AI Image Description & Translation Project

# 1. Introduction

This document provides detailed steps and information about the AI Image Description & Translation project. The project uses a Streamlit app to generate descriptions for uploaded images based on user prompts, leveraging Google Generative AI. Additionally, the app can translate the generated descriptions into various languages and convert them into audio using text-to-speech technology.

# 2. Project Description

The AI Image Description & Translation project aims to create an interactive web application where users can upload an image and provide a prompt. The app generates a description of the image based on the prompt using Google Generative AI. The description can be translated into multiple languages and converted into audio with a choice of male or female voices. This project is useful for creating multilingual content, enhancing accessibility, and exploring AI-driven content generation.

# 3. Prerequisites

Before starting, ensure that your system meets the following requirements:  
- Python 3.x  
- Streamlit  
- PIL (Pillow)  
- Google Generative AI (gemini-1.5-flash-001)  
- pyttsx3  
- googletrans

# 4. Setup Instructions

## Cloning the Repository

1. Clone the repository using the following command:

```bash  
git clone https://github.com/Kolli1999/AI-Image-Description-Translation.git  
```  
2. Navigate to the project directory:  
```bash  
cd AI-Image-Description-Translation  
```

## Installing Dependencies

3. Install the required dependencies:

```bash  
pip install -r requirements.txt  
```

## API Key Setup

4. Obtain an API key from Google Generative AI. Replace the placeholder API key in the script with your own API key:  
```python  
api\_key = 'your-api-key-here'  
```

# 5. Running the Project

1. Run the Streamlit app using the following command:  
```bash  
streamlit run app.py  
```  
2. The app will open in your web browser.  
3. Upload an image, enter a prompt, and the app will generate a description.

# 6. Audio and Translation

1. Choose a voice (Male/Female) for the text-to-speech conversion from the dropdown menu.  
2. Select a language to translate the generated description from the dropdown menu.  
3. The translated text will be displayed, and you can listen to the audio output.

# 7. Customization

## Modifying Prompts

You can customize the prompts to generate different types of descriptions for the images. Experiment with different prompts to see how the model responds.

## Changing Languages

The app supports multiple languages for translation. You can easily add or remove languages by modifying the language list in the script.

## Adjusting Text-to-Speech Settings

The text-to-speech settings, such as voice rate and volume, can be adjusted in the script. These settings allow you to fine-tune the audio output to your preference.

# 8. Conclusion

The AI Image Description & Translation project is a versatile tool for generating multilingual descriptions and audio outputs based on user-provided prompts. The app's flexibility allows for easy customization, making it suitable for a wide range of applications, including content creation, accessibility enhancements, and AI experimentation.

# 9. Appendix

For additional information and resources, refer to the following links:  
- [Google Generative AI Documentation](https://cloud.google.com/ai-platform/docs)  
- [Streamlit Documentation](https://docs.streamlit.io/)  
- [pyttsx3 Documentation](https://pyttsx3.readthedocs.io/en/latest/)  
- [Googletrans Documentation](https://py-googletrans.readthedocs.io/en/latest/)