IMAGE CAPTIONING AND TEXT TO IMAGE

BY: YAZEED ABDULLAH

OBJECTIVES

Enhance Visual Understanding through Automated Image Descriptions.

Generate High-Quality Images from Text Descriptions

Seamlessly Convert Text to Speech for Accessibility

SUPPORT MULTILINGUAL CAPTIONING AND SPEECH SYNTHESIS

Pipeline Implementation

Step 1: Load Models:

BLIP (Bootstrapping Language-Image Pre-training): Used for automatic image captioning

Stable Diffusion: Used for generating images from text descriptions

Google Translator: Used for translating captions into multiple languages

gTTS(Google Text-to-Speech): Converts the translated text into speech

Step 2: Image Captioning:

Upload an Image: The user uploads an image.

Generate Caption: The system generates a caption using the BLIP model.

Translate Caption: The caption is automatically translated using Google Translator.

Text-to-Speech: The translated caption is converted into speech using gTTS

Multilingual Support: Supports multiple languages including Arabic, English, French, etc

Step 3: Text-to-Image Generation

Text Input: The user provides a text description or selects from predefined suggestions

Image Input: The user can upload an image to enhance the output**.**

Generate Image: The Stable Diffusion model generates an image based on the text description and .optionally the uploaded image

Model Justification

Translation Models

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

CODE SNIPPETS

```
# Function to translate text to a target language
def translate_text(text, target_language):
    return GoogleTranslator(source='auto', target=target_language).translate(text)
```

The translate_textfunction uses the GoogleTranslator library to automatically translate text from the detected source language (source='auto’) to the specified target language.

Image Captioning Function

```python
# Function for generating captions from images, with text-to-speech
def generate_caption(image, target_language):
    inputs = processor(image, return_tensors="pt").to("cuda" if torch.cuda.is_available() else "cpu")
    out = model.generate(**inputs)

    caption = processor.decode(out[0], skip_special_tokens=True)

    translated_caption = translate_text(caption, target_language)

    tts = gTTS(text=translated_caption, lang=target_language)
    tts.save("output.mp3")

    return translated_caption, "output.mp3"
```

The **generate_captionfunction** processes an input image to generate a caption using the BLIP .model. It then translates the caption to the selected language using the **translate_textfunction** .Finally, it converts the translated caption into speech using Google TTS

Text-to-Image Function

```python
# Function for generating images from text descriptions
def generate_image_from_text(description, image_input=None):
    translated_description = translate_text(description, "en")

    if image_input is not None:
        inputs = processor(image_input, return_tensors="pt").to("cuda" if torch.cuda.is_available() else "cpu")
        out = model.generate(**inputs)
        image_caption = processor.decode(out[0], skip_special_tokens=True)
        translated_description += " " + image_caption  # Append the image caption to the text

    image = pipe(translated_description, num_inference_steps=70, guidance_scale=6.5).images[0]

    return image
```

The **generate_image_from_text function** translates a text description into English and generates an image using the Stable Diffusion model. If an image is provided, the function generates a caption for the image and appends it to the translated description before generating the final output image. The image is returned as the result.

Image Captioning Results

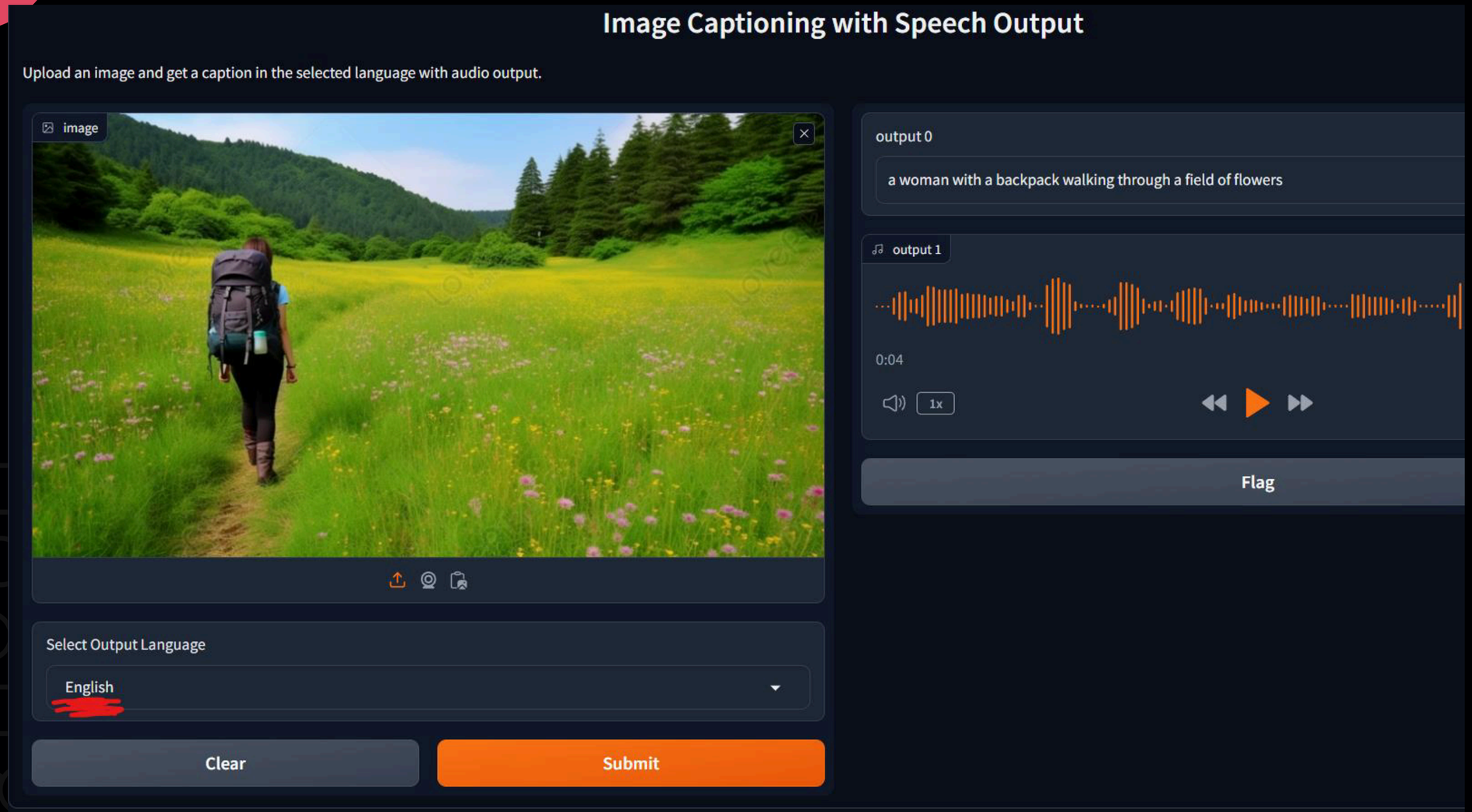
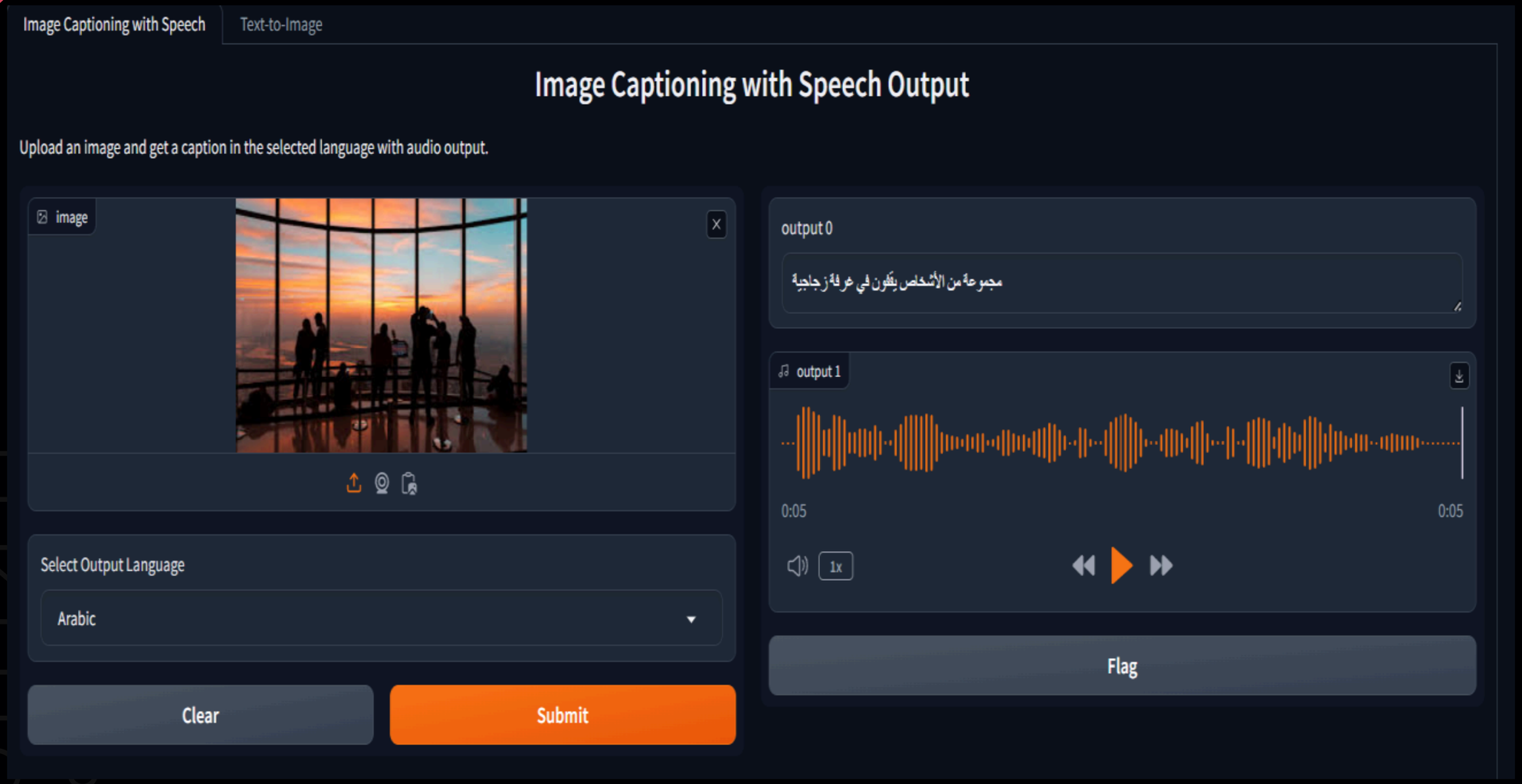
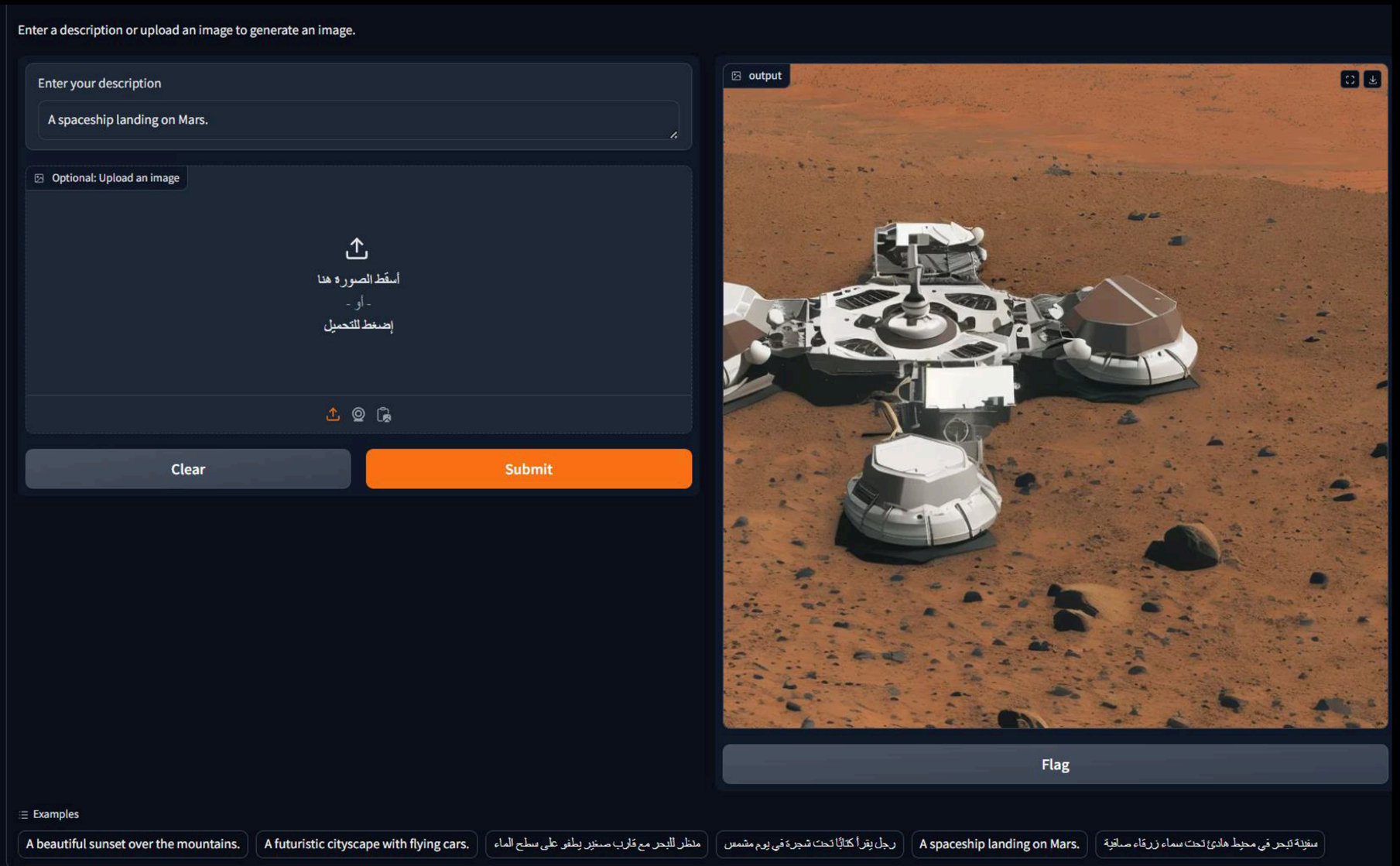


Image Captioning Results



Text-to-Image Generation Results



Previous Work: Text Translation Project

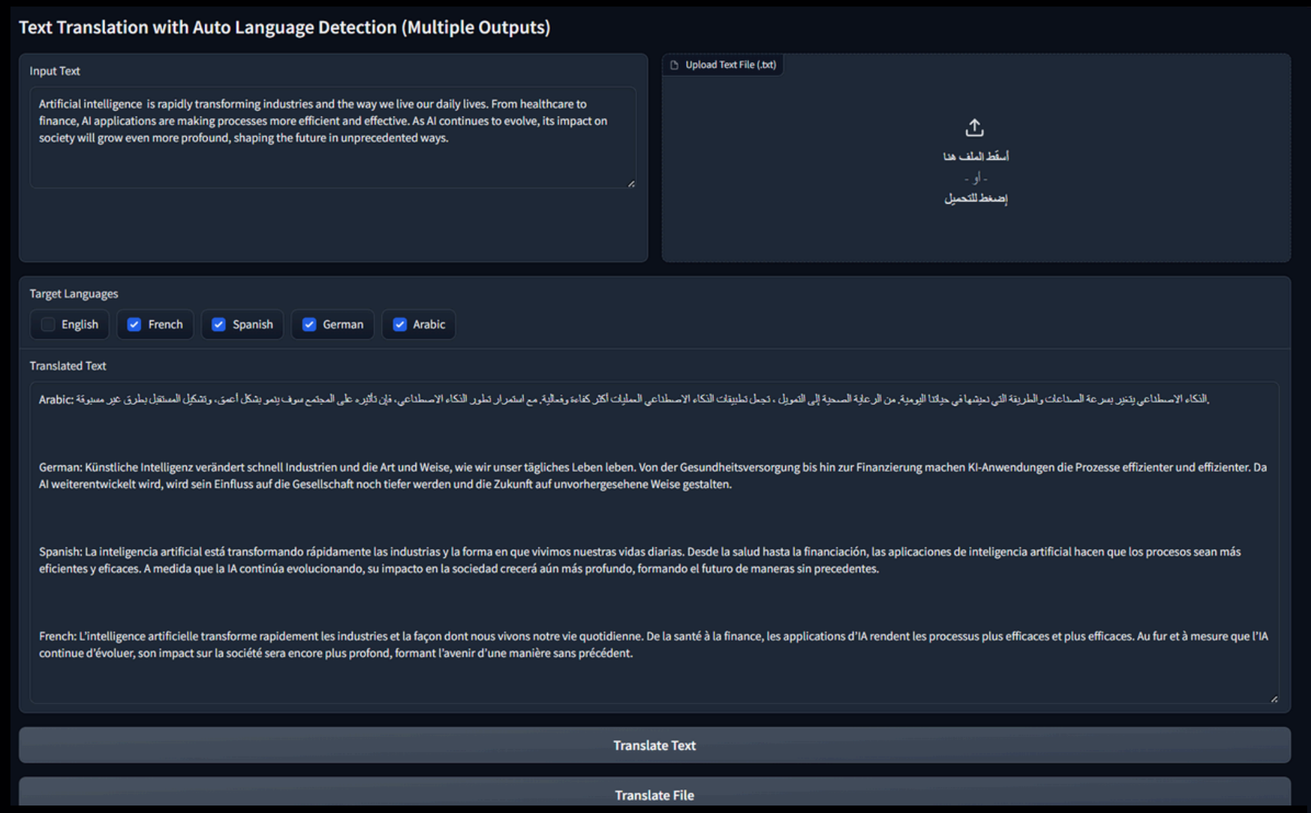
Project Overview:

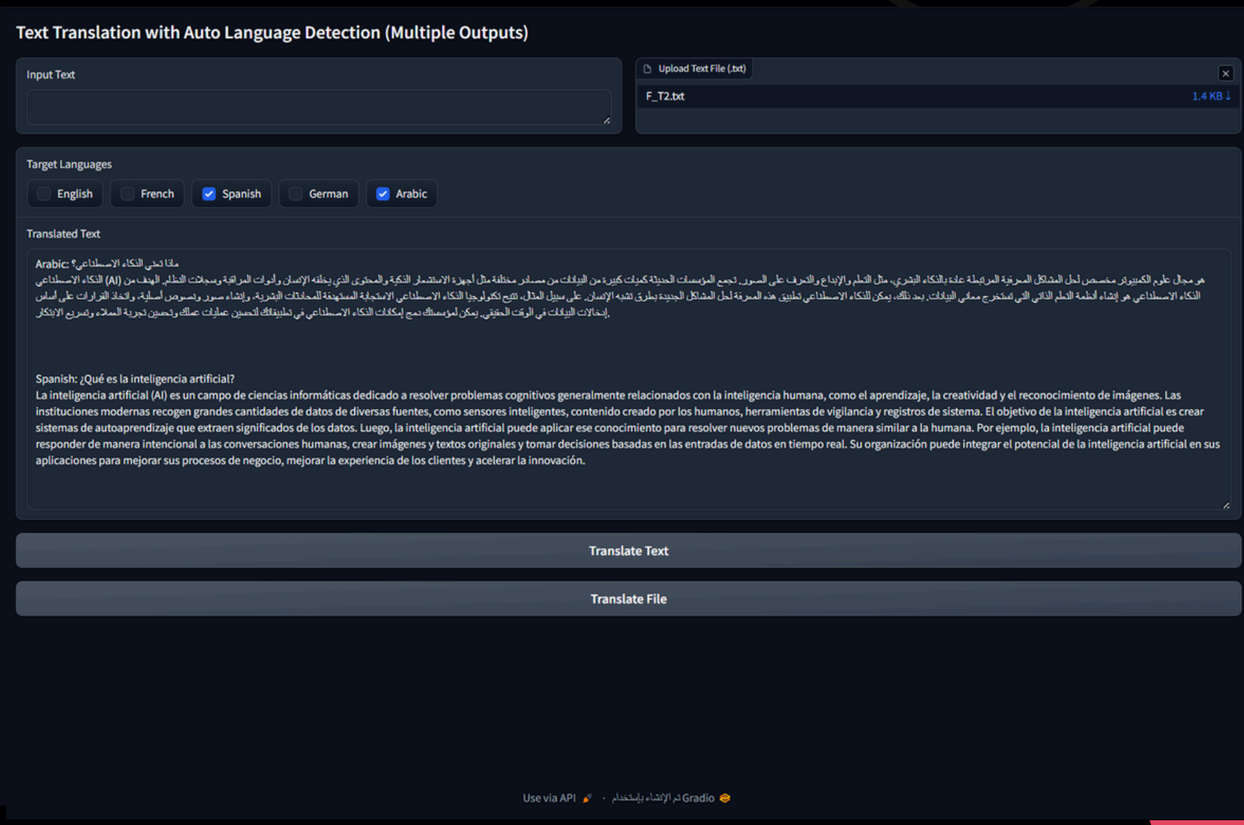
This project was focused on translating text using the **M2M100(418M)** model by Facebook AI, with automatic detection of the input language.

Features:

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

Text-Translation Results





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


THANK YOU