USER MANUAL AND INSTALLER GUIDE

I.HEAR.YOU: A WEB-BASED IMAGE-TO-SPEECH CONVERTER APPLICATION USING OPTICAL CHARACTER RECOGNITION AND SPEECH SYNTHESIS

By

Mulya Fajar Ningsih Alwi

001202000101

TABLE OF CONTENTS

INTRODUCTION

Welcome to the User Manual and Installer Guide for the Web-Based Image-to-Speech Converter Application. This user manual will guide you through the installation process and provide instructions on how to use the application and enjoy all the features offered in the web application. With this web application, you can effortlessly convert an image containing text into audible speech, opening up a world of possibilities for accessibility and convenience.

The developed Web-Based Image-to-Speech Converter Application implements Artificial Intelligence (AI) technology such as Optical Character Recognition (OCR) and Speech Synthesis by using its two main components, such as Natural Language Processing (NLP) and Digital Signal Processing (DSP). The web application allows you to convert an image containing text into audible speech and is equipped with a feature that can translate the detected text into various languages supported by the application along with its audio file.

The web application's name is I.HEAR.YOU (Image-to-Speech Converter). The web application was developed using the Python programming language with the help of the Flask framework to build a web application in a structured way and with the help of several Python libraries such as OpenCV, Tesseract, pytesseract, gTTS, pygame, googletrans, Flask-WTF, Flask-Dropzone, Flask-Session, etc.

SYSTEM REQUIREMENTS

Before installing the Web-Based Image-to-Speech Converter Application, it is important to ensure that your system meets the necessary requirements. This section outlines the prerequisites, including all the important software required to be installed. These dependencies are vital for the smooth functioning of the application, so please ensure that they are installed on your system. Below is a list of the several important pieces of software required for this web application, along with its function description.

No.	Software	Function Description
1.	Python Version 3.10	The main programming language used to develop the web application
2.	Visual Studio Code Version 1.77.0	A code editor used to develop all the script codes of the web application
3.	Windows PowerShell Version 5.1.22621.963	A terminal used to execute and run the application on a local server
4.	Chrome Version 112.0.5615.138 (Official Build) (64-bit)	A web browser used to run the web application
5.	Flask Version 2.2.3	A Python framework used to develop the web application in a structured way
6.	OpenCV Version 4.5.5.62	A Python library used to read and change the image color for better analyzing
7.	Tesseract Version v5.3.1.20230401	An Optical Character Recognition (OCR) engine used to recognize characters in the uploaded image
8.	pytesseract Version 0.3.8	A Python library used to recognize the uploaded image's content or text
9.	gTTS Version 2.2.3	A Python library used to convert the detected text into an audio file and speak it out in a human language
10.	googletrans Version 3.1.0a0	A Python library used to translate the detected text into various languages supported

INSTALLATION

Setting up the Web-Based Image-to-Speech Converter Application is a straightforward process. This section provides step-by-step instructions on how to install the required components. From installing the supporting software, including the Python programming language, Tesseract OCR engine, and Visual Studio Code, to activating the virtual environment and installing the necessary libraries that are listed in the requirements.txt file, you will have the application up and running in no time. Follow the instructions closely to ensure a successful installation.

Step 1: Download the Application and Supporting Software .zip folders and extract them to a folder on your system.

Step 2: Install all the supporting software provided in the Supporting Software folder (if not already installed), including the Python programming language, Tesseract OCR engine, and Visual Studio Code. For the installation steps of each required software, you can search and use the installation steps available on the internet.

Step 3: Open a command prompt or terminal and navigate to the web application folder that contains the existing environment and requirements.txt file, as shown below.

```
PS C:\Users\Lenovo Flex> cd Downloads\APPLICATION\APPLICATION
PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION>
```

Step 4: Activate the virtual environment. Depending on the operating system, the activation command can vary.

```
For Windows:
bash                                                    Copy code

  env\Scripts\activate

For macOS and Linux:
bash                                                    Copy code

  source env/bin/activate
```

For this web application, the activation command used the command for the Windows operating system, as shown below.

```
PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION> . venv/Scripts/activate
```

Step 5: Once the virtual environment is activated, you should see the environment name appear in your command prompt or terminal (the green word), as shown below.

```
(venv) PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION>
```

Step 6: Install the required dependencies using the command below.

```
(venv) PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION> pip install -r requirements.txt
```

This command will read the requirements.txt file (containing required libraries) and install all the listed dependencies or required libraries into the activated environment.

Step 7: Once the installation is complete, you can proceed to start the Flask development server (the steps will be explained in the next section) and use the Web-Based Image-to-Speech Converter Application as described in the user manual.

APPLICATION SETUP

Once you have installed all the necessary components, it is time to set up the Web-Based Image-to-Speech Converter Application. This section guides you through the process of preparing the web application for usage by starting the Flask development server and you will have the web application ready to convert your image into audible speech in just a few simple steps.

Step 1: Open a command prompt or terminal and navigate to the web application folder, as shown below.

```
PS C:\Users\Lenovo Flex> cd Downloads\APPLICATION\APPLICATION
PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION>
```

Step 2: Activate the virtual environment, as shown below.

```
PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION> . venv/Scripts/activate
(venv) PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION>
```

Step 3: Start the Flask development server by running the command below.

```
(venv) PS C:\Users\Lenovo Flex\Downloads\APPLICATION\APPLICATION> python3 run.py
```
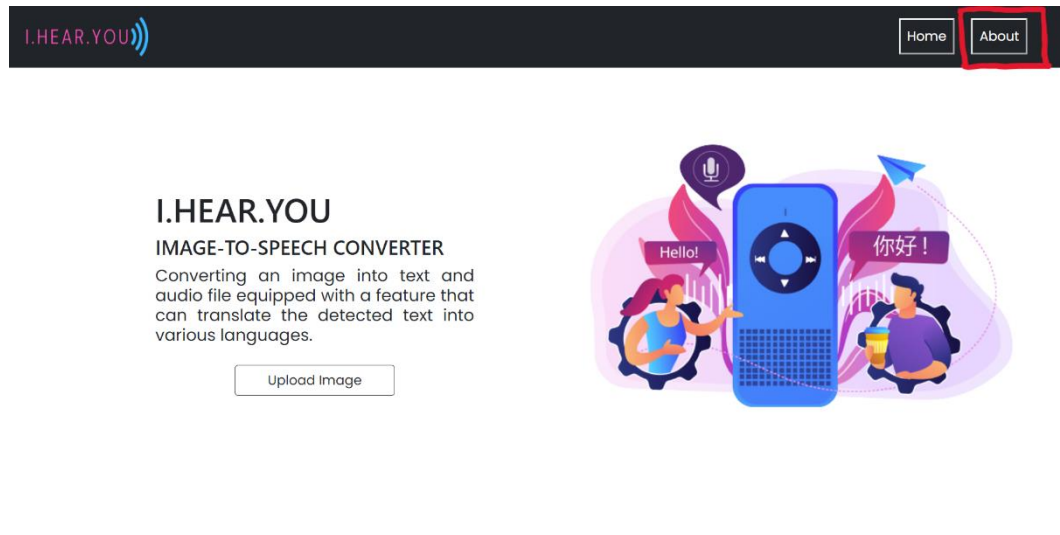
APPLICATION USAGE

With the Web-Based Image-to-Speech Converter Application fully set up, it is time to explore its functionalities. This section explains how to utilize the web application effectively. From accessing the web application interface to uploading an image file, hearing the converted speech of the uploaded image, and translating the detected text into another language supported, you will gain a clear understanding of how to use the web application for your image-to-speech needs in this section.

Step 1: Open a web browser and navigate to the Flask development server address displayed in your command prompt or terminal. Alternatively, you can press the CTRL key and click on the Flask development server link in your command prompt or terminal, which will automatically redirect you to your web browser and launch the web application. The Flask development server will be displayed as shown below.

**Step 2:** As the web application runs, you will be directed to the Home Page interface. From the Home Page, you can access the About Page by clicking the 'About' button in the navigation bar, as shown below.


Step 3: The About Page will appear as shown below. To return to the Home Page, simply click the 'Home' button in the navigation bar.

Step 4: To utilize the main features of the web application, start by accessing the Upload Page. Click the 'Upload Image' button available on the Home Page, as shown below.
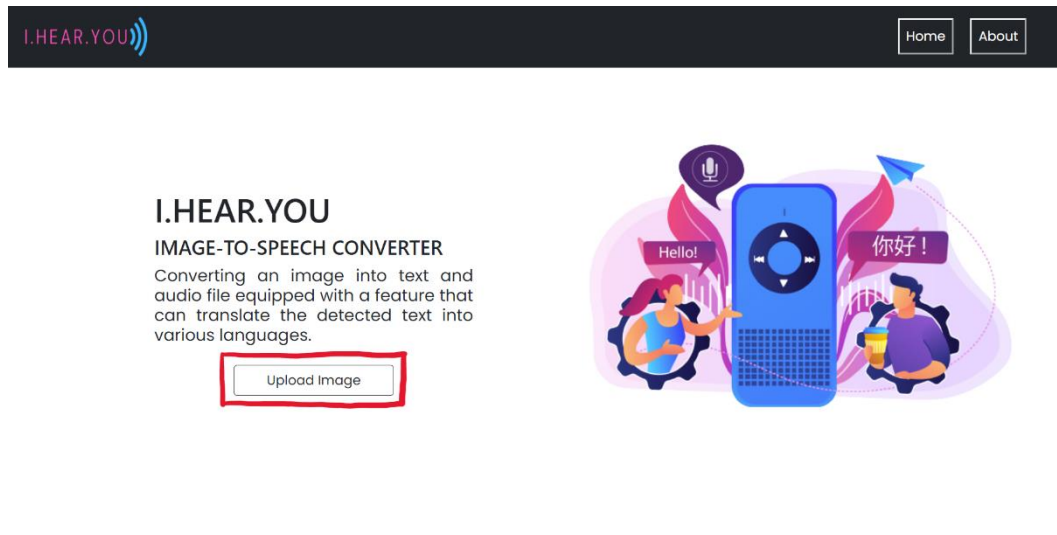

Step 5: On the Upload Page, you can upload your image file for processing. Use the provided Dropbox to upload or drop your image file by clicking Dropbox, as shown below. Select the desired image file from your computer (ensure it is in JPG, JPEG, or PNG format). Please note that only one image file can be uploaded at a time.
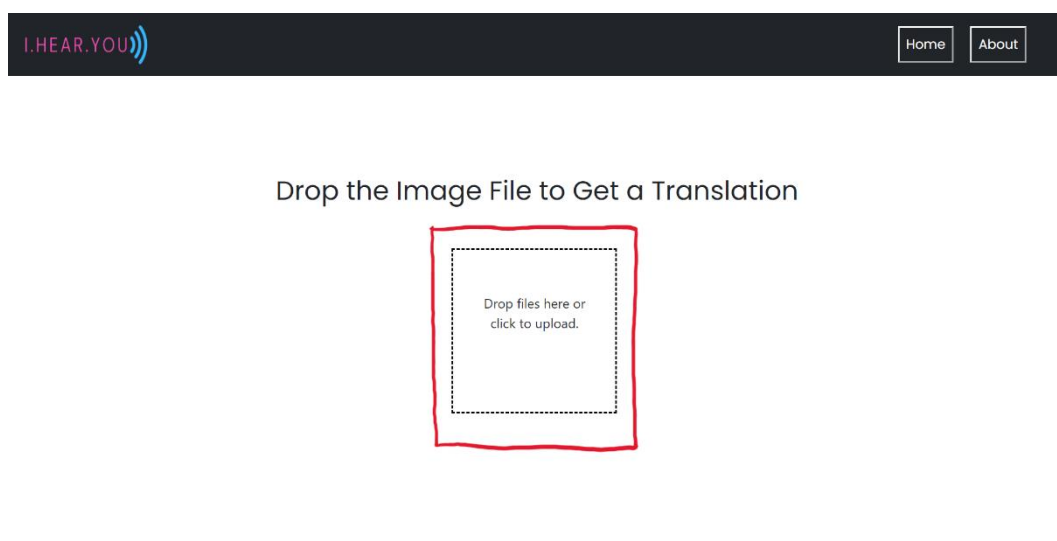
Step 6: If the image upload and decoding (converting the uploaded image into text) are successful, the application will automatically redirect you to the Image Decoded Page, and the application will autoplay the audio of the detected text in the uploaded image. The Image Decoded Page will be presented as shown below, providing detailed information about the uploaded image, including the display of the uploaded image, the language of the detected text, and the detected text itself.


Step 7: The web application also includes a translation feature for the detected text. To use this feature, select the desired translation language from the provided selection box, as shown below.
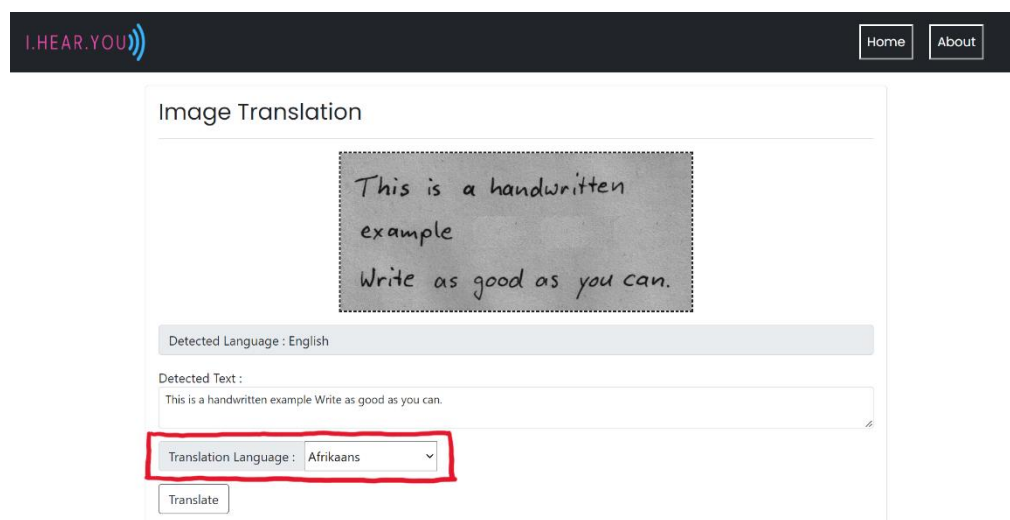
Step 8: After selecting the desired translation language, click the 'Translate' button, as shown below. The application will process your input and give you the desired translation result.
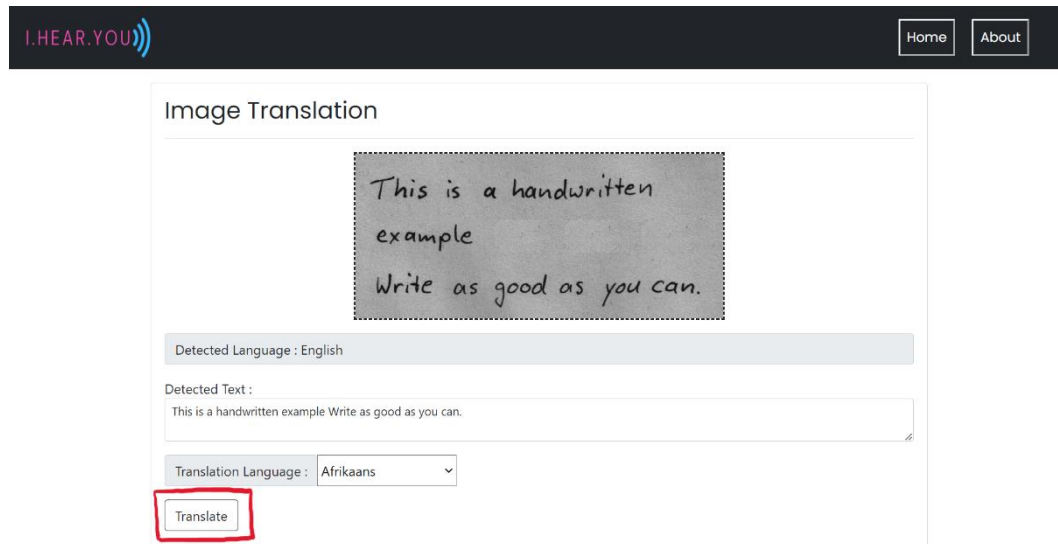

Step 9: If the translation process is successful, the application will display the translated text and autoplay the audio of the translated text with an audio playback option, as shown below.
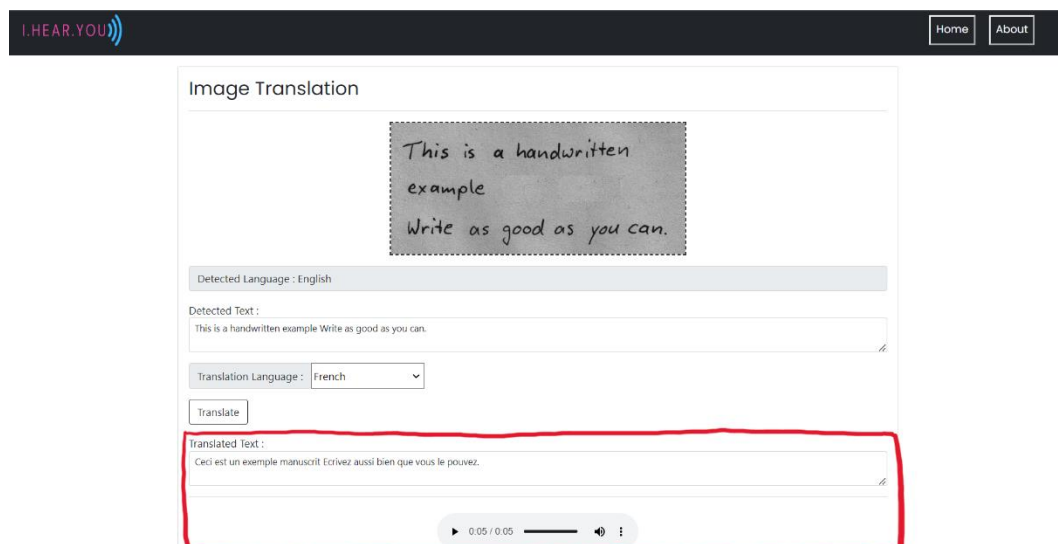
Step 10: To download the audio file of the translated text, click the three vertical dots in the audio playback option, and then select the download button, as shown below. This will download the audio file of the translated text and save it to your local folder.
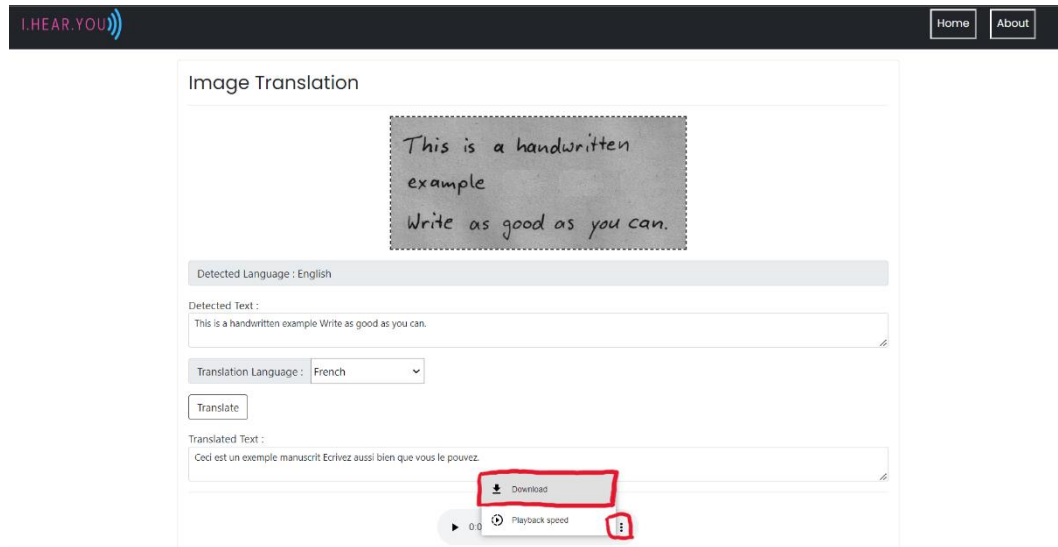
Step 11: To retry the web application's features by uploading another image file, simply click the 'Home' button in the navigation bar, as shown below. This will redirect you to the Home Page, where you can access the 'Upload Image' button and repeat the process from step 4.
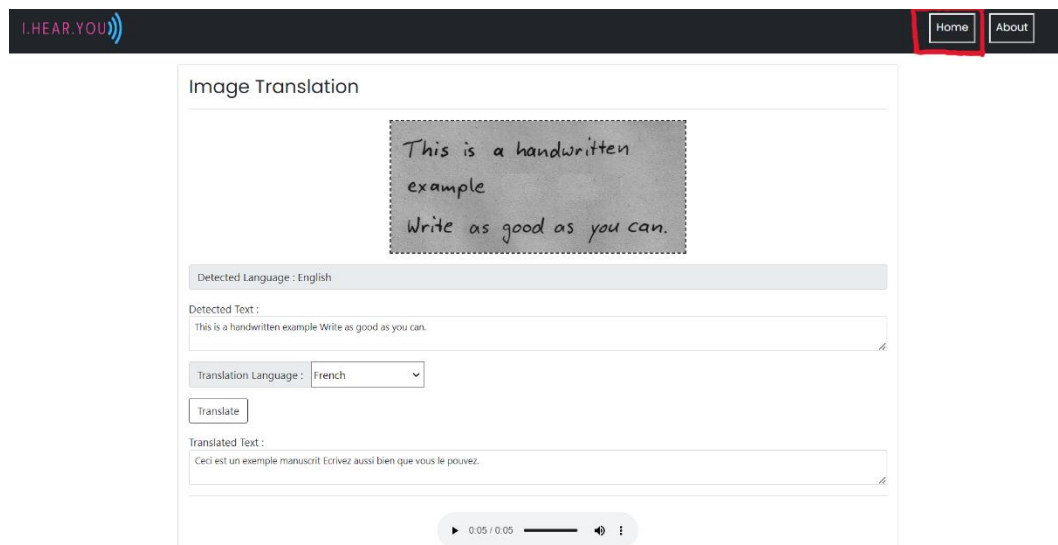
TROUBLESHOOTING

In case you encounter any difficulties while installing or using the web application, this section provides troubleshooting tips. From resolving installation errors to addressing issues with image conversion, you will find helpful suggestions to overcome common obstacles. Refer to this section if you need assistance in troubleshooting any problems you may encounter. If you encounter any issues while using the Web-Based Image-to-Speech Converter Application, refer to the following troubleshooting tips to resolve common problems:

a) Dependency Issues:

- If there are errors related to missing or incompatible dependencies during installation, make sure you have the required version of Python installed.

- Check if any specific system dependencies are needed and ensure they are installed properly.

- If the problem persists, refer to the documentation or seek support from the application's developer or community.

b) Configuration Errors:

- Double-check any configuration settings or environment variables required for the application to run properly.

- Verify that the necessary paths and file permissions are correctly set up.

c) Image Not Uploading:

- Ensure that the image file you are trying to upload is in a supported format (JPG, JPEG, or PNG).

- Check the file size of the image. If it is too large, try compressing the image or selecting a smaller image to upload.

- Verify that the image file is not corrupted. Try opening the image with an image viewer to ensure it is accessible.

d) Decoding Error:

- If the application fails to decode the uploaded image, ensure that the image contains clear and legible text.

- Check the language of the text in the image. The application may face difficulties decoding text in uncommon languages or with complex fonts.

- Consider adjusting the image quality or resolution. Very low-quality images may result in inaccurate or incomplete text recognition.

e) Translation Issues:

- Double-check that you have selected the desired translation language correctly from the available options.

- If the translation does not provide the expected results, try simplifying the text or breaking it down into shorter phrases.

- Keep in mind that machine translation may not always provide perfect accuracy. Complex or context-specific phrases may require manual translation.

f) Audio Playback Problems:

- Ensure that your device's audio is not muted and that the volume is set to an audible level.

- Check that your web browser supports audio playback and that you have the latest version installed.

- If the audio is not playing or is distorted, try refreshing the page or restarting your web browser.

g) Server Connection Issues:

- If you experience difficulties accessing the application or encounter server errors, check your internet connection and try reloading the page.

- Verify that the Flask development server is running correctly by checking the command prompt or terminal for any error messages.

- If the issue persists, contact the application's support team for further assistance.

h) General Troubleshooting:

- If you encounter any other issues not mentioned above, try restarting the application, clearing your browser cache, or using a different web browser.

- Check for any error messages or logs at the command prompt or terminal for further information on the problem.

CONCLUSION

Congratulations on successfully installing and using the Web-Based Image-to-Speech Converter Application! This concluding section acknowledges your achievement and encourages you to explore the web application further. Enjoy the convenience and accessibility it offers as you convert an image containing text into audible speech effortlessly.