VIDEO SUMMARIZER BOT

A MINI-PROJECT REPORT

Submitted By

BHARATH KUMAR L 2116220701043

in partial fulfilment for the course

OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

AUTONOMOUS, CHENNAI 602105

NOV/DEC,2024

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI – 602105

BONAFIDE CERTIFICATE

Certified that this project report **"TITLE"** is the Bonafide work of **"NAME(ROLLNO)"** who carried out the project work for the subject OAI1903-Introduction to Robotic Process Automation under my supervision.

SIGNATURE

Mrs. J. JINU SOPHIA,

SUPERVISOR

Assistant Professor(SG),

Computer Science & Engineering,

Rajalakshmi Engineering College

Thandalam , Chennai 602105

Submitted to Project and Viva Voce Examination for the subject OAI1903-Introduction to Robotic Process Automation held on ……

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Thiru. S. Meganathan, B.E., F.I.E.,** our Vice Chairman **Mr. M. Abhay Shankar, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, M.A., M.Phil., Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S. N. Murugesan, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. Kumar, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, **Mrs. J. Jinu Sophia, M.E., Ph.D.,** Assistant Professor (SG) Department of Computer Science and Engineering for their valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator Professor, **Dr. N. Duraimurugan, M.E., Ph.D.,** Associate Professor and **Mr. B. Bhuvaneswaran, M.E.,** Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project.

BHARATH KUMAR L - 2116220701043

ABSTRACT

Video content has become a cornerstone of modern education, offering students convenient access to lectures and learning materials. However, extracting key insights from lengthy videos poses a challenge, especially in fast-paced learning environments. This project addresses this issue by designing an automated, scalable video summarization solution tailored for educational purposes.

The system leverages UiPath Studio to automate the workflow, beginning with video-to-audio conversion using FFmpeg, followed by transcription via the Google Speech-to-Text API. The transcribed text is summarized using a natural language processing model from UiPath AI Center. Managed through UiPath Orchestrator, the workflow ensures task automation, error handling, and resource optimization, making the process efficient and reliable.

This solution empowers students to focus on essential concepts rather than navigating hours of video and supports educators by delivering concise summaries of their lectures. By combining AI tools with automation platforms, the project demonstrates a significant advancement in educational technology, paving the way for accessible and scalable summarization solutions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
API	Application Programming Interface
FFmpeg	Fast Forward Moving Picture Experts Group
ML	Machine Learning
UiPath	A Robotic Process Automation (RPA) tool
TFS	Text File Storage
STT	Speech-to-Text
MP4	MPEG-4 Video File Format
JSON	JavaScript Object Notation
API Key	Application Programming Interface Key
CSV	Comma Separated Values
WAV	Waveform Audio File Format

CHAPTER - 1

INTRODUCTION

1.1 GENERAL

The rapid advancement of digital technology has revolutionized the way educational content is delivered and consumed. Lecture videos, online tutorials, and e-learning modules have become integral to modern education, providing flexibility and accessibility to learners worldwide. However, the sheer volume and length of video-based content pose significant challenges to students and educators. Reviewing lengthy lecture videos to extract key insights is often tedious, time-consuming, and inefficient, particularly in fast-paced learning environments where quick comprehension is vital.

An automated solution for summarizing video content can significantly enhance the learning experience. By reducing the cognitive load required to sift through hours of material, students can focus on understanding core concepts, and educators can deliver condensed versions of their lessons more effectively. Such solutions not only save time but also provide a scalable approach to managing and utilizing video content in diverse educational settings.

1.2 OBJECTIVE

The primary objective of this project is to design and implement an **automated workflow** that can efficiently summarize lecture videos into concise, meaningful text. This system aims to bridge the gap between the growing reliance on video content and the need for quick information access by leveraging cutting-edge tools and technologies. Specifically, the project focuses on achieving the following goals:

1. **Video-to-Audio Conversion**: Use the open-source tool **FFmpeg** to extract audio from lecture videos while ensuring compatibility with transcription tools.
2. **Audio Transcription**: Utilize the **Google Speech-to-Text API**, a powerful transcription tool known for its accuracy and support for multiple languages and accents, to convert audio into text.
3. **Text Summarization**: Apply a **pre-built model from UiPath AI Center** to condense the transcribed text into concise summaries using advanced natural language processing (NLP) techniques.
4. **Future Enhancement – Workflow Scheduling**: Envision integrating **UiPath Orchestrator** to schedule, manage, and monitor the summarization process, ensuring a continuous and error-free pipeline in future iterations.

This project seeks to provide a robust and scalable solution that reduces manual effort, enhances efficiency, and supports both students and educators in navigating large volumes of educational video content.

1.3 EXISTING SYSTEM

Existing video summarization methods are largely manual or involve proprietary software tools with significant limitations:

1. **Manual Transcription**: In many educational settings, summarization involves manually transcribing lecture videos and extracting key points. This process is labor-intensive, error-prone, and time-consuming.
2. **Proprietary Tools**: Some available tools for video summarization are expensive and not easily accessible to educational institutions, particularly those with limited resources.
3. **Fragmented Workflow**: Current systems often lack integration, requiring separate tools for video processing, transcription, and summarization. This lack of cohesion results in inefficiencies and higher error rates.

These challenges highlight the need for a unified, automated system that combines all necessary components into a seamless workflow, eliminating dependency on manual effort and ensuring scalability.

1.4 PROPOSED SYSTEM

The proposed system offers a **semi-automated and integrated solution** for video summarization, addressing the limitations of existing methods. By combining powerful open-source tools and advanced AI models, the system achieves reliability, scalability, and efficiency. The key features of the proposed system are:

1. **End-to-End Processing**: Automates the process from video input to summarized text output, significantly reducing manual intervention.
2. **Integration of Advanced Tools**: Incorporates FFmpeg for video-to-audio conversion, Google Speech-to-Text API for accurate transcription, and UiPath AI Center for text summarization.
3. **Future-Ready Design**: Envisions using **UiPath Orchestrator** in future enhancements to implement scheduling and task management for continuous and error-free workflows.

By leveraging the latest advancements in automation and AI, the proposed system delivers a scalable and cost-effective solution that empowers students and educators to access essential information from lecture videos quickly and efficiently.

CHAPTER - 2

LITERATURE REVIEW

2.1 GENERAL

Video summarization has gained significant attention in recent years due to the explosion of digital content in various domains, such as media, education, and corporate training. The ability to extract and condense key information from lengthy videos offers tremendous benefits, including time savings, enhanced learning, and improved information accessibility. With the advancement of artificial intelligence (AI) and machine learning (ML), video summarization has evolved from manual techniques to automated processes leveraging state-of-the-art tools and algorithms.

The literature identifies three primary components essential for implementing an effective video summarization system: audio processing, speech-to-text conversion, and text summarization. These components are supported by a range of well-established technologies and methodologies:

Audio Processing Tools

Audio processing is a critical step in video summarization workflows, as it extracts the audio track from video content for further analysis. Among the tools available, **FFmpeg** stands out as a versatile and widely-used open-source software suite for handling multimedia data. FFmpeg supports numerous file formats and provides high performance for video-to-audio conversion, making it a popular choice in research and production environments. Its flexibility and command-line interface make it particularly suited for integration into automated pipelines.

Speech-to-Text Solutions

Speech-to-text conversion is another crucial element, enabling the transformation of spoken content into machine-readable text. The **Google Speech-to-Text API** is one of the most reliable and accurate solutions available today. It utilizes cutting-edge AI models to recognize and transcribe speech with high precision. Notably, it supports multiple languages and accents, making it ideal for diverse applications. Research highlights its ability to handle noisy environments and varying speech patterns, which are common in real-world educational video recordings. The API's scalability and robust cloud-based architecture also make it suitable for large-scale implementations.

Text Summarization Models

Text summarization represents the final step in the workflow, where transcribed text is condensed into concise and meaningful summaries. Modern text summarization approaches leverage **Natural Language Processing (NLP)** techniques, including pre-trained models such as GPT, BERT, and T5, which are known for their ability to understand context and generate coherent summaries. These models are trained on extensive datasets, enabling them to capture the nuances of language and extract key points effectively. In particular, pre-built summarization models available through platforms like **UiPath AI Center** provide ready-to-deploy solutions that integrate seamlessly into automation workflows. These models demonstrate high accuracy and efficiency, ensuring that the generated summaries are both informative and readable.

Integration and Workflow Optimization

Although the individual components of video summarization are well-researched, the integration of these tools into a seamless and automated workflow represents a relatively novel contribution. Existing systems often address isolated aspects of the process, such as transcription or summarization, but lack a unified framework. This fragmentation creates inefficiencies and limits scalability. By combining tools like FFmpeg, Google Speech-to-Text API, and UiPath AI Center into a cohesive pipeline, it is possible to deliver a solution that is both reliable and efficient.

CHAPTER – 3

SYSTEM DESIGN

3.1 SYSTEM FLOW DIAGRAM

The system flow diagram outlines the key steps in the video summarization process. It begins with the user providing the video file path and output audio file name. The system validates the file name, extracts audio using FFmpeg, and invokes a Python script for audio-to-text conversion. The generated text is processed by an ML Skill in UiPath for summarization. The summarized text is then displayed to the user. The diagram highlights decision points, validations, and integrations, ensuring a clear representation of the workflow.

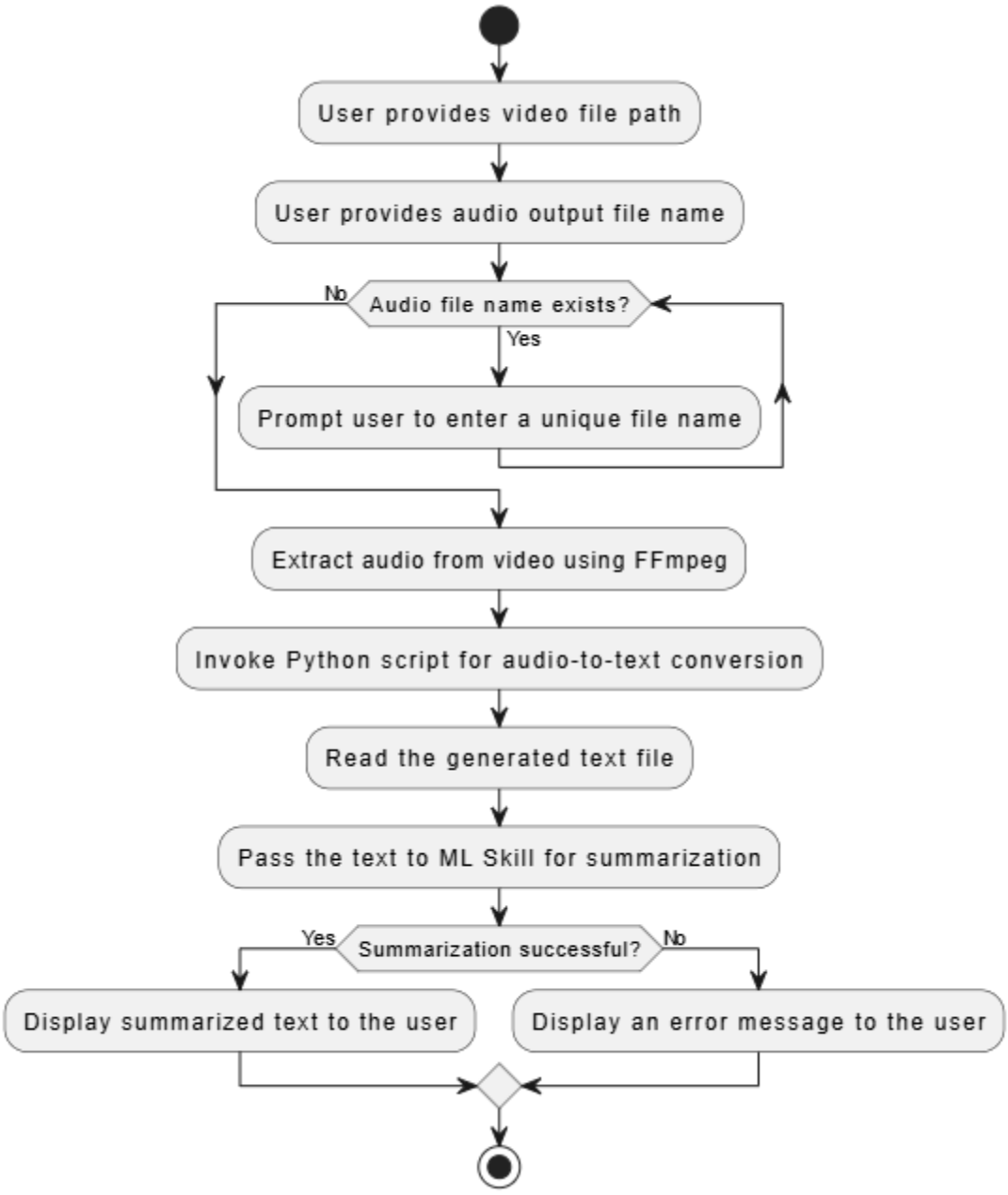

Figure 1 SYSTEM FLOW DIAGRAM

3.2 ARCHITECTURE DIAGRAM

The architecture diagram illustrates the automated video summarization workflow:

1. **Input**: The user provides the video and output audio file names.
2. **Conversion**: FFmpeg converts the video to audio.
3. **Transcription**: Audio is transcribed to text using Google Speech-to-Text API.
4. **Summarization**: Text is summarized using UiPath AI Center's NLP model.
5. **Output**: The summarized text is displayed to the user.
6. **Error Handling**: The system checks if the video exists and notifies the user if it doesn't.

This architecture automates the summarization process efficiently, ensuring quick and accurate results.

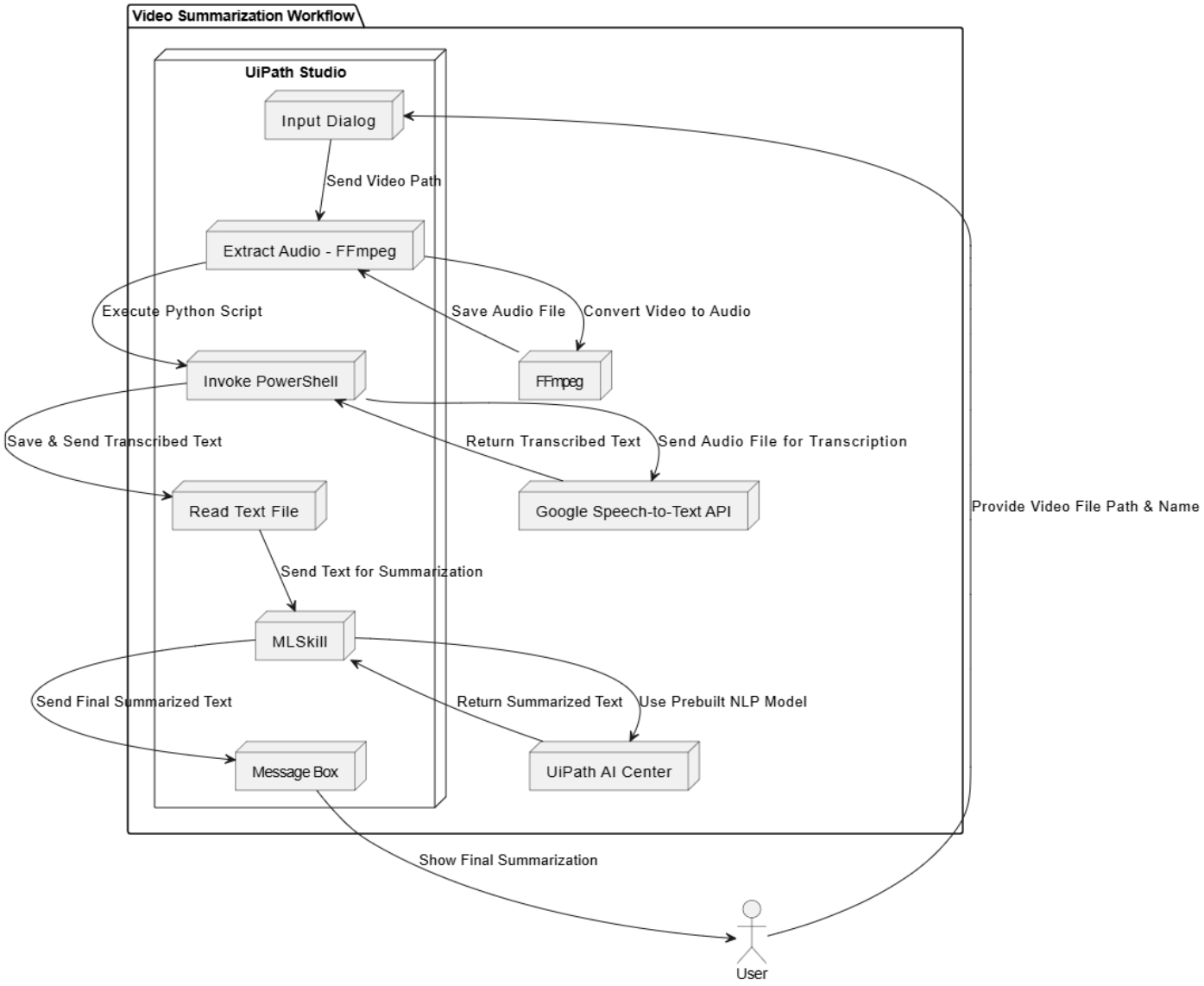

Figure 2 ARCHITECTURE DIAGRAM

3.3 SEQUENCE DIAGRAM

1. **User Interaction**:
   The user inputs the video file path and specifies the output audio file name.
2. **File Validation**:
   The system checks if the file name already exists and prompts the user for a unique name if needed.
3. **Audio Extraction**:
   The workflow uses FFmpeg to extract audio from the video file.
4. **Audio-to-Text Conversion**:
   The extracted audio is processed by a Python script to convert it into text and save it.
5. **Text Summarization**:
   The text is sent to an ML Skill for summarization, and the summarized text is returned.
6. **Output**:
   The workflow displays the summarized text to the user.

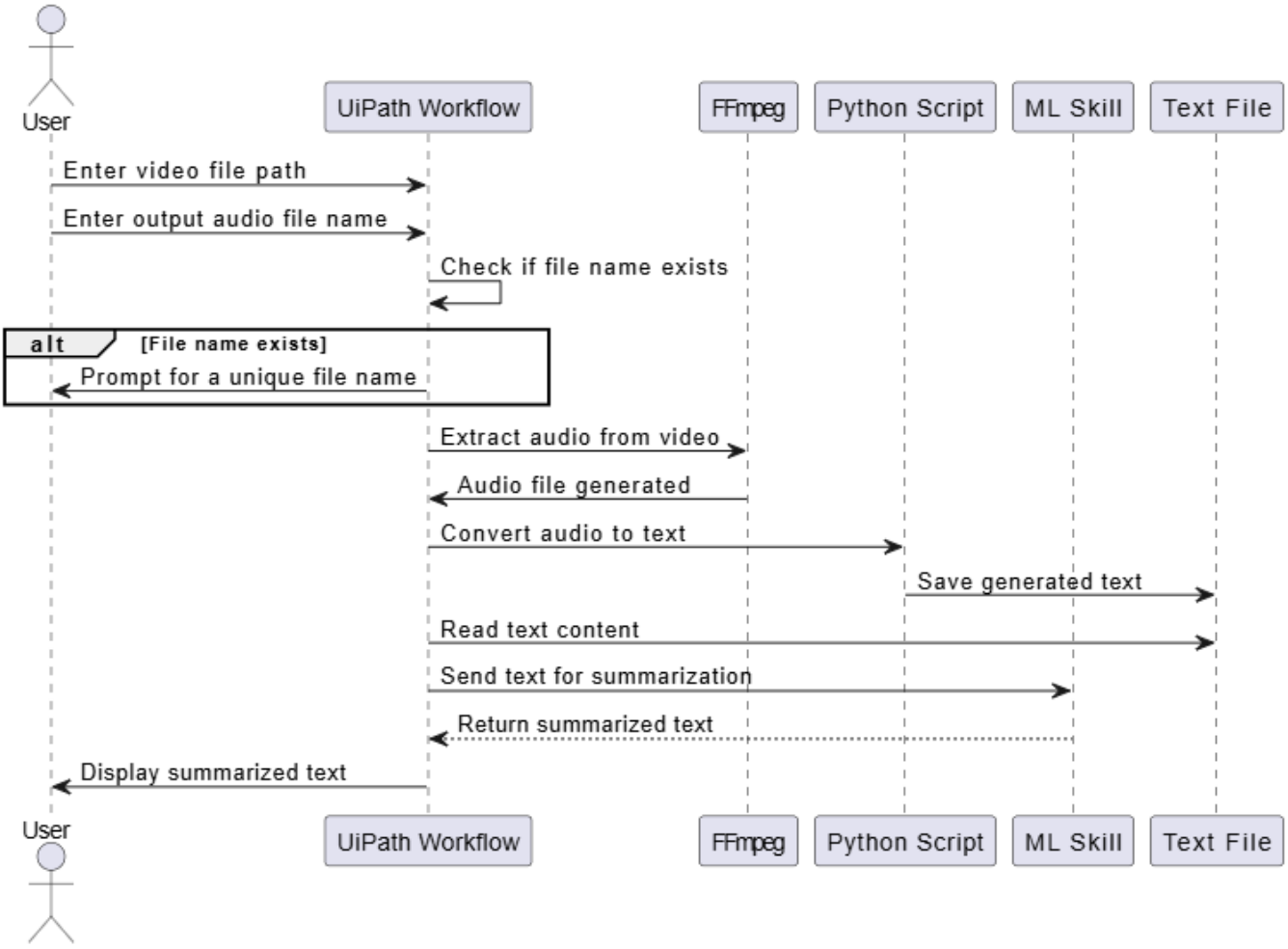

Figure 3 SEQUENCE DIAGRAM

CHAPTER – 4

PROJECT DESCRIPTION

4.1 METHODOLOGIES

This project integrates several advanced technologies and tools to automate the process of video summarization. The entire process is divided into three core modules, each designed to perform a specific task. These modules work together seamlessly to provide an efficient and scalable solution for summarizing educational video content.

4.1.1 MODULES

1. Video-to-Audio Conversion

- **Tool**: FFmpeg
- **Process**: The first module focuses on converting video files into audio. FFmpeg, a powerful open-source multimedia framework, is used to extract the audio stream from the video file. This step ensures the video content is in a format suitable for transcription. The audio is typically converted into a standard file format such as .wav or .mp3. FFmpeg is preferred for its speed, reliability, and compatibility with various video and audio formats. This module forms the foundation for the subsequent transcription step, ensuring the video content is accessible for further processing.

2. Audio Transcription

- **Tool**: Google Speech-to-Text API
- **Process**: Once the audio is extracted, the next step is transcribing the audio into text. This is achieved using the Google Speech-to-Text API, a highly accurate tool for converting spoken language into written text. It supports multiple languages and accents, ensuring wide accessibility for diverse educational content. The transcription process involves processing the audio file through the API, which uses advanced machine learning models to detect and transcribe speech accurately. The output is a text file containing the transcribed content, which will then be used for summarization.

3. Text Summarization

- **Tool**: UiPath AI Center pre-built model

- **Process**: After transcription, the next step is to summarize the transcribed text. This is done using a pre-built natural language processing (NLP) model from UiPath AI Center. The model leverages advanced machine learning techniques to analyze and condense the transcribed text into a concise summary. The NLP model is trained to understand the context of the text, identify the key points, and present them in a coherent format. This module is critical as it transforms lengthy transcriptions into easily digestible summaries, allowing users to quickly grasp the essential information from the video.

These modules form the core of the automated video summarization system, offering a complete solution for converting lecture videos into concise summaries. By combining the strengths of FFmpeg, Google Speech-to-Text, and UiPath AI Center, the project delivers an efficient, scalable, and accurate tool for educational content summarization.

CHAPTER – 5

OUTPUT SCREENSHOTS

5.1 IMPLEMENTATION WORKFLOW

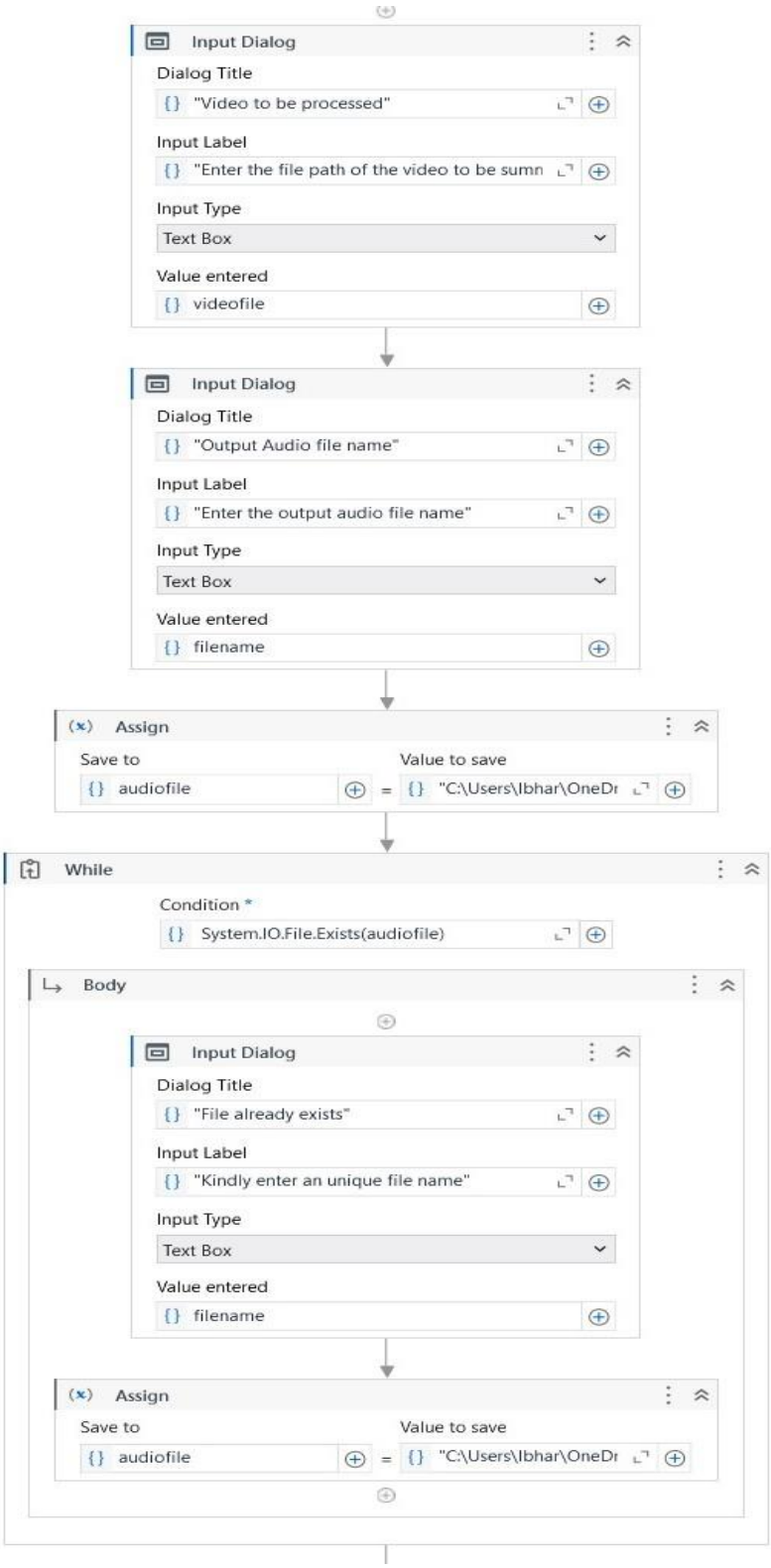

Figure 4.1 IMPLEMENTATION WORKFLOW

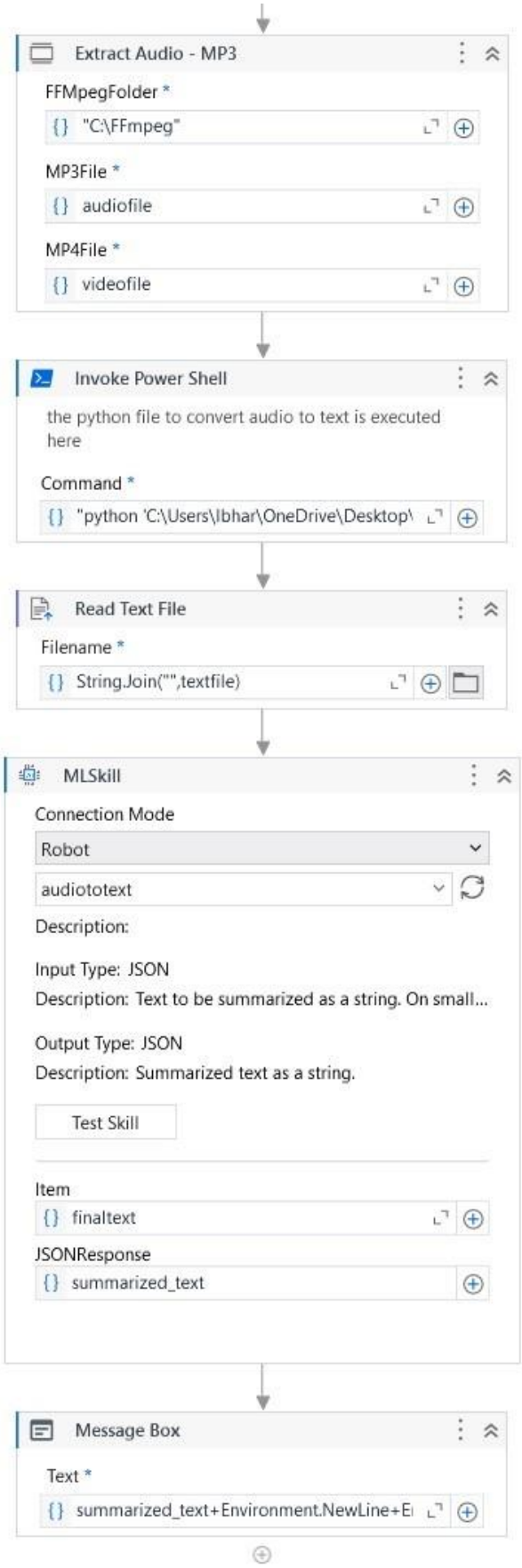

Figure 4.2 IMPLEMENTATION WORKFLOW

5.2 TRANSCRIBED TEXT

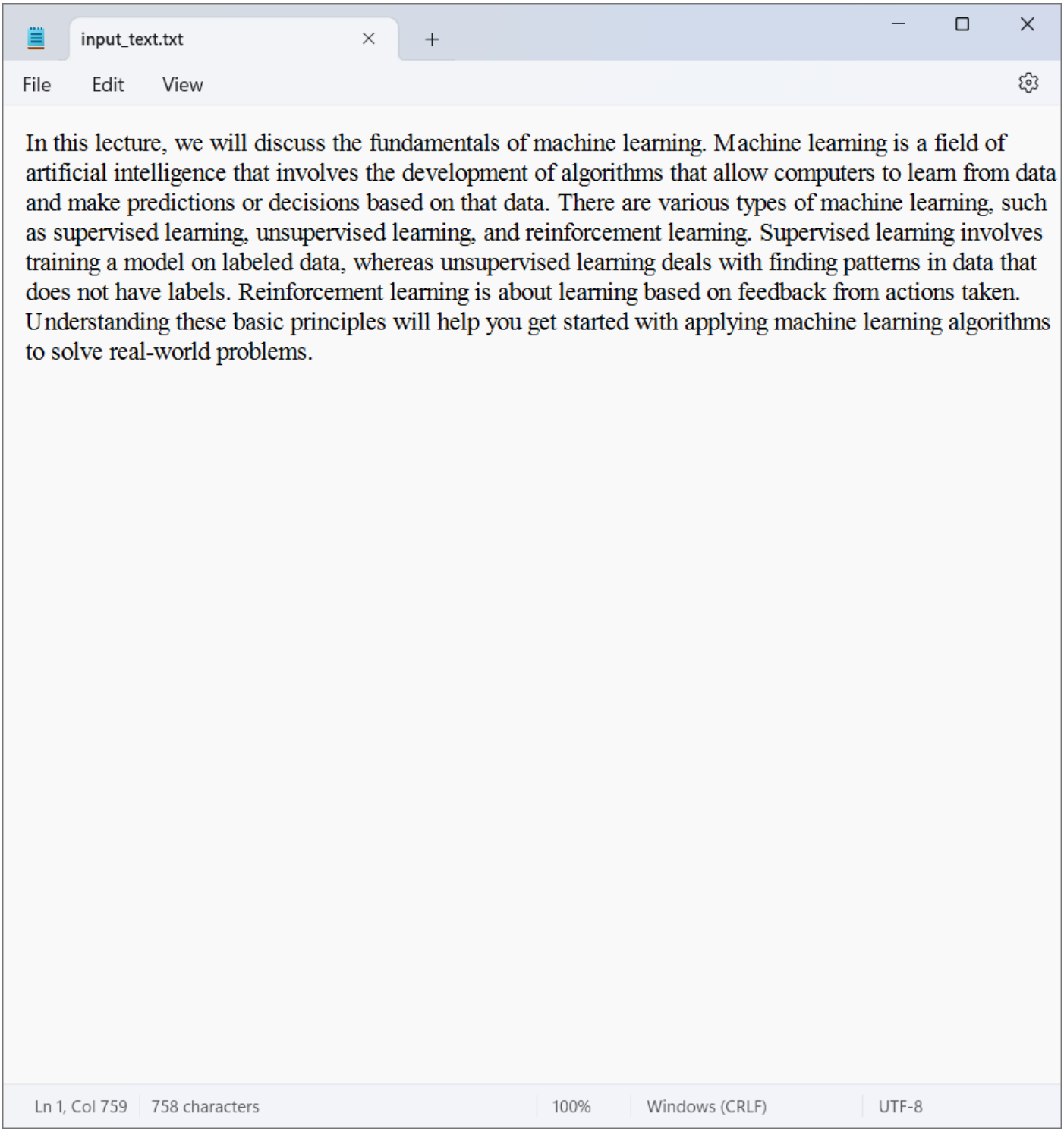

Figure 5 TRANSCRIBED TEXT

5.3 SUMMARIZED TEXT

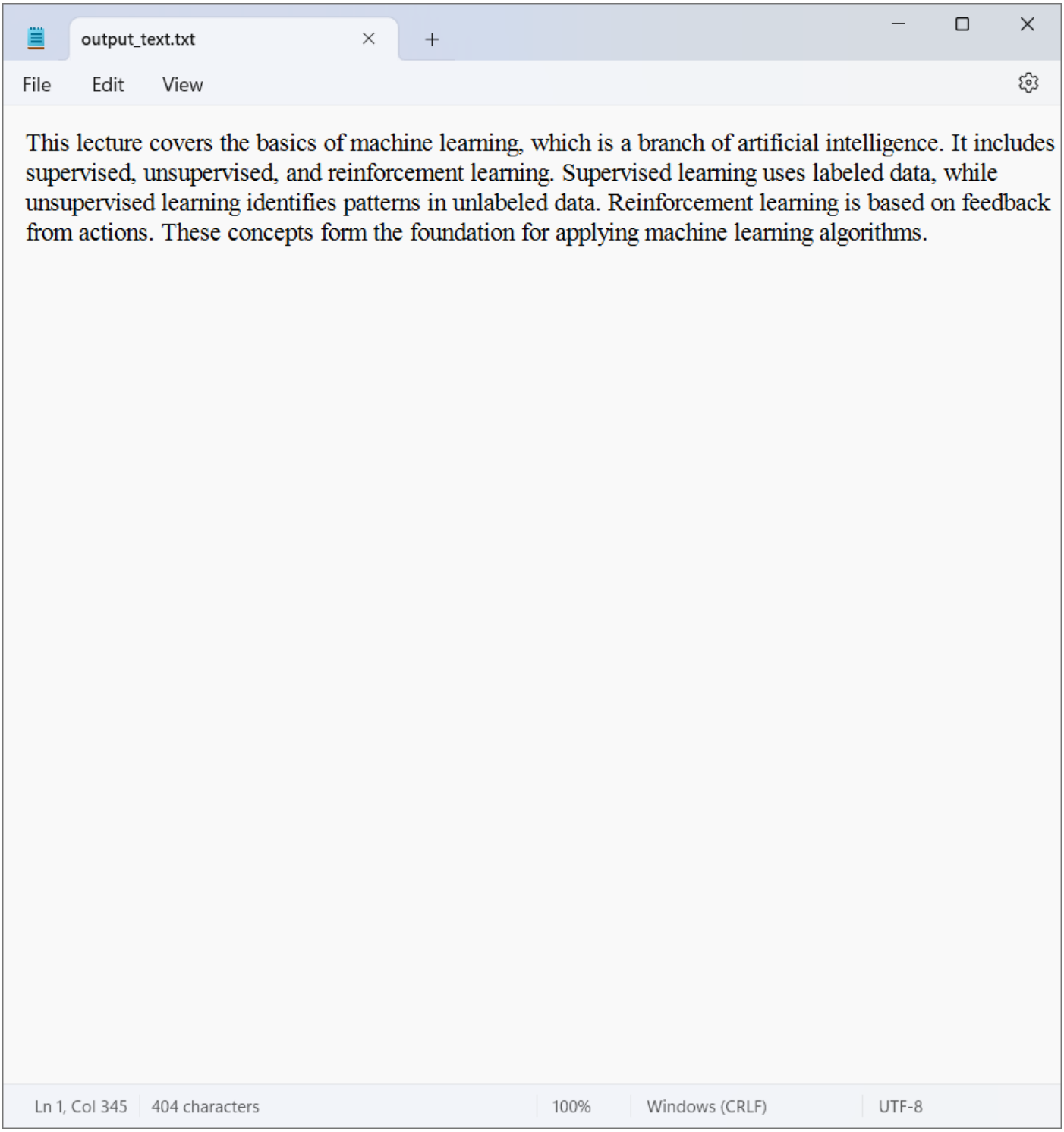

Figure 6 SUMMARIZED TEXT

CHAPTER – 6

CONCLUSIONS

6.1 GENERAL

This project successfully demonstrates the automation of video summarization using UiPath, FFmpeg, Python, and machine learning techniques. By combining these technologies, the system efficiently extracts audio from video files, converts speech to text, and summarizes the content, thus reducing manual effort and processing time. The workflow is streamlined, with key activities such as audio extraction, text conversion, and summarization being handled automatically, providing a seamless experience for the user.

The system's design is modular and adaptable, allowing for potential future enhancements and improvements. The integration of UiPath provides a flexible automation environment that can easily be expanded to handle more complex scenarios or larger datasets. This solution is beneficial not only for personal or educational use but also for industries that require quick content summarization, such as media, education, and corporate sectors.

As the demand for processing large volumes of video content continues to grow, the ability to automate tasks like video summarization becomes increasingly important. This system offers a reliable and scalable solution that can be expanded and enhanced to meet evolving needs, providing value through time-saving automation and improved productivity in content analysis.

APPENDICES

Appendix A: *Workflow Design*

This section includes the design and workflow of the video summarization process. The steps are as follows:

1. User inputs the video file path and specifies the audio output file name.
2. The system validates the file name to avoid overwriting.
3. FFmpeg is used to extract the audio from the video file.
4. A Python script converts the extracted audio to text.
5. The generated text is summarized using an ML Skill in UiPath.
6. The summarized text is displayed to the user.

Refer to the **system flow diagram***(Figure 1)* and **sequence diagram***(Figure 3)* in the report for detailed representations of the process.

Appendix B: *Tools and Technologies Used*

1. **UiPath Studio**: For designing and automating the workflow.
2. **FFmpeg**: For extracting audio from video files.
3. **Python Script**: For converting audio to text using speech-to-text libraries.
4. **ML Skill**: Used for text summarization in UiPath.
5. **File System**: For validating file paths and storing generated files.

Appendix C: *System Requirements*

1. **Software**:
    - UiPath Studio (Version 24.10.5 or above)
    - Python 3.7 or above
    - FFmpeg
2. **Hardware**:
    - Processor: Intel i3 or above
    - RAM: 4 GB minimum
    - Disk Space: 500 MB for tools and dependencies

Appendix D: *Python Script (Audio-to-Text Conversion)*

```
def transcribe_audio(file_path):
    client = speech.SpeechClient()
    with io.open(file_path, "rb") as audio_file:
        content = audio_file.read()
    audio = speech.RecognitionAudio(content=content)
    config = speech.RecognitionConfig(encoding=speech.RecognitionConfig.
            AudioEncoding.LINEAR16, sample_rate_hertz=16000, language_code="en-US")
    try:
        operation = client.long_running_recognize(config=config, audio=audio)
        response = operation.result(timeout=600)
        full_transcript = []
        for result in response.results:
            transcript = result.alternatives[0].transcript
            full_transcript.append(transcript)
        output_file_path = file_path.rsplit('.', 1)[0] + "_transcript.txt"
        with open(output_file_path, 'w', encoding='utf-8') as f:
            for line in full_transcript:
                f.write(line + '\n')
        print(f"Transcription complete. Output saved to {output_file_path}")
    except Exception as e:
        print(f"Error during transcription: {e}")
```

Figure 7 PYTHON SCRIPT

Appendix E: *Example Input and Output*

1. **Input**:
    - Video file path: C:\Videos\example.mp4
    - Output audio file name: example_audio.mp3
2. **Intermediate Files**:
    - Extracted audio: example_audio.mp3
    - Converted text: example_text.txt
3. **Output**:
    - Summarized text: "This video explains the process of automation using UiPath."

REFERENCES

[1] https://docs.uipath.com/ai-center/automation-cloud/latest/user-guide/out-of-the-box-packages

[2] https://dataintegration.info/using-ffmpeg-with-google-cloud-speech-to-text

[3] https://ffmpeg.org/documentation.html

[4] https://docs.uipath.com/

[5] https://pypi.org/project/SpeechRecognition/

[6] https://cloud.google.com/getting-started

[7] https://cloud.google.com/docs/authentication/api-keys

[8] https://cloud.google.com/speech-to-text/docs/auth

[9] https://cloud.google.com/speech-to-text/docs/quickstart