Image Captioning and Text to Image

Name: Yazeed Abdullah Al nashib

Date: 1\10\2024

Objectives

- **Automated Image Captioning**: Automatically generate a caption for a given image.

- **Text-to-Image Generation**: Create images from user-provided text descriptions or suggestions.

- **Speech Output**: Convert the generated captions to speech.

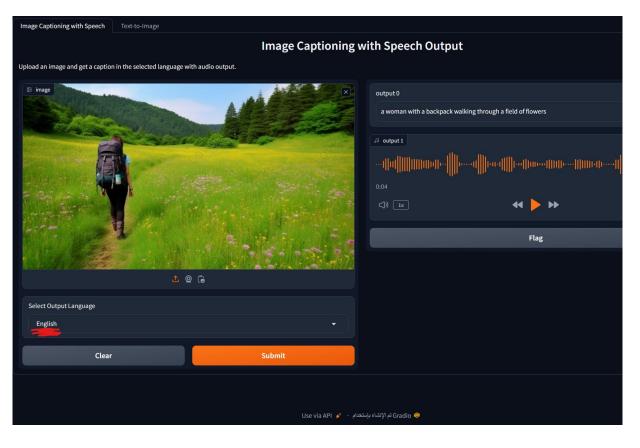- **Multilingual Support**: Provide caption and speech output in multiple languages.
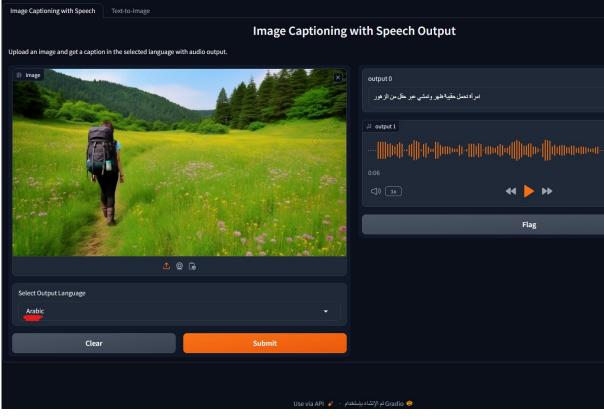
Pipeline Implementation

- **Step 1: Load Models**:
- BLIP (Bootstrapping Language-Image Pre-training) for image captioning.
- Stable Diffusion for text-to-image generation.

- **Step 2: Image Captioning**:
- User uploads an image.
- The system generates a caption using BLIP.
- The caption is translated and converted to speech using GoogleTranslator and gTTS.

- **Step 3: Text-to-Image**:
- User inputs a text description or selects a suggestion.
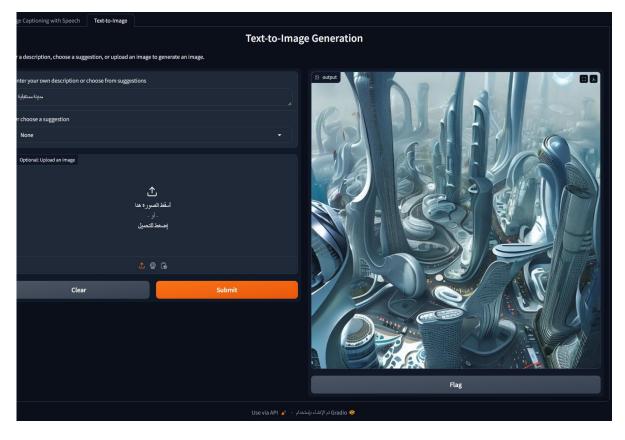- Stable Diffusion generates an image based on the input description.

Results

- **Image Captioning Example**:
- Input: Image of a sunset.
- Output: "A beautiful sunset over the mountains." (with speech)

- **Text-to-Image Example**:
- Input: "A robot playing chess with a human."
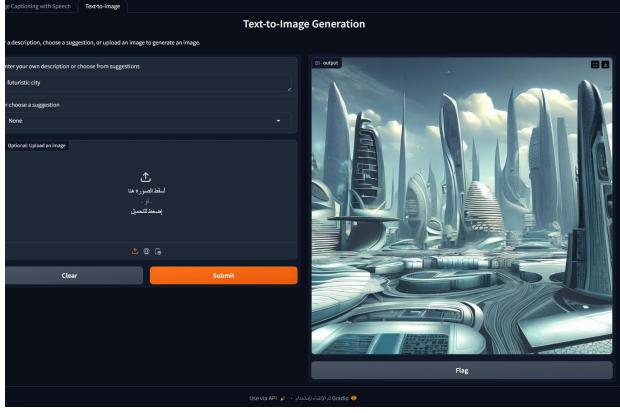- Output: Generated image of a robot playing chess.

Image Captioning Results (Arabic , English)

Text-to-Image Generation Results (Arabic , English)

Previous Work: Text Translation Project
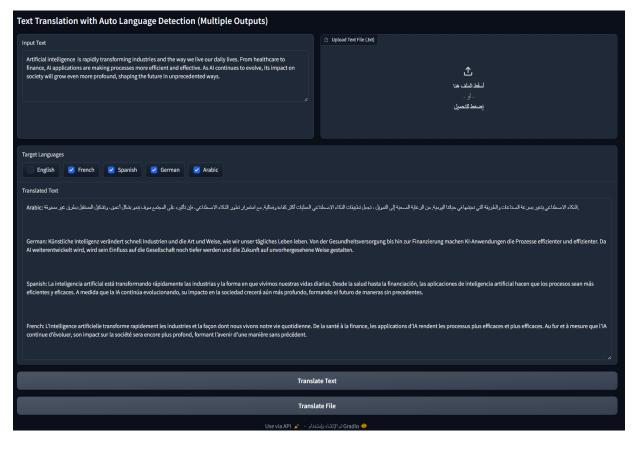
- **Project Overview:**

This project was focused on translating text using the NLLB (No Language Left Behind) model by Facebook AI, with automatic detection of the input language.
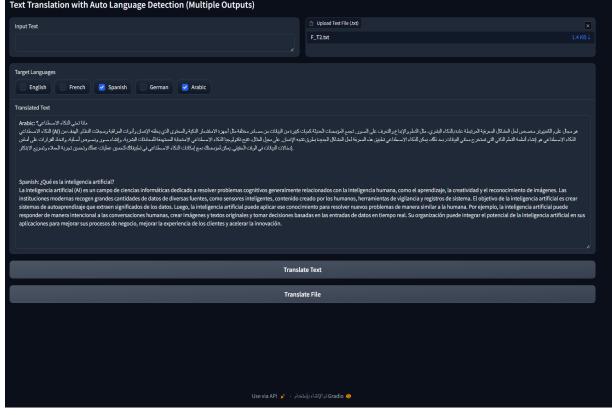
Features:

Automatic language detection for translating text into multiple languages (English, French, Spanish, German, Arabic)

Easy-to-use web interface using Gradio.

Text-Translation Results

Model Justification

- **Why BLIP for Captioning:**

- BLIP provides accurate and efficient image-to-text generation.

- **Why Stable Diffusion for Text-to-Image:**

- Stable Diffusion offers high-quality, detailed images based on text descriptions.

- **Why Google Translator:**

- Used for automatic language detection and translation of captions.

- Supports multiple languages to reach a broader audience.

GitHub & Hugging Face Links

- GitHub Repository:

- https://github.com/Yaz1-e/Image-Captioning-Text2Image

- Hugging Face Space:

- https://huggingface.co/spaces/Yaz1-e/Image-Captioning-Text2Image