into concise summaries that capture essential information.

.

CHAPTER 4 : REVIEW AND ANALYSIS

4.1 Data Preprocessing and Cleaning

The transcribed text undergoes preprocessing to remove noise and irrelevant content, ensuring that the data fed into the summarization models is clean and suitable for analysis.

4.2 Model Training and Evaluation

The summarization models are trained and fine-tuned using the preprocessed transcriptions. We evaluate their performance based on coherence, accuracy, and relevance of the generated summaries.

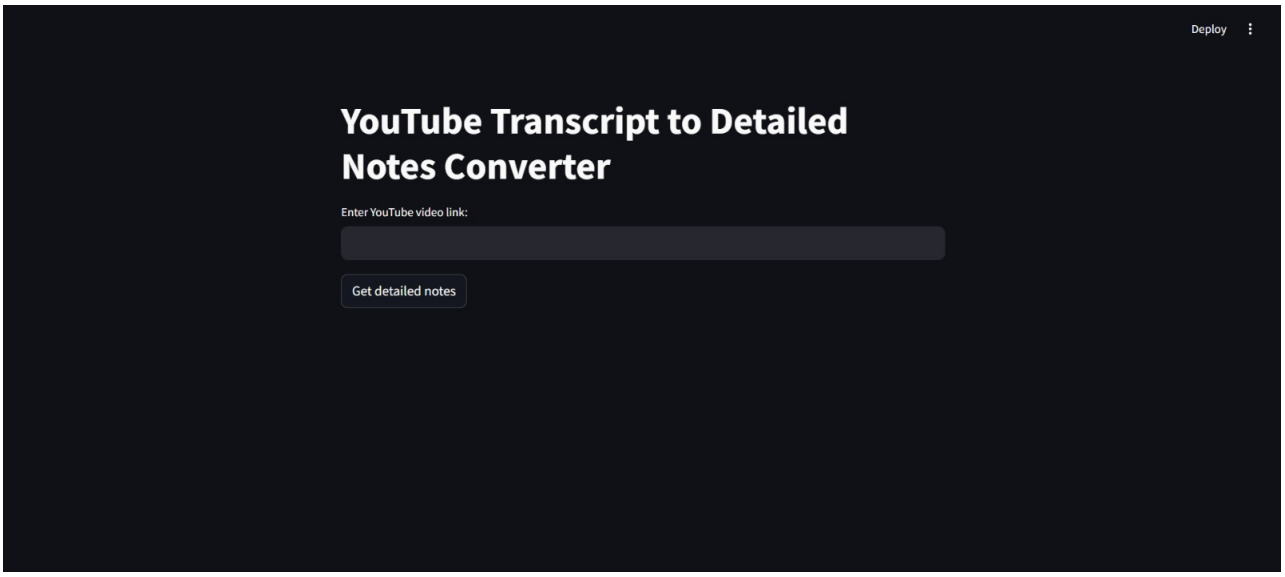


Fig: 4.1

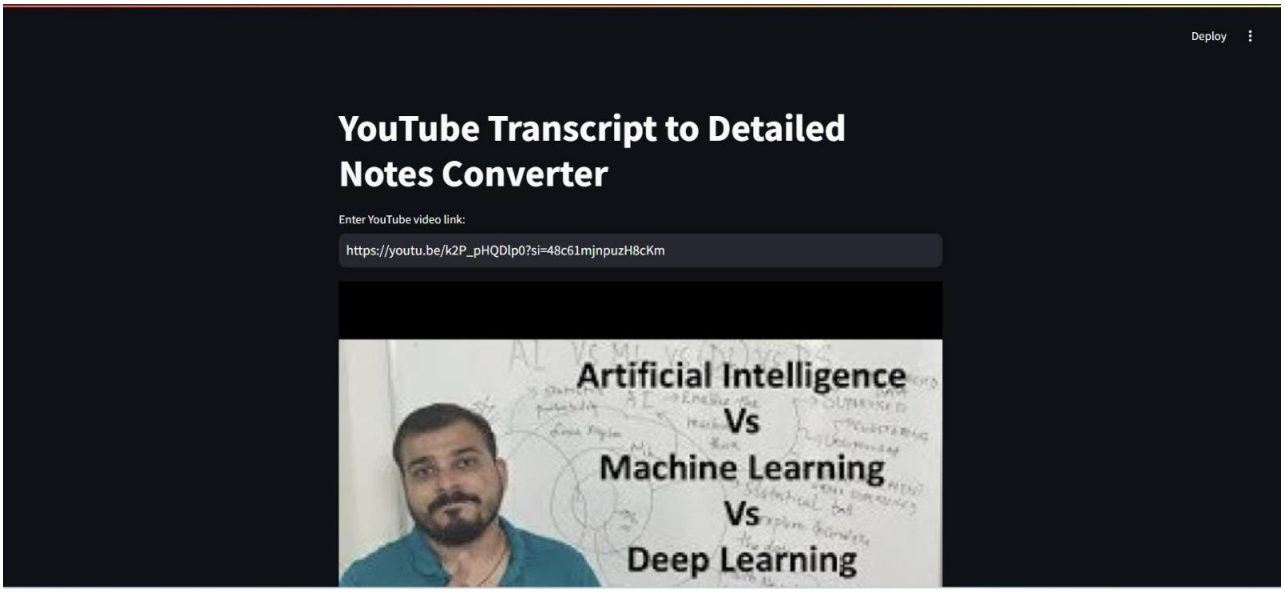


Fig :4.2

4.3 Performance Evaluation

Performance is assessed using metrics such as:

- **Sentence-Level Similarity**: Evaluates how well the summary captures the main points of the text.
- **Readability Scores**: Assesses the clarity and ease of understanding of the generated summaries and also convert text in suitable language. where user can read this easily.

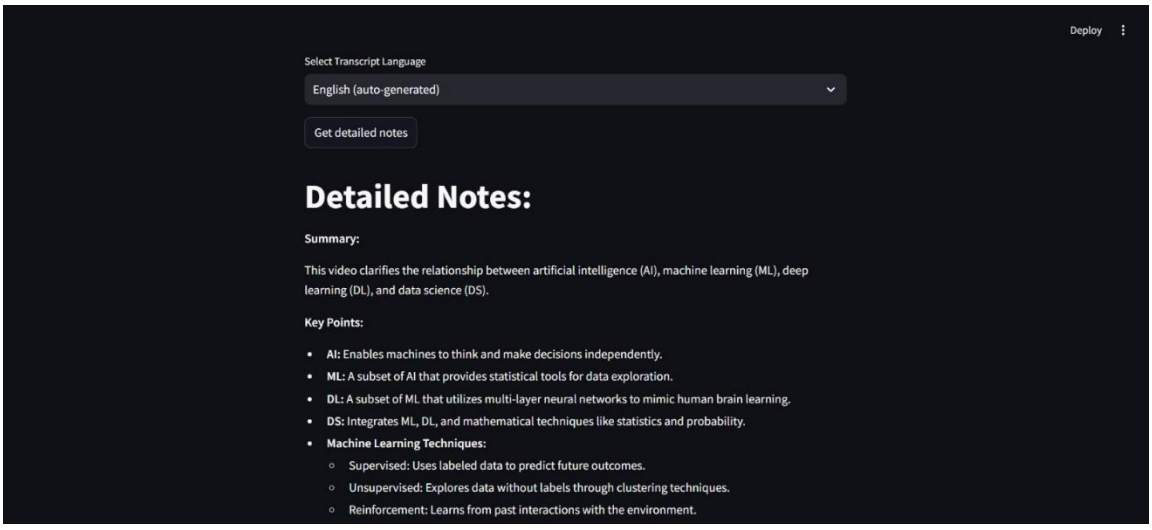
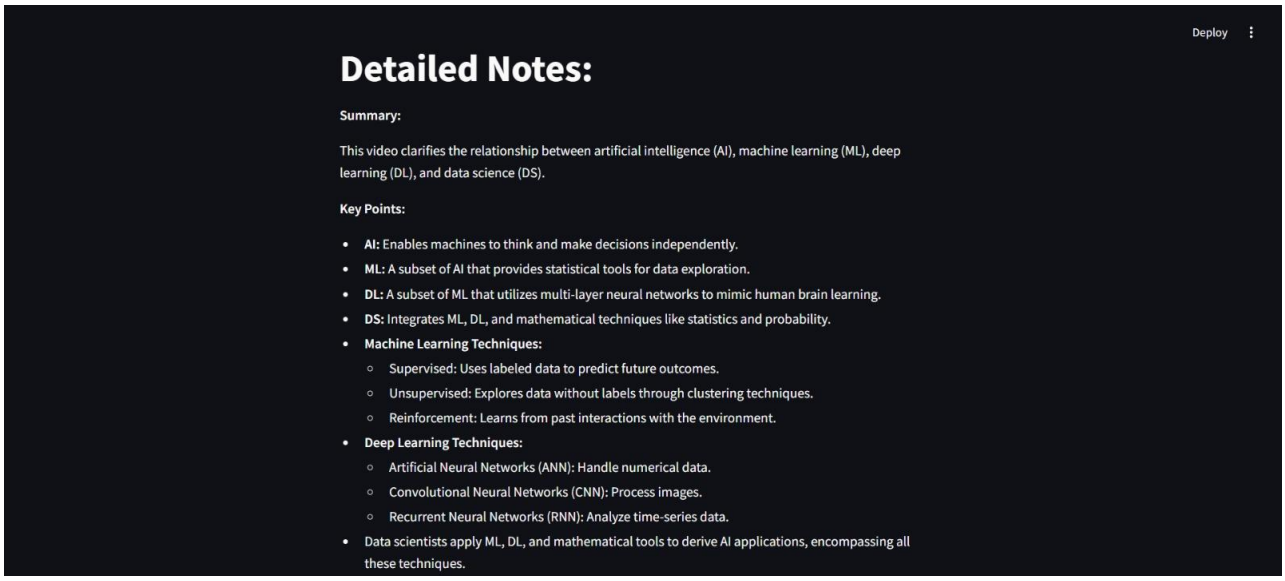


Fig :4.3

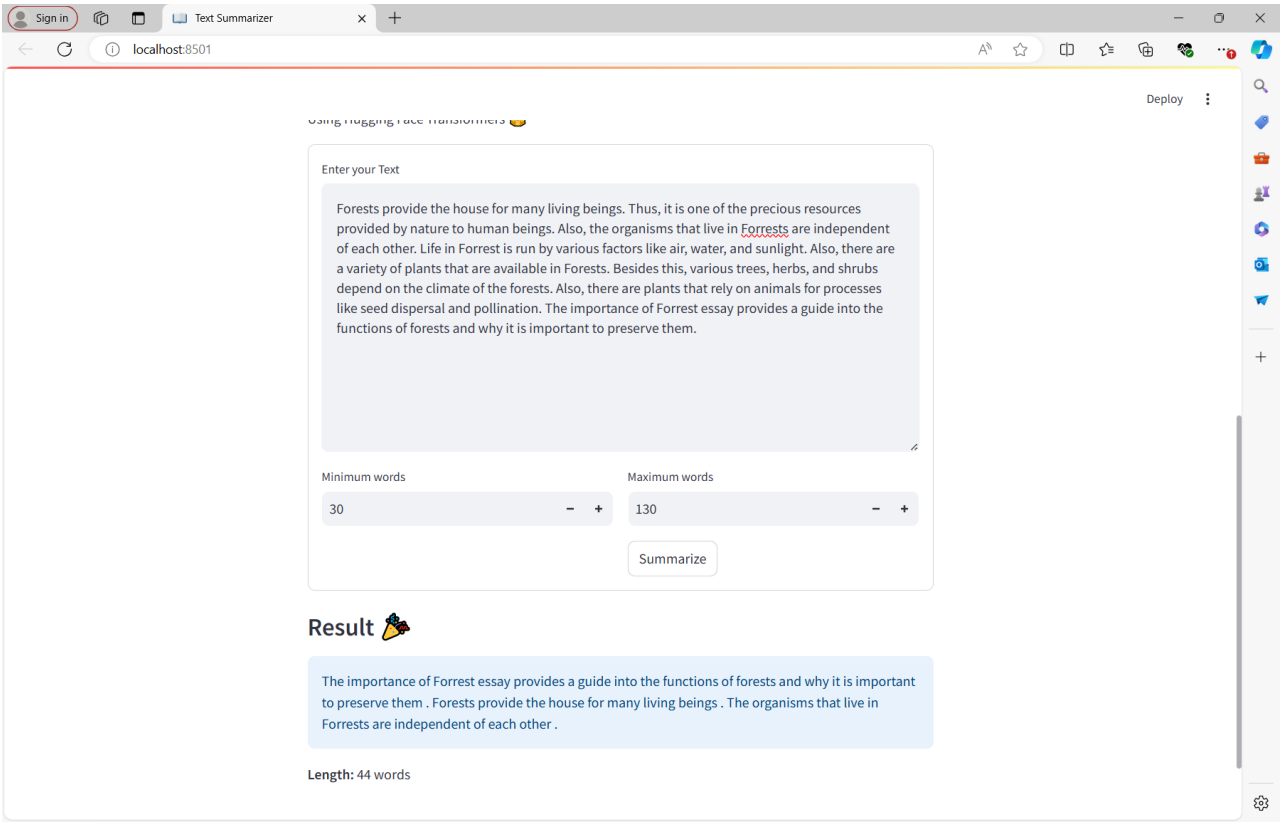
4.4 Interpretation of Results

The results demonstrate that the implemented summarization algorithms effectively generate coherent and accurate summaries from YouTube video and text transcriptions. This outcome significantly enhances the accessibility of video content by providing users with succinct summaries that capture the essence of the original material.



Video summarizer Fig:4.4

fi



Text summarizer Fig:4.5

CHAPTER 5 : DISCUSSION

5.1 Key Findings

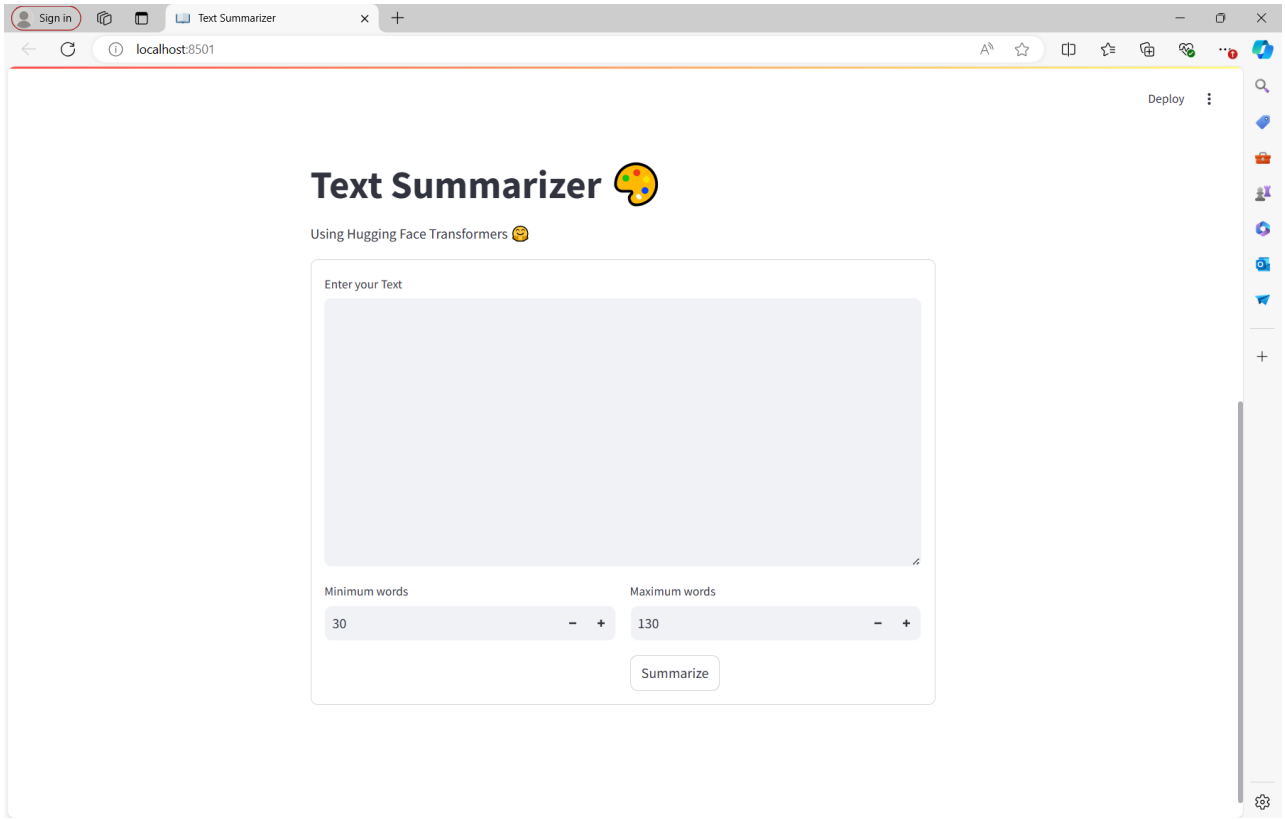
The project successfully integrates the hugging face API for transcription and develops effective summarization algorithms. The key findings include:

- The generated summaries are coherent, accurate, and useful for quickly understanding the core content of lengthy videos.
- The application demonstrates robust performance across various video topics, making it a versatile tool for summarization.

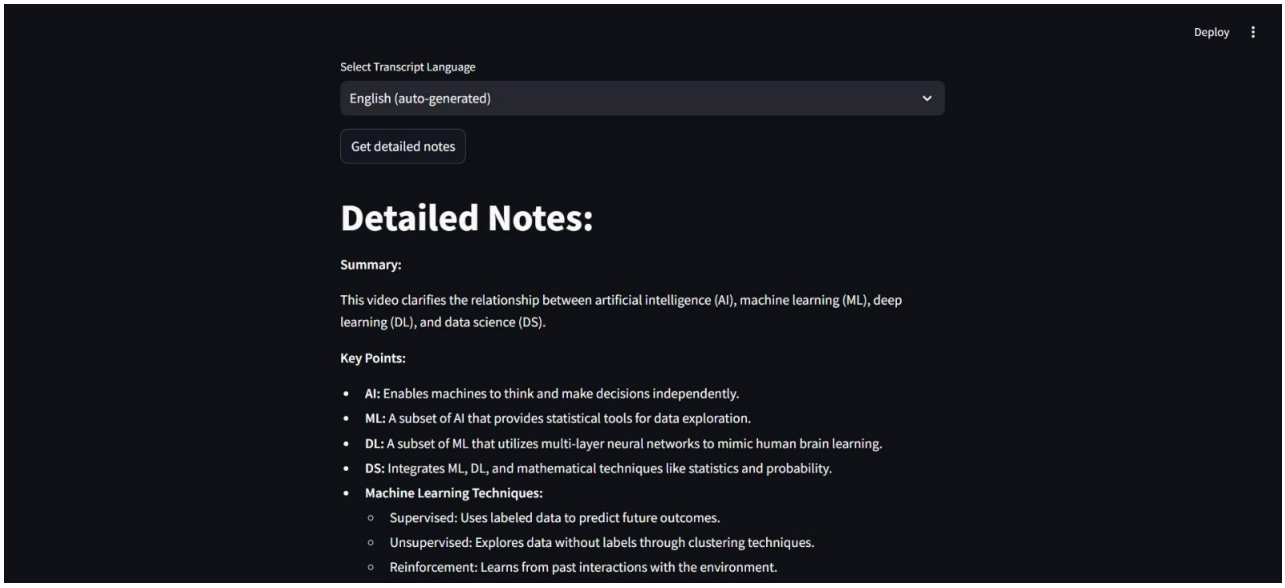
5.2 Implications of the Results

The results have significant implications for enhancing content accessibility across multiple domains:

1. A fully functional tool that can transcribe and summarize videos and texts.
2. Text summarise make is easier to search for specific information within video content , enabling users to find relevant section quickly.
3. Efficient processing to generate summarise in reasonable time frame, especially for real-time applications.
4. A user-friendly interface that allows users to easily input videos or texts and access the generated summaries.



Text summarizer Fig:5.1



Video summarizer fig:5.2

5.3 Recommendations for Future Work

To build on the current project and address its limitations, the following recommendations are proposed:

- **Enhanced Preprocessing**: Improve preprocessing techniques to handle noisy data more effectively, ensuring cleaner input for the summarization models.
- **Advanced Summarization Models**: Explore and integrate more advanced summarization models, such as those based on newer transformer architectures, to further enhance the accuracy and coherence of summaries.
- **Multilingual Support**: Expand the application to support multiple languages, broadening its accessibility and usability across different linguistic groups.
- **Diverse Content Types**: Adapt the application to handle a wider range of content types, including technical, scientific, and creative materials, to make the tool more versatile and effective in various contexts.

CHAPTER 6 : CONCLUSION

6.1 Summary of the Work

This project aim to develop a video and text summarizer that utilizes artificial intelligence to generate summary of video and text as well.

- **VIDEO SUMMAIZER AI**: This technology analysis video content to extract key frames, segments and highlights producing a shorter version that encapsulates the main ideas and events.
- **TEXT SUMMARIZER AI:** This system processes textual content, identifying important sentences and concepts to create brief summary that retains the core message.

6.2 Achievements of the Project

The project achieved several significant milestones:

- **Accurate Video Transcription**: Successfully integrated the hugging face , transformer, pipeline and you tube transcript API, enabling accurate and reliable transcription of YouTube videos and text files.
- **Effective Summarization Algorithms**: Developed robust summarization algorithms that produce coherent, concise, and meaningful summaries from transcribed text.
- **User-Friendly Application**: Created a user-friendly interface using Streamlit, allowing users to easily input video links or upload text files, and receive detailed summaries and text summary outputs.
- **Performance Metrics**: Evaluated the summarization tool using practical metrics such as length ratio, sentence-level similarity, and readability scores, ensuring the generated summaries are accurate and easy to understand.

6.3 Concluding Remarks

The project demonstrates the transformative potential of generative AI in enhancing content accessibility. By automating the transcription and summarization of video and text content, the developed tool provides a valuable resource for users in education, research, and entertainment,

enabling quick and efficient access to critical information. The success of this project lays a strong foundation for future advancements, including improved preprocessing techniques, the incorporation of more advanced summarization models, and the expansion to support multiple

languages and diverse content types. This work underscores the importance of leveraging AI to make digital content more accessible and user-friendly, paving the way for broader applications and continuous improvement in the field of content summarization.

6.4References

- https://youtu.be/HFfXvfFe9F8?feature=shared
- Text Summarization in NLP - GeeksforGeeks
- Summarization - Hugging Face NLP Course
- Chatgpt