

AI-Powered Multi-Language Translator - Project Report

Introduction

This project is a web-based language translation app that enables users to translate text between various languages using pre-trained NLP models from Hugging Face (Helsinki-NLP/opus-mt). The frontend is built with Streamlit, providing a simple and interactive user experience.

Objectives

- Build a multilingual translation tool.
- Use Hugging Face's MarianMTModel for neural machine translation.
- Deploy a clean and interactive UI using Streamlit.

Tools and Technologies

- Programming Language: Python
- Libraries:
 - streamlit: for building the UI
 - transformers: for loading NLP models
- Model Source: Hugging Face (Helsinki-NLP/opus-mt) models

Supported Languages

English (en), Hindi (hi), French (fr), Spanish (es), German (de), Italian (it), Russian (ru), Japanese (ja), Chinese (zh)

Core Functionalities

- Model Loading: Dynamically loads MarianMT models for selected source-target language pair.
- Translation: Uses tokenization and model inference to generate the translated text.
- Caching: Implements `@st.cache_resource` to speed up model loading.
- User Interface: Allows users to input text and select languages via dropdowns.

Code Summary

AI-Powered Multi-Language Translator - Project Report

Key Components:

- `load_model()`: Loads tokenizer and model from Hugging Face.
- `translate_text()`: Tokenizes input and decodes output from model.
- `st.text_area`, `st.selectbox`, and `st.button`: Core Streamlit components to capture user input and display results.

How to Run

1. Create virtual environment:

```
python -m venv venv
```

```
source venv/bin/activate (Windows: venv\Scripts\activate)
```

2. Install dependencies:

```
pip install streamlit transformers
```

3. Save the code as `multi_lang_translator.py`

4. Run the app:

```
streamlit run multi_lang_translator.py
```

Limitations

- Model loading might be slow for first-time language pairs.
- Limited to pre-defined languages available in opus-mt models.
- Internet connection required for first-time model downloads.

Future Improvements

- Add speech-to-text input and text-to-speech output.
- Expand supported languages dynamically via API.
- Save translation history for logged-in users.