# Multi-Modal Translator with Text and Speech Translation in NLP

## Project Overview and Purpose

This NLP project provides a user-friendly, multi-modal translation application that translates English input (text or audio) into Spanish. Users can input text directly, upload a text file, or upload an audio file. The application provides options for output in either text or audio formats, allowing flexibility in translation formats and catering to diverse user needs.  
  
The application is built using a combination of the MarianMT Model for machine translation, Google Text-to-Speech (gTTS) for audio output, and SpeechRecognition for processing audio inputs. The entire process is managed in a streamlined web-based interface, implemented with Flask and enhanced by front-end HTML and JavaScript for improved user experience.

## How It Works

### 1. Input Handling

Users select the input type (text, audio, or file). If the input is text, the user can enter English text directly. If the input is audio, the user uploads an audio file in English, which is processed using SpeechRecognition to extract text. If the input is a text file, users can upload a .txt file containing English text. This text is read and prepared for translation.

### 2. Translation Process

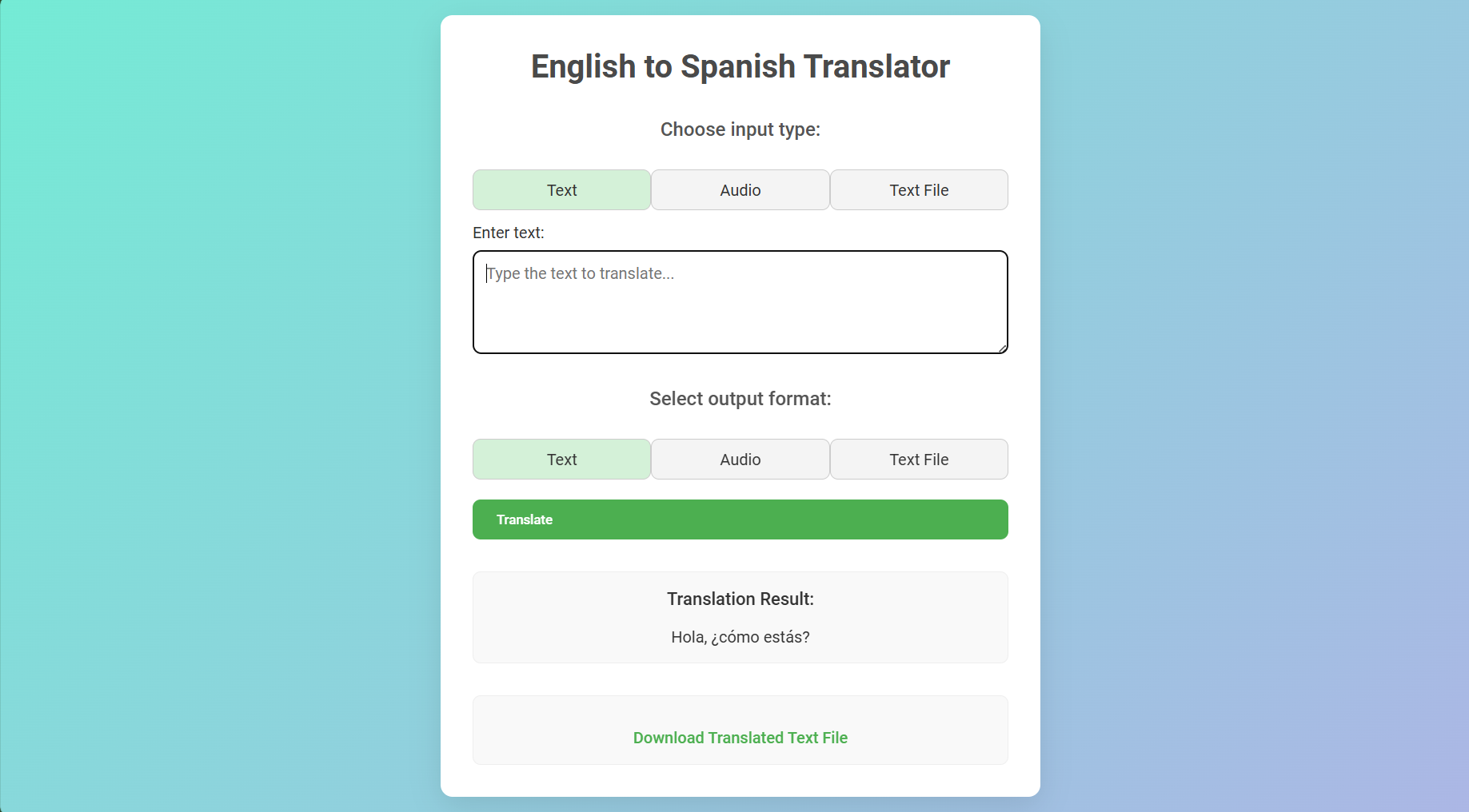
The application leverages the MarianMTModel from the transformers library, specifically the Helsinki-NLP/opus-mt-en-es model, which is designed for high-quality English-to-Spanish translations. To handle long sentences, the translation function segments text into manageable sentences and processes each one individually, preventing truncation and ensuring full translation output.

### 3. Output Options

Text Output: Translated text is displayed directly in the web interface.  
Audio Output: The application uses the gTTS library to convert translated text into spoken Spanish audio, which users can download or listen to.  
File Output: If the user selects file output, the translation is saved as a .txt file, which can be downloaded.

## User Interface

The web interface is built with HTML and CSS for a clear and intuitive experience. Users can toggle between input and output options, upload files, and view/download their results directly. JavaScript handles the dynamic switching of input types based on user selection, ensuring a seamless experience.



The above picture shows the front end

## Achievements of the Program

This program demonstrates an effective use of NLP techniques and model-based translation in a multi-modal format. By providing options for both text and audio input/output, the project highlights the versatility of translation applications in real-world scenarios. Furthermore, the use of Flask allows the application to be hosted and accessed through a web server, making it accessible to users without technical expertise in NLP.

## Conclusion

The Multi-Modal Translator with Text and Speech Translation in NLP successfully combines translation, text-to-speech, and speech recognition technologies into a unified platform. It showcases the potential of modern NLP models and provides an adaptable solution for multi-format translation tasks. Through this project, users can experience seamless translation, whether interacting via text or audio, demonstrating the practical applications of NLP in communication tools.