



CapTransVoice



02

Project Overview

Workflow Pipeline:

1. Image Input: Users upload an image.
2. Caption Generation: AI model generates descriptive captions.
3. Translation: Captions are translated into desired languages using the LAMA model.
4. Speech Generation: Translated captions are converted to audio using the gTTS library.



Image Captioning

Utilize a pre-trained image captioning
model (specially BLIP processor)

Translation with LAMA Model

Using a robust language model
designed for accurate translations.
(LAMA Model)

- **Input:** Generated caption in the source language (English)
- **Output:** Translated caption in the target language(Arabic)

Text-to-Speech with gTTS Library

Google Text-to-Speech, a Python library interfacing with Google Translate's text-to-speech API.



Result and Impact

- -Transforms visual content into audio descriptions, aiding visually impaired individuals in understanding image content.

- Social Media Enhancement:

Enables content creators to generate accessible and multilingual content effortlessly.

- Business and Customer Service

- Global Reach: & Empowers businesses to cater to international markets by providing localized audio descriptions of their products or services.

OP



Thank You!