



# St. PETER'S ENGINEERING COLLEGE

UGC - AUTONOMOUS



Affiliated to JNTUH, Approved by AICTE, Accredited by NAAC with "A" Grade, NBA Programme Accredited (EEE, CSE, ECE)

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING AND INFORMATION TECHNOLOGY MINI PROJECT PRESENTATION

SECTION : IT-B

BATCH : 4

	Name	Roll No.
1.	PUNNA SAI GANESH	21BK1A1298
2.	VANJARAPU BHANU CHARAN	21BK1A12C3
3.	BELLAMKONDA RAGHU	22BK5A1203

Dr. K. Little Flower

PROJECT GUIDE

HOD CSE - IT

# **SPEECH RECOGNITION SYSTEM USING GOOGLE API**

Giving Wings To Thoughts

# PROBLEM STATEMENT

- Most industries need speech-to-text transcription to be accurate and efficient, but the solutions developed so far are error-prone, expensive, and painstaking. Advanced tools fail with complicated speech or noise in the background. There is thus a great need for a very robust yet quite straightforward system to work by quick and accurate speech-to-text conversion while allowing flexibility about where you can store and access the transcriptions. This project is intended to provide a dual-interface solution that would have the functionality of saving text into Notepad, in addition to an option for downloading that could be saved as a PDF file. It will address all the needs of the user, personal as well as professional applications, while working on making the application accessible for users with speech impairments.

Giving Wings To Thoughts

# PROPOSED SOLUTION

- The proposed system makes use of an easy to use speech-to-text tool provided by Google Cloud Speech API.
  - This tool gives two kinds of interfaces: the Notepad Interface, which translates speech into a Notepad file saved as `.txt` document, and the Web Interface, whereby users can download their transcription as PDF files for easier sharing and archiving.
  - The flexibility of the system allowed users to switch interfaces according to current needs. It provided high accuracy and reliability even in noisy environments with advanced speech recognition, audio pre-processing, and language modelling.
  - Designed with accessibility in mind, this system is supportive of people with speech impairments and thus fosters better cross-professional and private communication.
  - Combining the aspects of accurate transcription with versatile output options, it constitutes an efficient, powerful solution for vast numbers of users.
-

# EXISTING SOLUTIONS VS PROPOSED SOLUTION

There are several existing systems of speech recognition that are widely used today such as Google Speech-to-Text, Microsoft Azure Speech to Text, Apple Siri etc. The Google Cloud Speech-to-Text API offers reliable and scalable speech recognition capabilities, making it suitable for a wide range of applications, including transcription services, voice assistants, call centre analytics, voice command processing, and more.

## DRAWBACKS OF EXISTING SYSTEM

- Accuracy Limitations
  - Privacy and Data Security
  - Language Support and Model Limitations
  - Customization Complexity
-

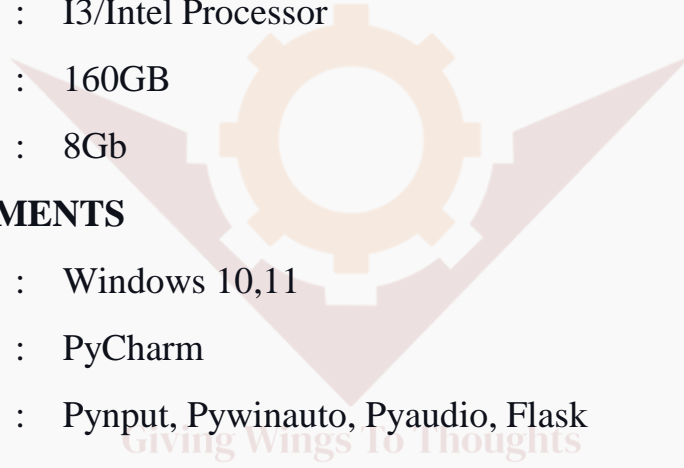
# REQUIREMENTS

## HARDWARE REQUIREMENTS

- Processor : I3/Intel Processor
- Hard Disk : 160GB
- RAM : 8Gb

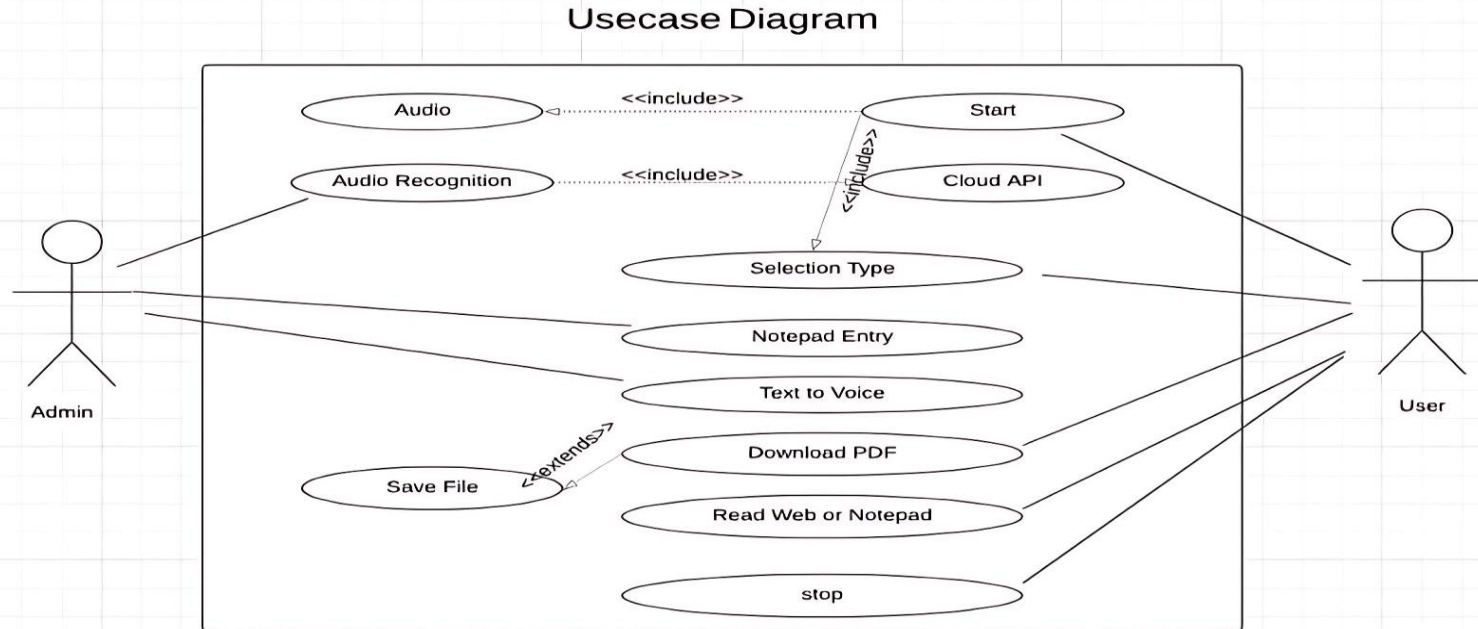
## SOFTWARE REQUIREMENTS

- Operating System : Windows 10,11
- IDE : PyCharm
- Libraries Used : Pynput, Pywinauto, Pyaudio, Flask
- Technology : PYTHON, HTML, CSS, JAVA SCRIPT



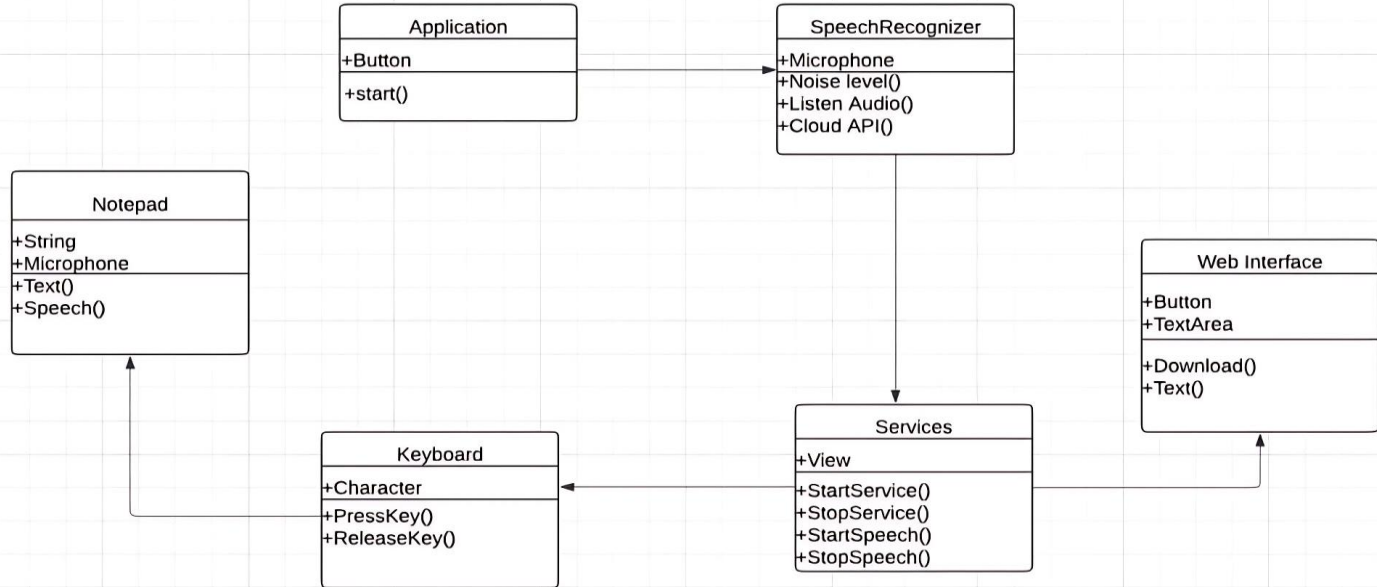
# UML DIAGRAMS

## USE-CASE DIAGRAM



# UML DIAGRAMS

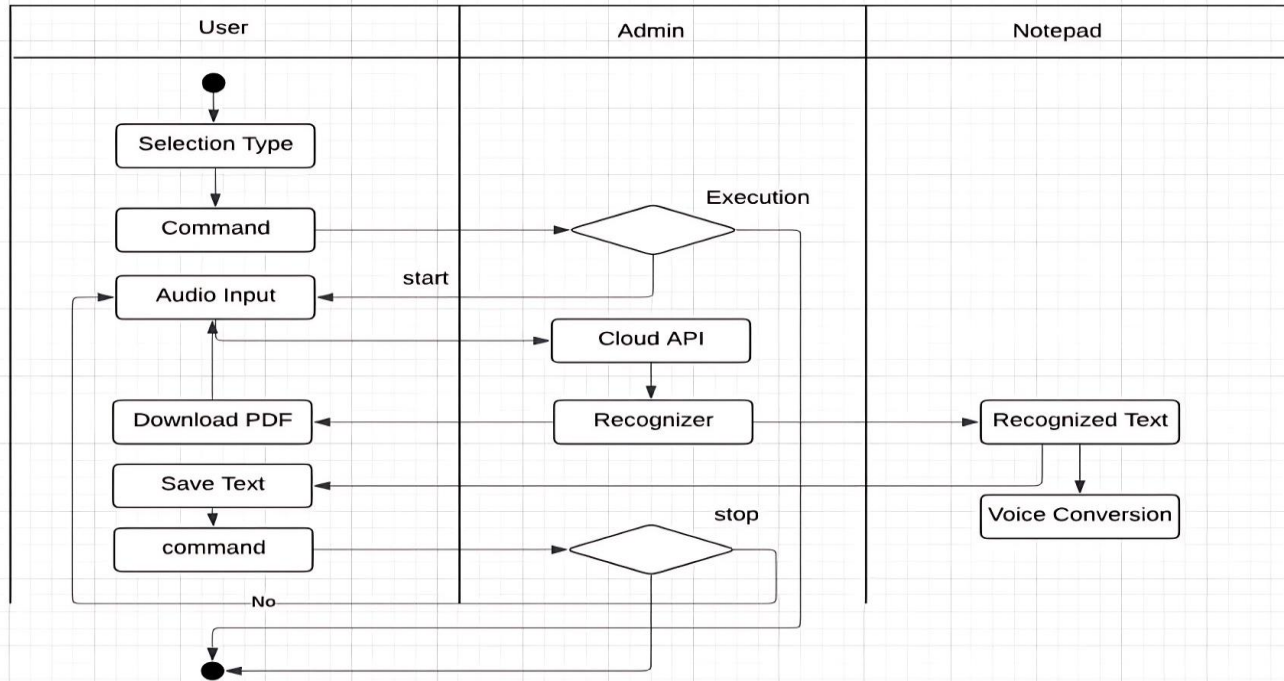
## CLASS DIAGRAM





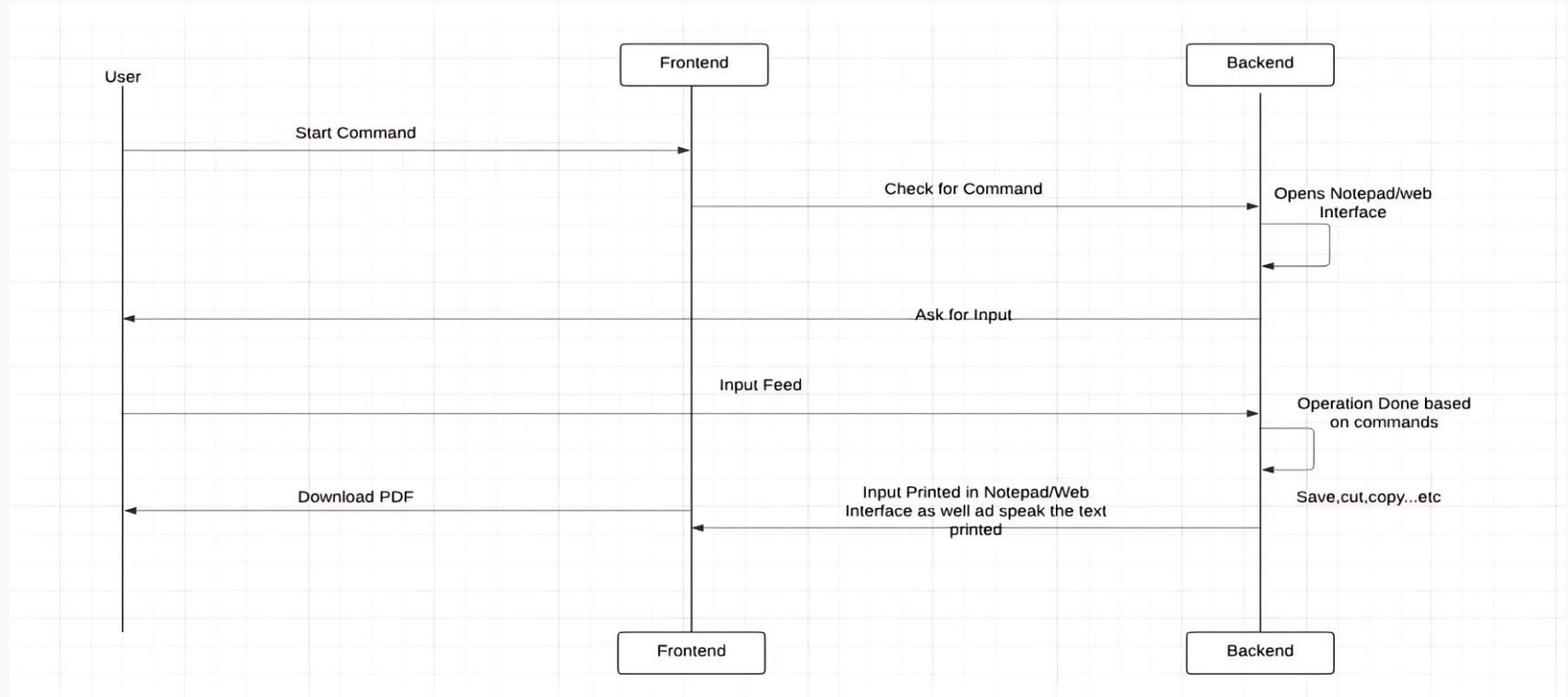
# UML DIAGRAMS

## ACTIVITY DIAGRAM



# UML DIAGRAMS

## SEQUENCE DIAGRAM



# PROJECT OUTPUT

## SPEECH TRANSCRIPTION

### Welcome to Speech-to-Text Application

#### Voice to Text Converter

hello this is Albert

Show to Text

#### Web Interface

The Web Interface offers a seamless, dynamic online experience, perfect for users looking to convert speech to text directly in a web environment. It's accessible from any device with internet connectivity and integrates all major functionalities of the app.

Explore Web Interface

#### Notepad Interface

The Notepad Interface is designed for users who prefer a direct text integration. Convert speech to text quickly and efficiently in a Notepad-style view, ideal for rapid note-taking and editing.

Explore Notepad Interface

#### AUDIO TO TEXT USING GOOGLE API



The Speech-to-Text application is designed to convert spoken words into text, providing an accessible solution. For developers, this application offers a comprehensive set of features and functionalities. The application is built using the Google Cloud Speech API, which provides a robust and scalable solution for converting speech to text. The application is designed to be easy to use and integrate into existing systems. It supports a wide range of audio formats and can be used to transcribe audio files, live audio streams, and more. The application is also designed to be secure and reliable, ensuring that your data is always protected. For more information, please visit our website or contact us directly.

Show to Text

# Voice to Text Converter



Voice to Text

Download as PDF

# Voice to Text Converter

this is Gilbert

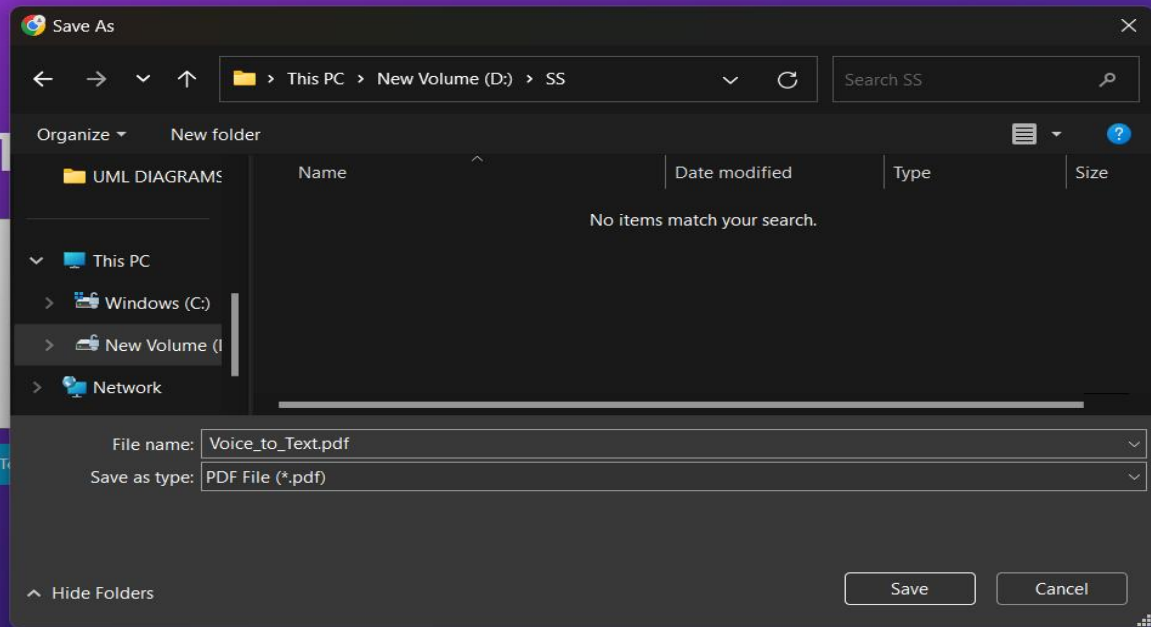
Voice to Text

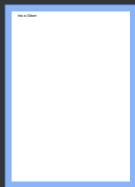
Download as PDF

# Voice to T

this is Gilbert

Voice to T

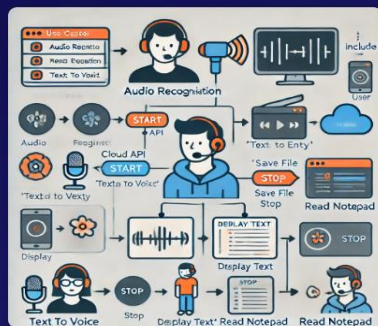




1

this is Gilbert

# AUDIO TO TEXT USING GOOGLE API



The Speech-to-Text application is designed to convert spoken words into written text, providing an accessible solution for individuals with impairments and enhancing communication across various industries. By leveraging the Google Cloud Speech API, the system captures audio input and transcribes it into text. It integrates with Notepad using the pynput library to simulate keyboard actions, the speech recognition library to process voice commands, and pywinauto to control the Notepad application. The process involves several steps, including audio input, preprocessing, feature extraction, acoustic modeling, and language modeling, enabling accurate and real-time transcription that can be saved as text in Notepad.

Start



# AUDIO TO TEXT USING GOOGLE AP

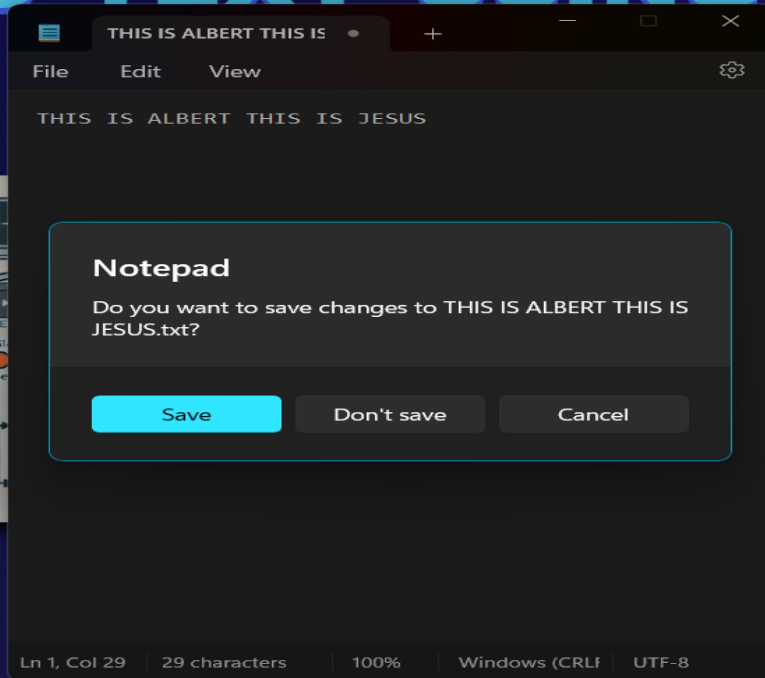
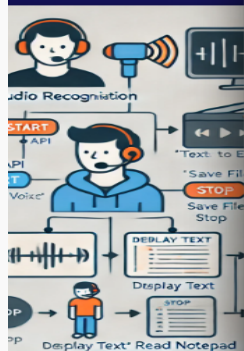


```
THIS IS ALBERT
```

Ln 1, Col 16 | 15 characters | 100% | Windows (CRLF) | UTF-8

n is designed to convert spoken words  
ding an accessible solution for  
and enhancing communication across  
ging the Google Cloud Speech API, the  
and transcribes it into text. It  
ng the pynput library to simulate  
recognition library to process voice  
control the Notepad application. The  
steps, including audio input,  
tion, acoustic modeling, and language  
and real-time transcription that can

# 0 TO TEXT USING GOOGLE API



n is designed to convert spoken words  
ding an accessible solution for  
and enhancing communication across  
ging the Google Cloud Speech API, the  
and transcribes it into text. It  
ng the pynput library to simulate  
recognition library to process voice  
control the Notepad application. The  
steps, including audio input,  
tion, acoustic modeling, and language  
and real-time transcription that can

# BENEFITS AND LIMITATIONS

- Efficient note-taking and transcription tool.
- Voice assistant with natural text-to-speech output.
- Enhanced accessibility for individuals with speech impairments.
- Accurate speech-to-text conversion for precise transcription.
- Versatile applications in note-taking, voice assistance.
- Seamless text-to-speech conversion for natural-sounding audio output.

Giving Wings To Thoughts

---

# CONCLUSION

The Google Cloud Speech API facilitates the speech recognition system through Notepad and Web interfaces. It employs high innovations such as HMM and Deep Learning to ensure perfect performance in all noisy environments as well as accommodate users with any form of speech impediment. Use of libraries such as Pynput and Pywinauto enhances the user experience of navigating applications on both the desktop and the web in a fluid and natural way. Further advancements could include the addition of language support, which will allow real-time transcription, include voice command functionality to enable hands-free control, noise-cancellation feature refinement for better accuracy in noisy conditions, mobile platforms integration, and cloud storage integration for the easy access and viewing of files from any device. Personalization of user customizations should be incorporated to fine-tune speech recognition while ensuring enhanced encryption and secure communication for improved data security.

---

A logo consisting of a light orange gear with eight teeth, centered behind a pair of light pink wings. The wings are spread outwards and upwards, with a darker pink outline. The entire logo is semi-transparent.

# THANK YOU

Giving Wings To Thoughts

---