TEXT SUMMERIZATION BOT

A PROJECT REPORT

Submitted by

Arun Mohan S (220701029)

in partial fulfillment for the course

OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM

CHENNAI – 602 105

NOVEMBER 2024

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report **" TEXT SUMMERIZATION BOT "** is the bonafide work of **"Arun Mohan S (220701029)"** who carried out the project work for the subject OAI1903-Introduction to Robotic Process Automation under my supervision.

Mrs. J. JINU SOPHIA,
SUPERVISOR
Assistant Professor
Department of
Computer Science and Engineering
Rajalakshmi Engineering College
Rajalakshmi Nagar
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

ABSTRACT

The "**Text Summarization Bot**" is an RPA solution designed to automate the extraction, summarization, and sharing of student-submitted notes in PDF format. Leveraging UiPath's capabilities, the bot begins by extracting text from PDFs using the Extract PDF activity. The extracted text is then processed through UiPath's Generative AI activity, which condenses the content into concise, meaningful summaries. These summaries are stored locally for easy access and future use.

The final step involves automating the distribution of the summarized content via email using the SMTP protocol. This streamlined workflow minimizes manual effort, enhances productivity, and ensures timely delivery of educational content. By integrating RPA and AI technologies, the project simplifies the management of large volumes of information, making it an efficient tool for educational and administrative purposes.

The bot is designed with scalability in mind, allowing it to handle a wide variety of input formats and large datasets. The integration of Generative AI ensures that the summaries retain the essence of the original text, offering valuable insights in a condensed form. This makes it particularly beneficial for educators and administrators who need quick access to key information.

Additionally, the bot's automated process reduces the likelihood of errors associated with manual work, improving accuracy and reliability. The solution is also versatile and can be adapted for other domains such as business reporting, legal document processing, and research analysis. By automating repetitive tasks, the project demonstrates the power of combining RPA and AI to deliver impactful results in a variety of settings.

.

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 Department of Computer Science and Engineering for their valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators, **Dr. N.Durai Murugan, 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.

Arun Mohan S (220701029)

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	**iii**
	LIST OF FIGURES	**vi**
1.	**INTRODUCTION**	**1**
	1.1 GENERAL	1
	1.2 OBJECTIVE	2
	1.3 EXISTING SYSTEM	2
	1.4 PROPOSED SYSTEM	3
2.	**LITERATURE REVIEW**	**5**
3.	**SYSTEM DESIGN**	**8**
	3.1.1 SYSTEM FLOW DIAGRAM	8
	3.1.2 ARCHITECTURE DIAGRAM	9
	3.1.3 SEQUENCE DIAGRAM	10
4.	**PROJECT DESCRIPTION**	**11**
	4.1 MODULES	11
	4.1.1 PDF Text Extraction Module	11
	4.1.2 Text Summarization Module	11
	4.1.3 Local Storage Module	12
	4.1.4 Email Sending Module	12
	4.1.5 Manager Module	12
5.	**OUTPUT**	**13**
6.	**CONCLUSION AND FUTURE ENHANCEMENT**	**18**
	6.1 Conclusion	18
	6.2 Future Enhancement	19
	REFERENCES	**20**

LIST OF FIGURES

Figure No	Figure Name	Page No
3.1	System Flow Diagram	8
3.2	Architecture Diagram	9
3.3	Sequence Diagram	10
5.1	Output 1	13
5.2	Output 2	14
5.3	Output 3	14
5.4	Output 4	15
5.5	Output 5	15
5.6	Output 6	16
5.7	Output 7	16
5.8	Output 8	17

CHAPTER 1
INTRODUCTION

1.1. GENERAL

In an increasingly digital academic environment, managing and processing large volumes of student-submitted educational content can be a challenging task. Manually extracting, summarizing, and sharing these documents, often submitted in PDF format, is both time-consuming and prone to human error. This project, **Text Summarization Bot**, utilizes UiPath's Robotic Process Automation (RPA) capabilities to automate these tasks, creating a streamlined process for handling student notes.

The project automates the workflow by first extracting text from PDFs using UiPath's Extract PDF activity. The extracted text is then summarized using UiPath's Generative AI activity, which produces concise and meaningful summaries of the original content. These summaries are stored locally for easy access and subsequently sent to designated recipients via email using the SMTP protocol.

This automated process eliminates the need for manual intervention, ensuring faster processing, improved accuracy, and timely delivery of summarized content. By condensing extensive information into concise formats, the bot enables educators and administrators to focus on more critical tasks, thereby improving the efficiency of managing and sharing educational content.

Through this RPA-driven approach, the Text Summarization Bot demonstrates how automation and artificial intelligence can simplify repetitive, labor-intensive tasks. This system enhances productivity and reliability, providing educational institutions with a powerful tool to optimize their workflows and better manage their resources.

1.2. OBJECTIVE

The objective of this project is to automate the extraction, summarization, and sharing of student-submitted notes in PDF format. Using UiPath for robotic process automation (RPA), the system will extract text from PDFs, process the data, and apply Generative AI to summarize the content into a concise format. This approach eliminates manual effort, enhances operational efficiency, and provides a streamlined method for managing educational content. By automating the workflow, the system ensures accurate and timely processing of documents, making it easier for educators and administrators to access summarized information. The summarized content is shared via email using the SMTP protocol, enabling quick dissemination to the intended recipients. This automation not only simplifies the handling of large volumes of data but also enhances productivity, improves accuracy, and supports efficient communication, ultimately fostering a more effective educational environment.

1.3. EXISTING SYSTEM

The existing system for processing and summarizing student-submitted notes is largely manual, involving the extraction of text from PDFs and summarizing it by hand. This process is labor-intensive, time-consuming, and prone to human error, particularly when dealing with large volumes of submissions. While some tools may allow for basic text extraction or editing, they do not offer automated summarization or streamlined workflows for content sharing. As a result, educators and administrators face significant challenges in managing and distributing summarized content efficiently. The lack of automation leads to delays, inconsistent results, and increased workload, limiting their ability to focus on other critical tasks. This inefficiency hampers productivity and makes it difficult to deliver timely and concise information in an educational setting.

1.4. PROPOSED SYSTEM

The proposed system for the **Text Summarization Bot** is designed to automate the workflow of extracting, summarizing, and sharing student-submitted notes, utilizing UiPath's Robotic Process Automation (RPA) capabilities. The system is composed of four primary modules, each responsible for different stages of the process to ensure accuracy, efficiency, and timely delivery of summarized content..

1. **PDF Text Extraction Module**: This module automates the extraction of text from student-submitted PDFs using UiPath's Extract PDF acti*vity*. It handles structured and unstructured PDF formats, ensuring accurate extraction of textual content for further processing. This module serves as the initial stage of the workflow, where raw data is captured and prepared for summarization.

2. **Text Summarization Module:** After extracting the text, the summarization module leverages UiPath's Generative AI capabilities to condense the content into a concise and meaningful summary. The module is designed to retain the essential details while significantly reducing the length of the text.

3. **Local Storage Module**: Once the summarization is complete, the summarized content is stored locally in a structured format. This module ensures that the content is securely saved for future access or further use. The storage system is designed to handle a large volume of summarized documents efficiently, ensuring data consistency and reliability.

4. **Email Sending Module**: The final stage of the workflow automates the distribution of summarized notes to designated recipients via email. Using UiPath's SMTP protocol, this module ensures timely delivery of the content while allowing customization of email templates and recipient lists. This enables educators and administrators to communicate efficiently and effectively.

The proposed system eliminates the need for manual intervention, significantly reducing the time and effort required for processing student notes. By automating the end-to-end workflow, this project enhances productivity, ensures accuracy, and simplifies the dissemination of educational content. The system demonstrates the potential of RPA and AI in optimizing academic workflows and improving the management of educational resources.

CHAPTER 2

LITERATURE REVIEW

The application of automation and artificial intelligence (AI) in the processing of documents, particularly for text extraction and summarization, has seen significant advancements in recent years. In educational settings, managing and distributing student notes is a labor-intensive process that benefits greatly from automation. Several studies have explored the integration of Robotic Process Automation (RPA) and AI in enhancing the efficiency of document handling, summarization, and distribution.

The extraction of text from PDFs is a critical step in many document management systems. Numerous studies have focused on optimizing the extraction process to handle diverse document formats and content. Liu et al. [1] present an in-depth analysis of methods for extracting text from PDF files, comparing various extraction tools and highlighting the challenges posed by complex layouts, mixed content, and image-based documents. The study emphasizes the importance of using advanced techniques like Optical Character Recognition (OCR) for image-based PDFs and the use of AI for text. Once the text is extracted, summarization becomes the next challenge. AI-powered text summarization has been widely explored in the context of educational content management. Research by Gupta and Sharma [2] discusses the application of machine learning models to summarize educational documents, including course materials and student notes.

They describe how natural language processing (NLP) models like transformers can be trained to summarize long documents into concise summaries while maintaining the essence of the original content. These AI-driven models have been shown to significantly reduce the manual effort required to condense large amounts of text.

Further research by Zhang et al. [3] explores the application of deep learning techniques in automatic summarization, particularly focusing on the use of generative models. Their work demonstrates how AI systems, specifically those utilizing GPT-based architectures, can produce human-like summaries by understanding the underlying structure and context of the text. These advancements in AI-based summarization have paved the way for automated systems like the one in this project, which can process educational PDFs, extract relevant content, and generate concise summaries.

The automation of educational content distribution has been a growing area of interest. The integration of RPA with AI for handling and distributing student notes is discussed in a study by Patel et al. [4], which explores how automated systems can streamline the management of student materials. Their system uses RPA to extract text from PDF files, apply AI for summarization, and send the summarized content via email to designated recipients. This research emphasizes how RPA reduces the workload on administrative staff and improves the efficiency of managing large volumes of student submissions.

Similarly, Sharma and Yadav [5] propose a framework for automating the processing of student-provided documents, which includes the use of RPA for text extraction and email automation for content distribution. Their system integrates AI summarization techniques to generate summaries that are then shared with students and faculty. The study highlights the benefits of automation in reducing human error, enhancing the speed of document processing, and ensuring timely delivery of summarized content to recipients.

Email automation is another critical component of this project, as it enables the automatic distribution of summarized content to recipients. Research by Singh et al. [6] explores the role of RPA in email communication, particularly in automating the sending of reports and educational content. Their study discusses how RPA bots can be configured to send personalized emails, attach relevant documents, and ensure that recipients receive the information promptly. This capability is essential for educational institutions where large volumes of content need to be distributed quickly and accurately.

The integration of RPA with SMTP protocols for email automation has been widely recognized for its ability to streamline communication. According to Kumar et al. [7], automating email distribution using RPA not only reduces manual effort but also enhances operational efficiency by ensuring that emails are sent on time and without error. This has proven particularly beneficial in educational environments where the timely distribution of information is crucial.

CHAPTER 3

SYSTEM DESIGN

3.1. SYSTEM FLOW DIAGRAM

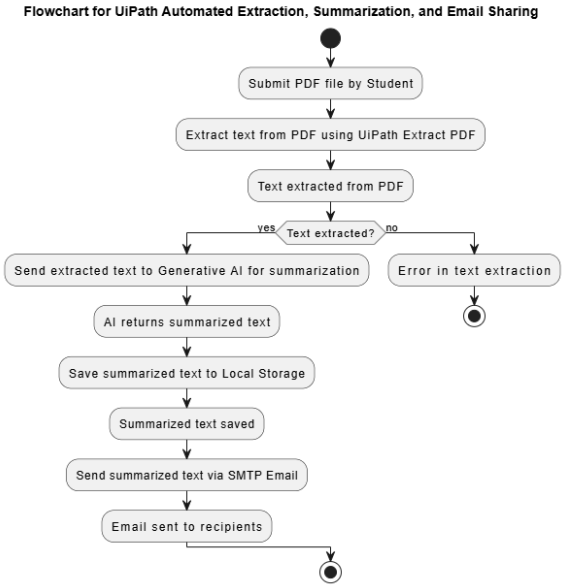

Fig. 3.1. System Flow Diagram

This workflow outlines an automated process designed to extract, summarize, and share text from PDF files using UiPath and Generative AI. It begins with a student submitting a PDF document for processing. UiPath then extracts text from the file. A decision point checks if the extraction was successful; if not, an error is logged, and the workflow terminates. If the text is successfully extracted, it is sent to a Generative AI model (e.g., GPT) for summarization. The AI generates a concise summary, which is saved to local storage for future use or reference. Finally, the summarized text is sent to designated recipients via email using an SMTP service, completing the process. This workflow streamlines document handling, reducing manual effort while ensuring reliability.

3.2. ARCHITECTURE DIAGRAM

An architecture diagram is a graphical representation of a set of concepts, that are part of an architecture including their principle, elements and components.

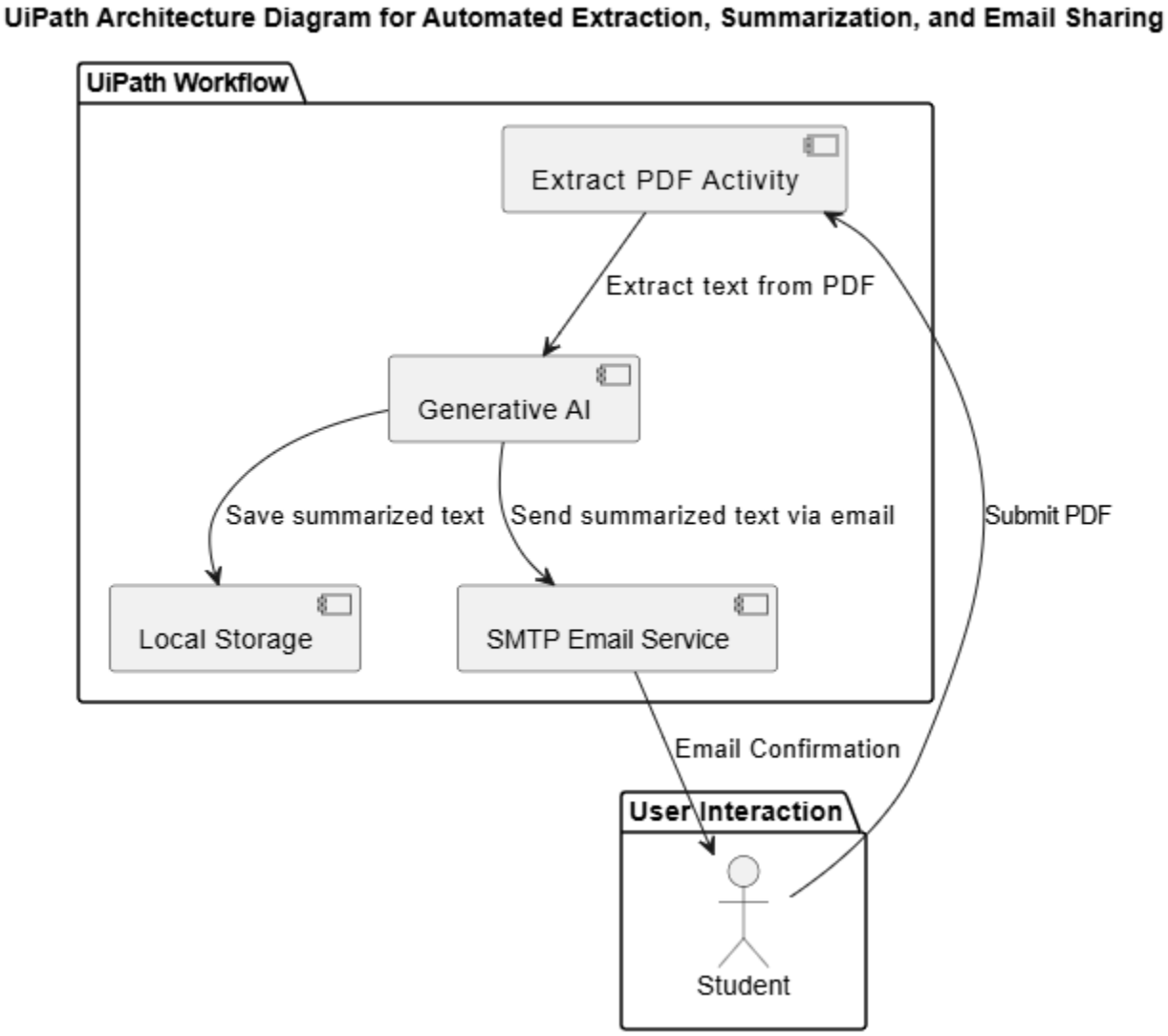

Fig. 3.2. Architecture Diagram

3.3. SEQUENCE DIAGRAM

This sequence diagram illustrates the interaction between various entities involved in automating the extraction, summarization, and email-sharing of PDF content using UiPath and Generative AI. Below is an explanation of each step:

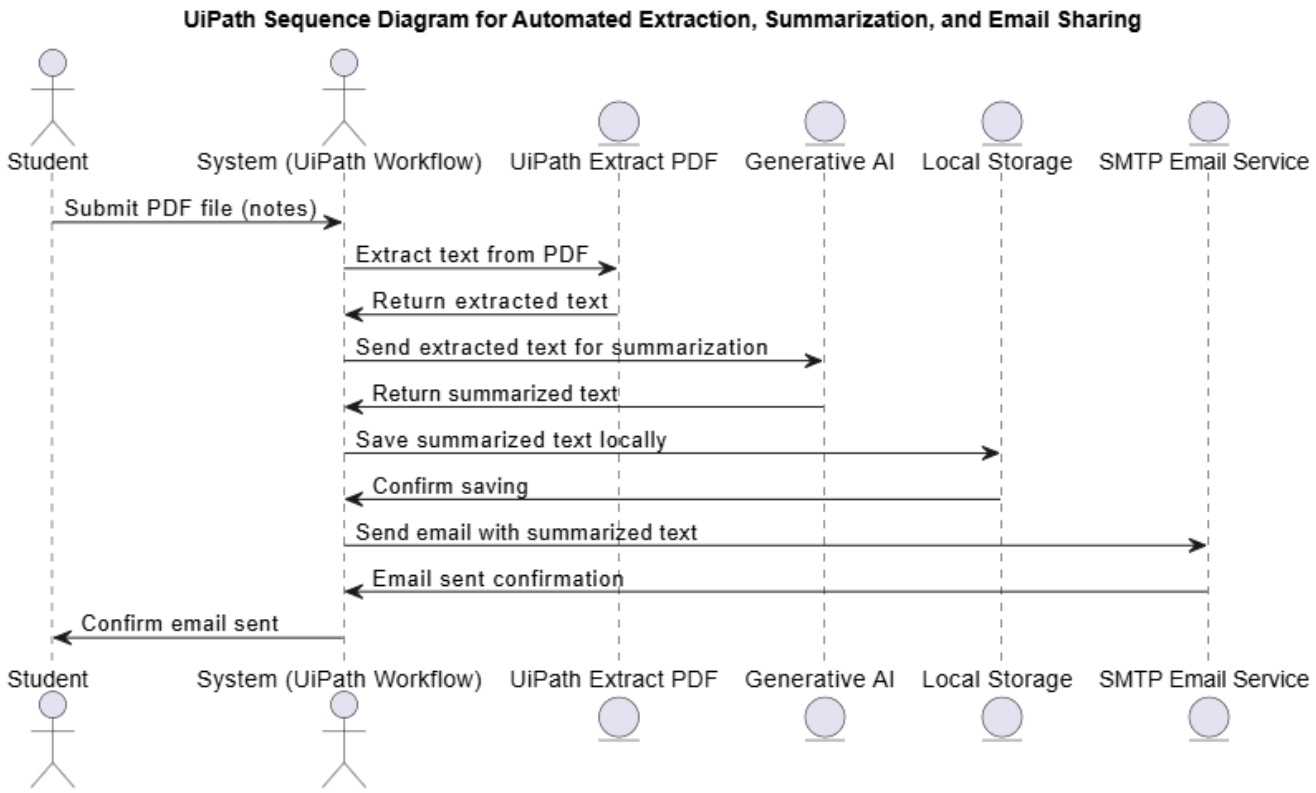

Fig. 3.3. Sequence Diagram

1. Submit PDF File

- Actor(s): Student and System (UiPath Workflow)
- The process begins with the student submitting a PDF file (e.g., notes) to the UiPath automation system for processing.

2. Extract Text from PDF

- Actor(s): UiPath Workflow and UiPath Extract PDF
- UiPath's Extract PDF module processes the submitted file to extract the text

3. Return Extracted Text

- Actor(s): UiPath Extract PDF and UiPath Workflow
- After extracting the text, the extracted content is returned to the UiPath workflow for further processing.

4. Send Extracted Text for Summarization

- Actor(s): UiPath Workflow and Generative AI
- The extracted text is sent to a Generative AI model (e.g., GPT) for summarization. This step involves formatting the content and transmitting it to the AI for processing.

5. Return Summarized Text

- Actor(s): Generative AI and UiPath Workflow
- The Generative AI processes the text and returns a summarized version to the UiPath workflow.

6. Save Summarized Text Locally

- Actor(s): UiPath Workflow and Local Storage
- The summarized text is saved to a local storage system for record-keeping or future access.

7. Confirm Saving

- Actor(s): Local Storage and UiPath Workflow
- Local storage confirms to the UiPath workflow that the summarized text has been successfully saved.

8. Send Email with Summarized Text

- Actor(s): UiPath Workflow and SMTP Email Service
- The summarized text is sent to the intended recipients through an email service configured with SMTP.

CHAPTER 4

PROJECT DESCRIPTION

4.1. MODULES

PDF TEXT EXTRACTION MODULE:

This module is responsible for extracting text content from PDF files submitted by students. Using UiPath's Extract PDF activity, the system processes student-provided PDF documents, regardless of format or structure, to retrieve textual data. It ensures that the extracted content is accurately captured, even from complex layouts, and prepares it for further processing. The module is capable of handling both text-heavy and image-heavy PDFs efficiently, providing the foundation for subsequent summarization and sharing tasks.

TEXT SUMMARIZATION MODULE:

Once text is extracted from the PDF, this module sends the content to a Generative AI model for summarization. The system uses AI capabilities to condense the extracted text into concise, meaningful summaries, ensuring the information is easy to understand while retaining critical details. This module streamlines the process of converting verbose notes into digestible content, reducing manual summarization effort and enhancing usability for recipients.

LOCAL STORAGE MODULE:

The Local Storage module stores the summarized text securely for future reference or use. After the summarization process, the output is saved in an organized format (e.g., text files or database entries) on local or cloud-based storage systems. This module ensures the content is accessible, enabling easy retrieval for sharing or archiving. Proper storage also supports traceability and audit requirements if needed.

EMAIL SENDING MODULE

This module facilitates the automated sharing of summarized content with designated recipients via email. Using SMTP protocols, the system formats and sends the summarized text to specified email addresses. It ensures seamless communication by automating the email-sending process, including error handling and confirmation of delivery. This module is particularly useful for sharing summarized educational content with students, instructors, or other stakeholders.

ADMIN MODULE

The Admin module allows administrators or managers to oversee the process and manage the workflow. Administrators can track submissions, monitor summarization outputs, and ensure that emails are sent to the correct recipients. This module also includes error reporting and troubleshooting features, ensuring that the entire process remains smooth and reliable. It empowers managers to optimize and improve the workflow based on feedback and insights from the system.

CHAPTER 5

OUTPUT SCREENSHOTS

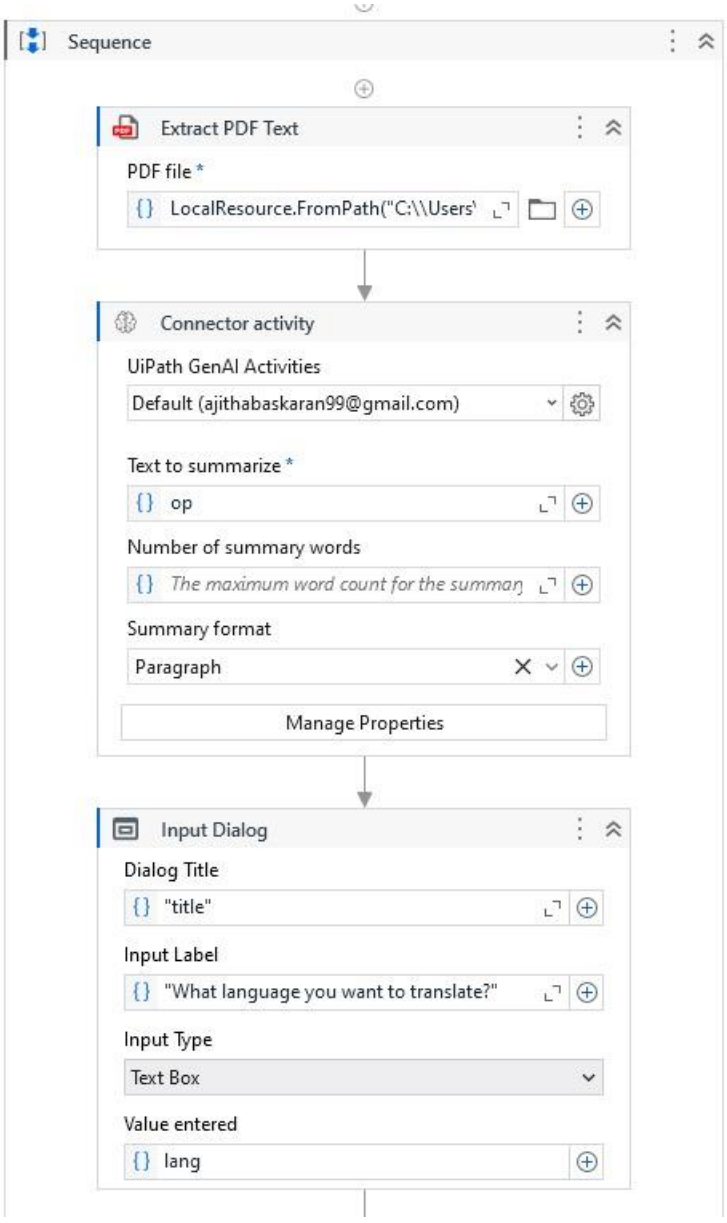

Fig. 5.1.

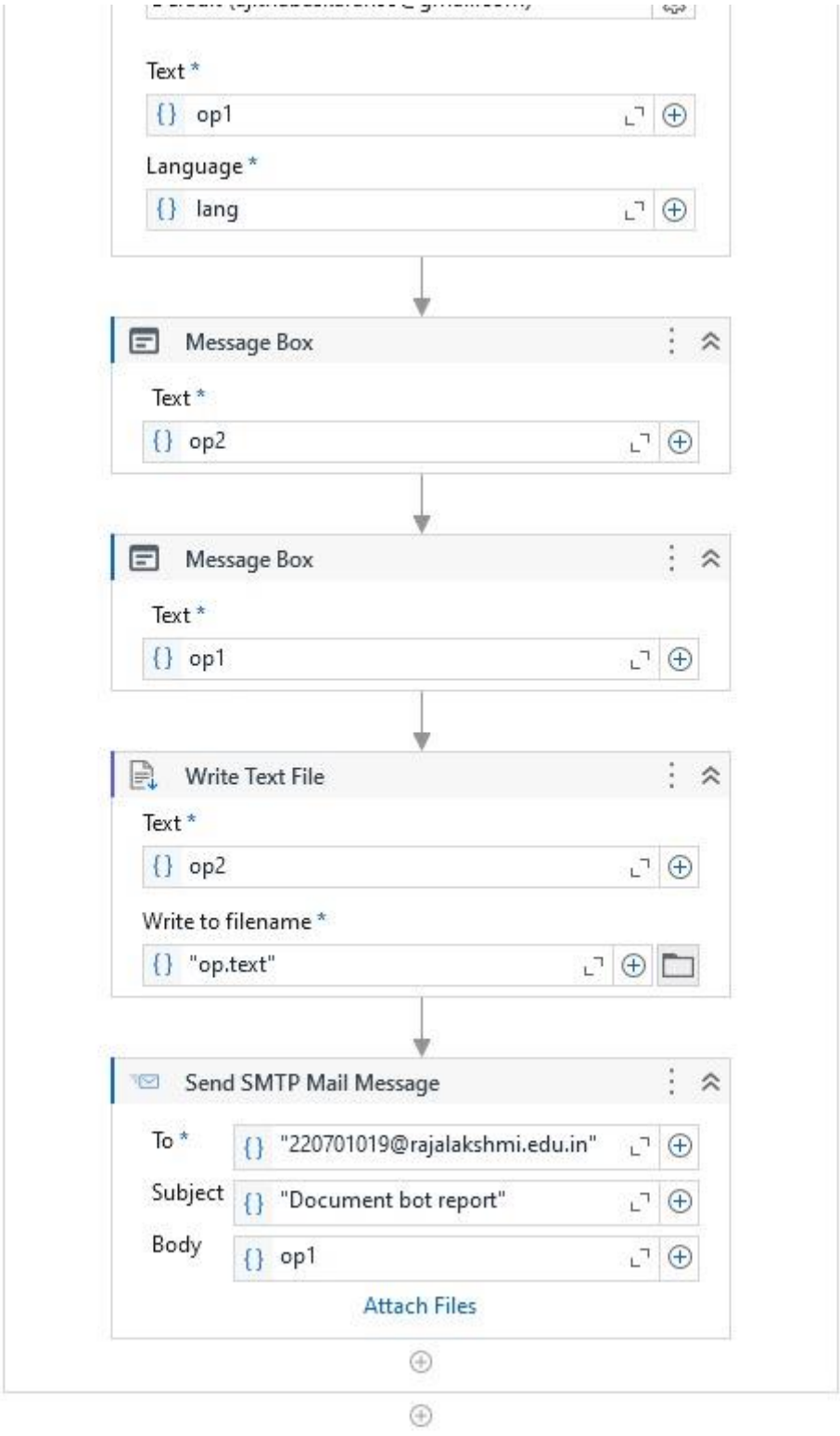
Text *
op1
Language *
lang
Message Box
Text *
op2
Message Box
Text *
op1
Write Text File
Text *
op2
Write to filename *
"op.text"
Send SMTP Mail Message
To *
"220701019@rajalakshmi.edu.in"
Subject
"Document bot report"
Body
op1
Attach Files

Fig. 5.2

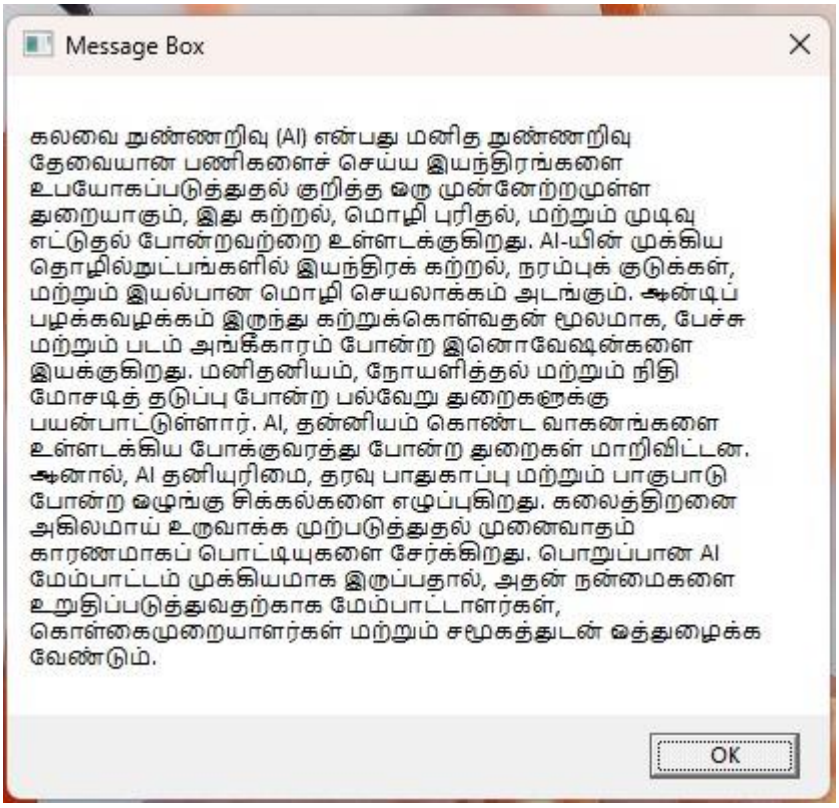
Message Box
கலவை நுண்ணறிவு (AI) என்பது மனித நுண்ணறிவு தேவையான பணிகளைச் செய்ய இயந்திரங்களை உபயோகப்படுத்துதல் குறித்த ஒரு முன்னேற்றமுள்ள துறையாகும், இது கற்றல், மொழி புரிதல், மற்றும் முடிவு எட்டுதல் போன்றவற்றை உள்ளடக்குகிறது. AI-யின் முக்கிய தொழில்நுட்பங்களில் இயந்திரக் கற்றல், நரம்புக் குடுக்கள், மற்றும் இயல்பான மொழி செயலாக்கம் அடங்கும். ஆன்டிப் பழக்கவழக்கம் இருந்து கற்றுக்கொள்வதன் மூலமாக, பேச்சு மற்றும் படம் அங்கீகாரம் போன்ற இனொவேஷன்களை இயக்குகிறது. மனிதனியம், நோயளித்தல் மற்றும் நிதி மோசடித் தடுப்பு போன்ற பல்வேறு துறைகளுக்கு பயன்பாட்டுள்ளார். AI, தன்னியம் கொண்ட வாகனங்களை உள்ளடக்கிய போக்குவரத்து போன்ற துறைகள் மாறிவிட்டன. ஆனால், AI தனியுரிமை, தரவு பாதுகாப்பு மற்றும் பாகுபாடு போன்ற ஒழுங்கு சிக்கல்களை எழுப்புகிறது. கலைத்திறனை அகிலமாய் உருவாக்க முற்படுத்துதல் முனைவாதம் காரணமாகப் பொட்டியுகளை சேர்க்கிறது. பொறுப்பான AI மேம்பாட்டம் முக்கியமாக இருப்பதால், அதன் நன்மைகளை உறுதிப்படுத்துவதற்காக மேம்பாட்டாளர்கள், கொள்கைமுறையாளர்கள் மற்றும் சமூகத்துடன் ஒத்துழைக்க வேண்டும்.
OK

Fig. 5.3.

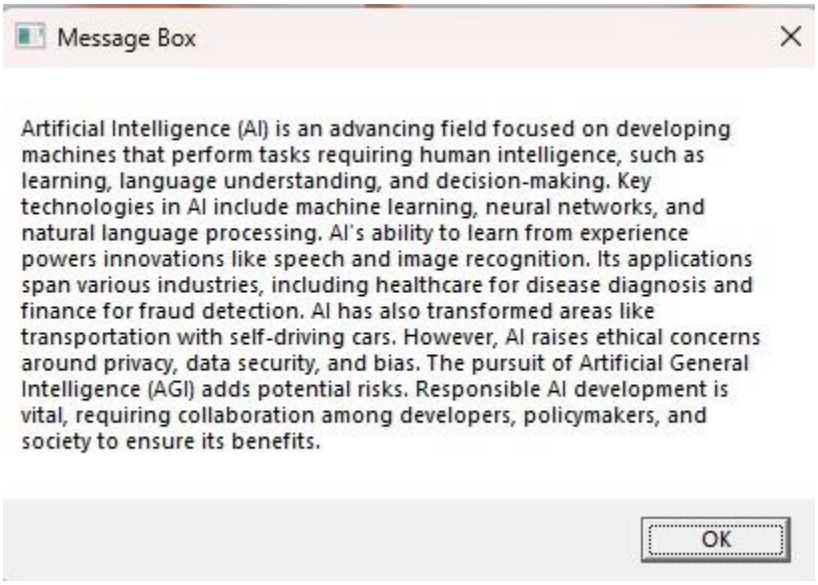
Message Box
Artificial Intelligence (AI) is an advancing field focused on developing machines that perform tasks requiring human intelligence, such as learning, language understanding, and decision-making. Key technologies in AI include machine learning, neural networks, and natural language processing. AI's ability to learn from experience powers innovations like speech and image recognition. Its applications span various industries, including healthcare for disease diagnosis and finance for fraud detection. AI has also transformed areas like transportation with self-driving cars. However, AI raises ethical concerns around privacy, data security, and bias. The pursuit of Artificial General Intelligence (AGI) adds potential risks. Responsible AI development is vital, requiring collaboration among developers, policymakers, and society to ensure its benefits.
OK

Fig. 5.4.

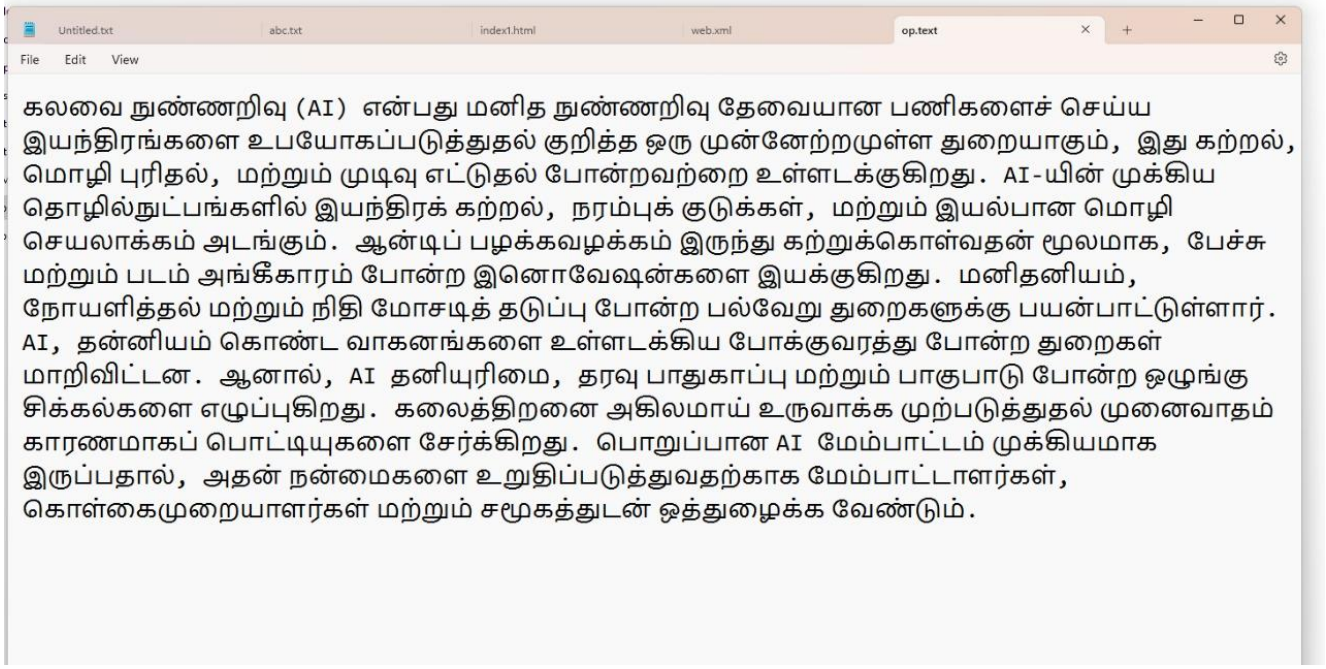
Untitled.txt
abc.txt
index1.html
web.xml
op.text
File Edit View
கலவை நுண்ணறிவு (AI) என்பது மனித நுண்ணறிவு தேவையான பணிகளைச் செய்ய இயந்திரங்களை உபயோகப்படுத்துதல் குறித்த ஒரு முன்னேற்றமுள்ள துறையாகும், இது கற்றல், மொழி புரிதல், மற்றும் முடிவு எட்டுதல் போன்றவற்றை உள்ளடக்குகிறது. AI-யின் முக்கிய தொழில்நுட்பங்களில் இயந்திரக் கற்றல், நரம்புக் குடுக்கள், மற்றும் இயல்பான மொழி செயலாக்கம் அடங்கும். ஆன்டிப் பழக்கவழக்கம் இருந்து கற்றுக்கொள்வதன் மூலமாக, பேச்சு மற்றும் படம் அங்கீகாரம் போன்ற இனொவேஷன்களை இயக்குகிறது. மனிதனியம், நோயளித்தல் மற்றும் நிதி மோசடித் தடுப்பு போன்ற பல்வேறு துறைகளுக்கு பயன்பாட்டுள்ளார். AI, தன்னியம் கொண்ட வாகனங்களை உள்ளடக்கிய போக்குவரத்து போன்ற துறைகள் மாறிவிட்டன. ஆனால், AI தனியுரிமை, தரவு பாதுகாப்பு மற்றும் பாகுபாடு போன்ற ஒழுங்கு சிக்கல்களை எழுப்புகிறது. கலைத்திறனை அகிலமாய் உருவாக்க முற்படுத்துதல் முனைவாதம் காரணமாகப் பொட்டியுகளை சேர்க்கிறது. பொறுப்பான AI மேம்பாட்டம் முக்கியமாக இருப்பதால், அதன் நன்மைகளை உறுதிப்படுத்துவதற்காக மேம்பாட்டாளர்கள், கொள்கைமுறையாளர்கள் மற்றும் சமூகத்துடன் ஒத்துழைக்க வேண்டும்.

Fig. 5.5

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1. CONCLUSION

This project successfully demonstrates the automation of extracting, summarizing, and sharing educational content from PDF files submitted by students. By leveraging UiPath's capabilities, such as the *Extract PDF* activity and integration with Generative AI, the workflow provides an efficient, reliable, and scalable solution for managing large volumes of student notes. The automated process eliminates the need for manual intervention, reducing errors and improving overall productivity.

The integration of advanced modules like text summarization, local storage, and automated email sharing ensures that summarized content is concise, accessible, and effectively distributed to the intended recipients. Additionally, the implementation of error-handling mechanisms and admin controls enhances the system's robustness, making it suitable for real-world applications in educational institutions or organizations dealing with extensive document management.

This project demonstrates how RPA (Robotic Process Automation) and AI can be combined to streamline complex workflows, highlighting their potential for improving operational efficiency and user satisfaction. Future enhancements could include extending the system's capabilities to handle additional file formats, incorporating advanced AI models for more accurate summarization, or introducing analytics to further improve decision-making processes.

.

6.2. FUTURE ENHANCEMENT

The automated system for extracting, summarizing, and sharing student notes can be further enhanced to meet evolving needs and provide greater functionality. Future enhancements could include extending the system to support additional file formats, such as Word documents, PowerPoint presentations, or scanned images using Optical Character Recognition (OCR) technology. This would broaden its usability and make it compatible with diverse educational content formats.

To improve the quality of text summarization, advanced Natural Language Processing (NLP) models, such as BERT, GPT-4, or other domain-specific AI models, could be integrated. These models would enable more accurate and context-aware summarization, handling complex content more effectively. Additionally, incorporating multilingual capabilities would allow the system to process and summarize content in multiple languages, catering to a more diverse user base.

The reporting aspect can be enhanced by developing an interactive dashboard that provides real-time insights into the content processing pipeline. Administrators could use these dashboards to monitor submissions, track email delivery statuses, and analyze trends in the summarized content, ensuring better oversight and decision-making.

Integrating the system with Learning Management Systems (LMS) or cloud platforms like Google Drive or OneDrive would simplify content sharing and collaboration among students and educators. A mobile-friendly interface or dedicated mobile app could also be introduced, allowing users to submit PDFs, view summaries, and monitor email reports on the go.

To ensure scalability and data security, migrating the system to a cloud infrastructure would enable handling large volumes of data while adhering to privacy regulations like GDPR. Automating error-handling mechanisms and introducing AI-based quality checks for extracted and summarized content would enhance the system's robustness and reliability.