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 examples.
- **Speech Output**: Convert the generated captions to speech.
- **Multilingual Support**: Provide caption and speech output in multiple languages.

Model Justification

- Translation Models
- **Challenge**: Difficulty finding a reliable translation model on Hugging Face.
- **Decision**: Chose **GoogleTranslator** for accurate and consistent translations.

- Text-to-Speech Models
- **Challenge**: Hugging Face TTS models either produced errors or had limited language support.
- **Decision**: Opted for **gTTS (Google Text-to-Speech)** for high-quality multilingual speech synthesis

Pipeline Implementation

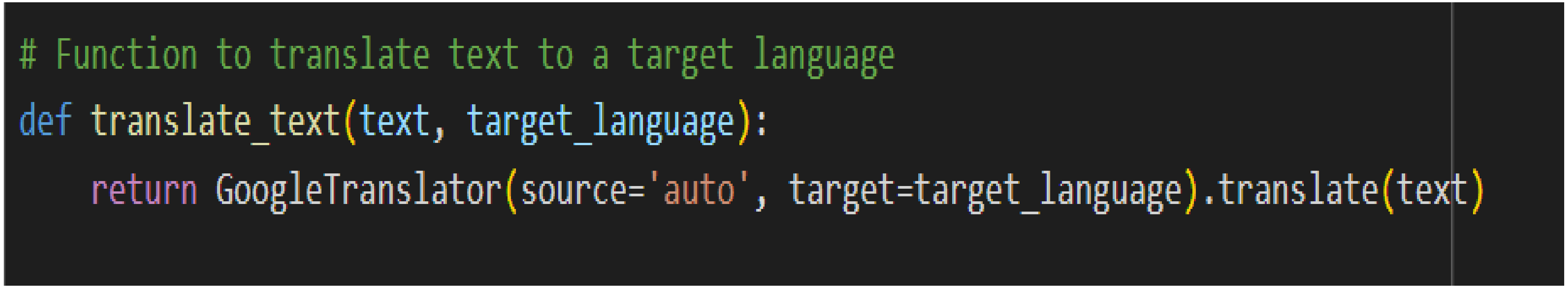
- **Step 1: Load Models**:
- **BLIP (Bootstrapping Language-Image Pre-training)**: Used for automatic image captioning
- **Stable Diffusion**: Used for generating images from text descriptions
- **Google Translator**: Used for translating captions into multiple languages
- **gTTS (Google Text-to-Speech)**: 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

Pipeline Implementation (Continued)

- **Step 3: Text-to-Image Generation:**
- **Text Input**: The user provides a text description or selects from predefined suggestions.
- **Optional 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.

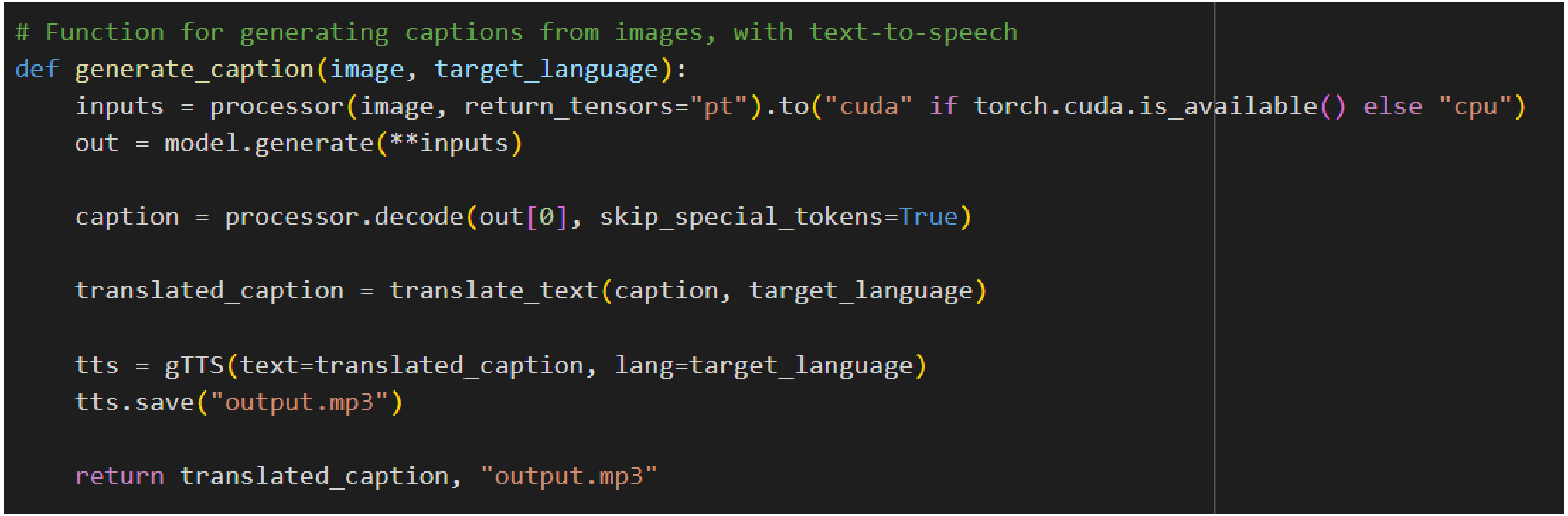
Code Snippets



```
# 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_text function uses the **GoogleTranslator** library to automatically translate text from the detected source language (source='auto’) to the specified target language.

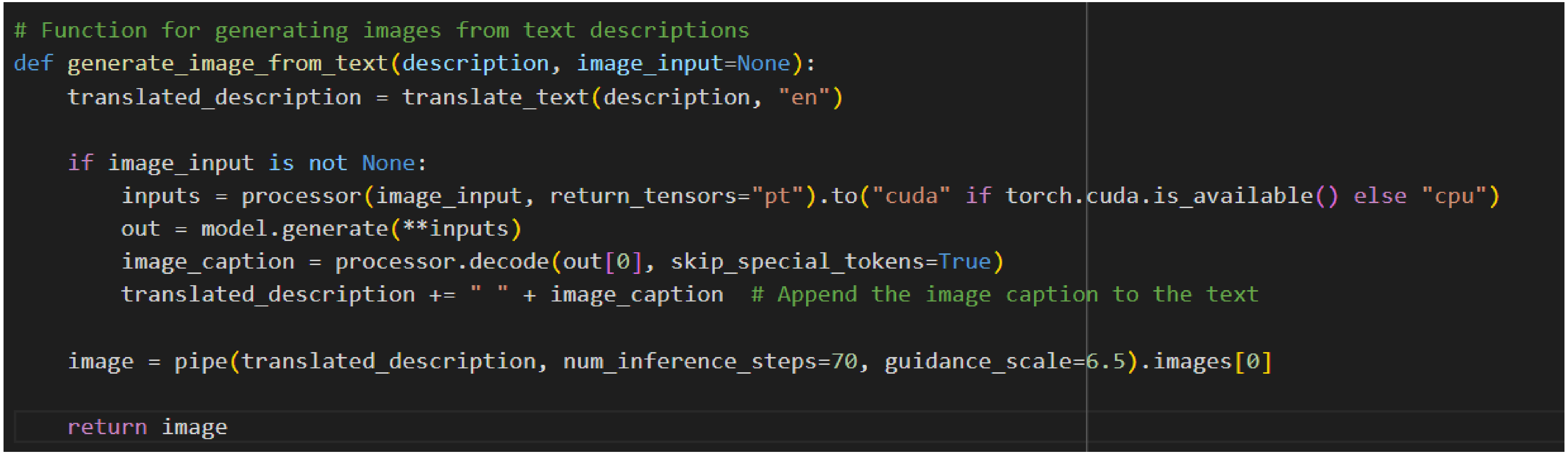
Image Captioning Function



```
# 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_caption** function processes an input image to generate a caption using the BLIP model. It then translates the caption to the selected language using the **translate_text** function. 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 (Arabic , English)

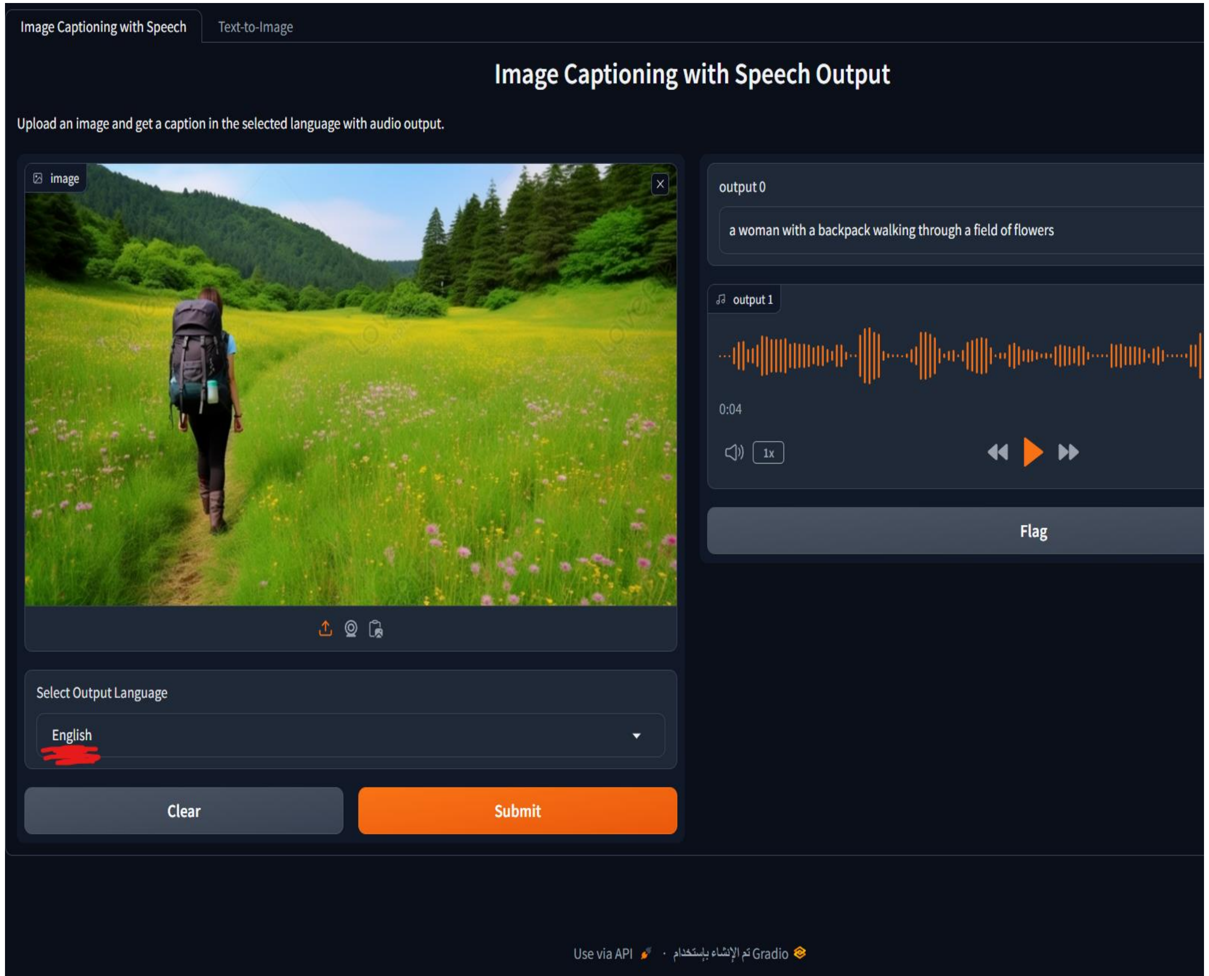


Image Captioning with Speech | Text-to-Image

Image Captioning with Speech Output

Upload an image and get a caption in the selected language with audio output.

output 0

a woman with a backpack walking through a field of flowers

Select Output Language: English

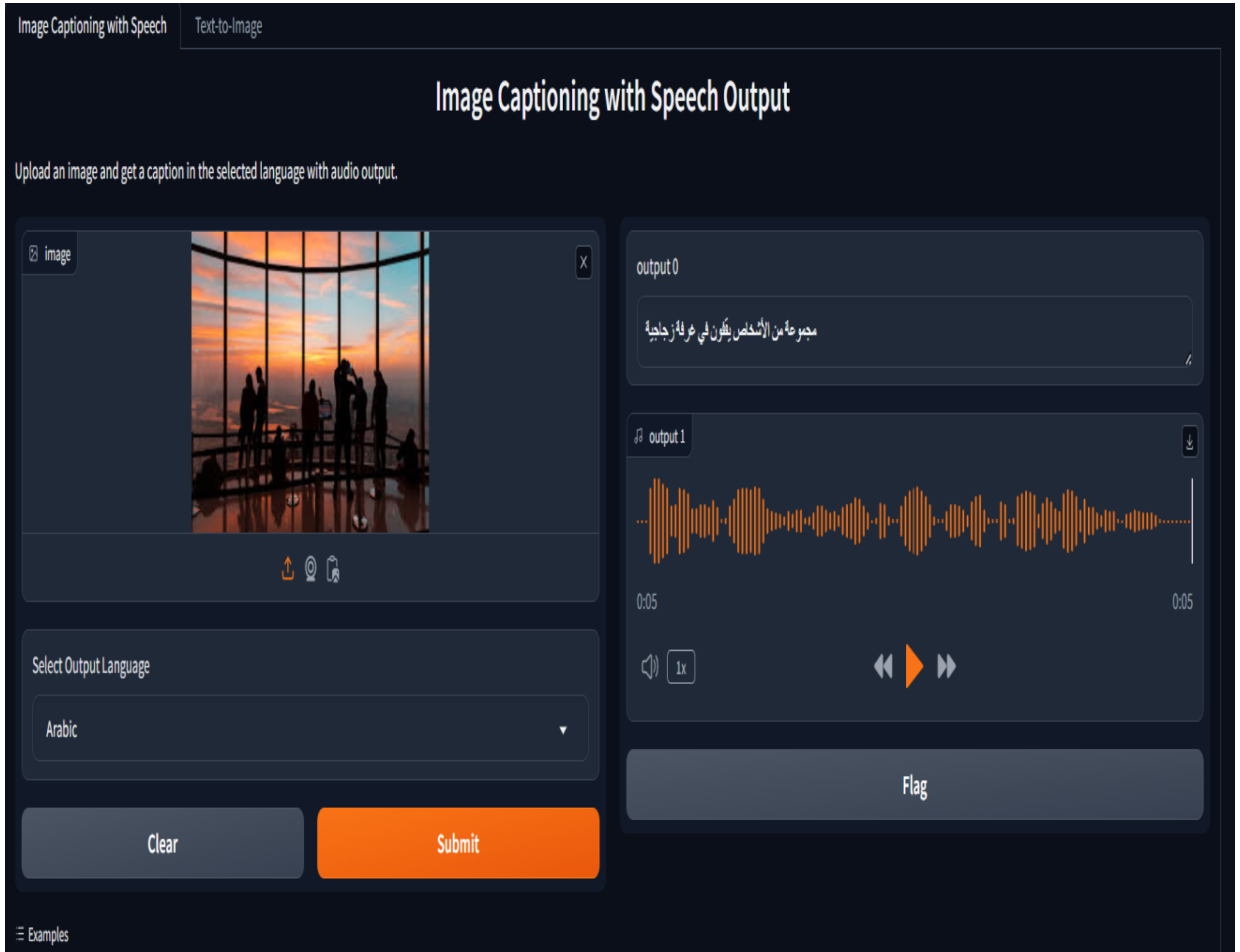


Image Captioning with Speech | Text-to-Image

Image Captioning with Speech Output

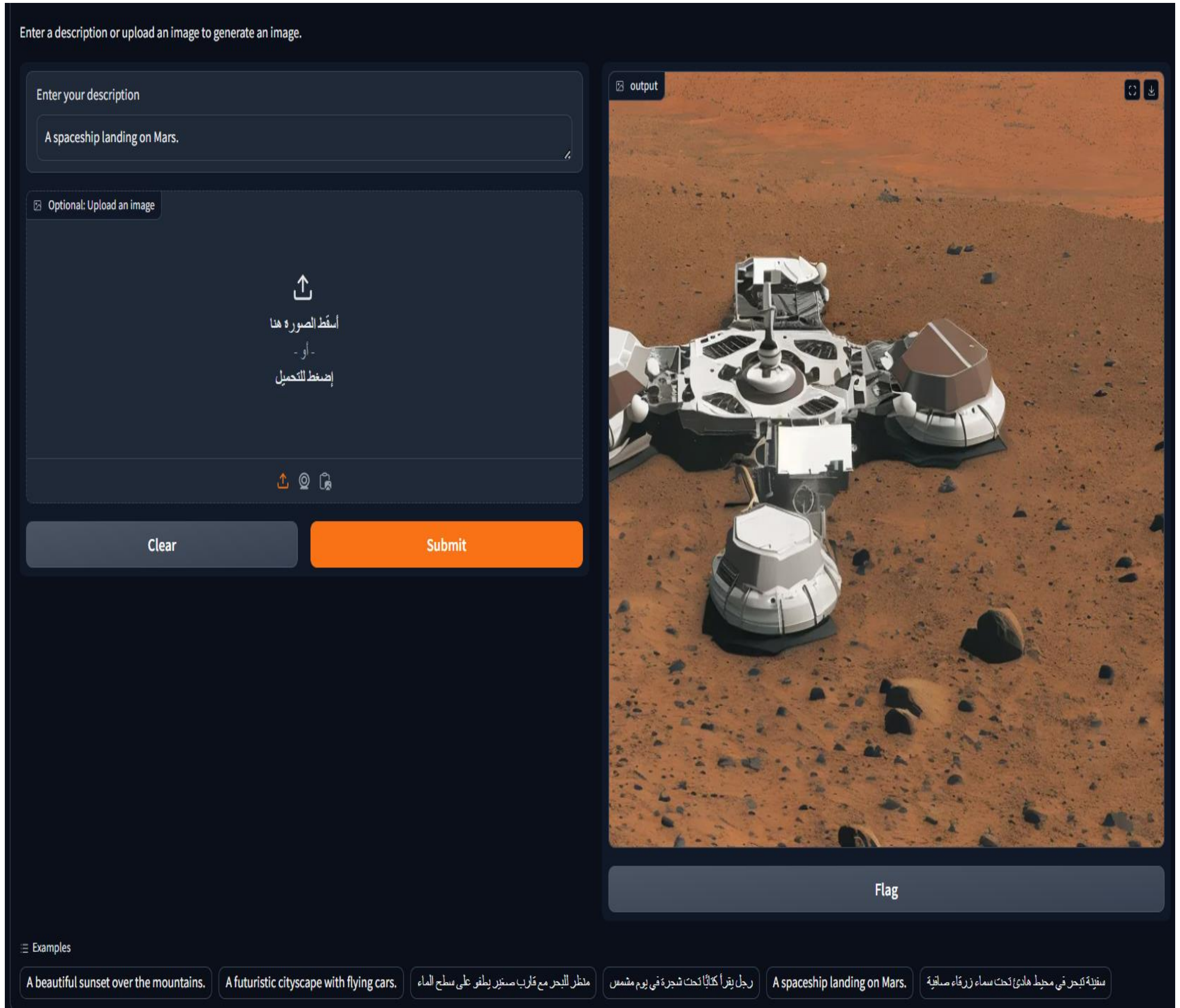
Upload an image and get a caption in the selected language with audio output.

output 0

مجموعة من الأشخاص يقفون في غرفة زجاجية

Select Output Language: Arabic

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



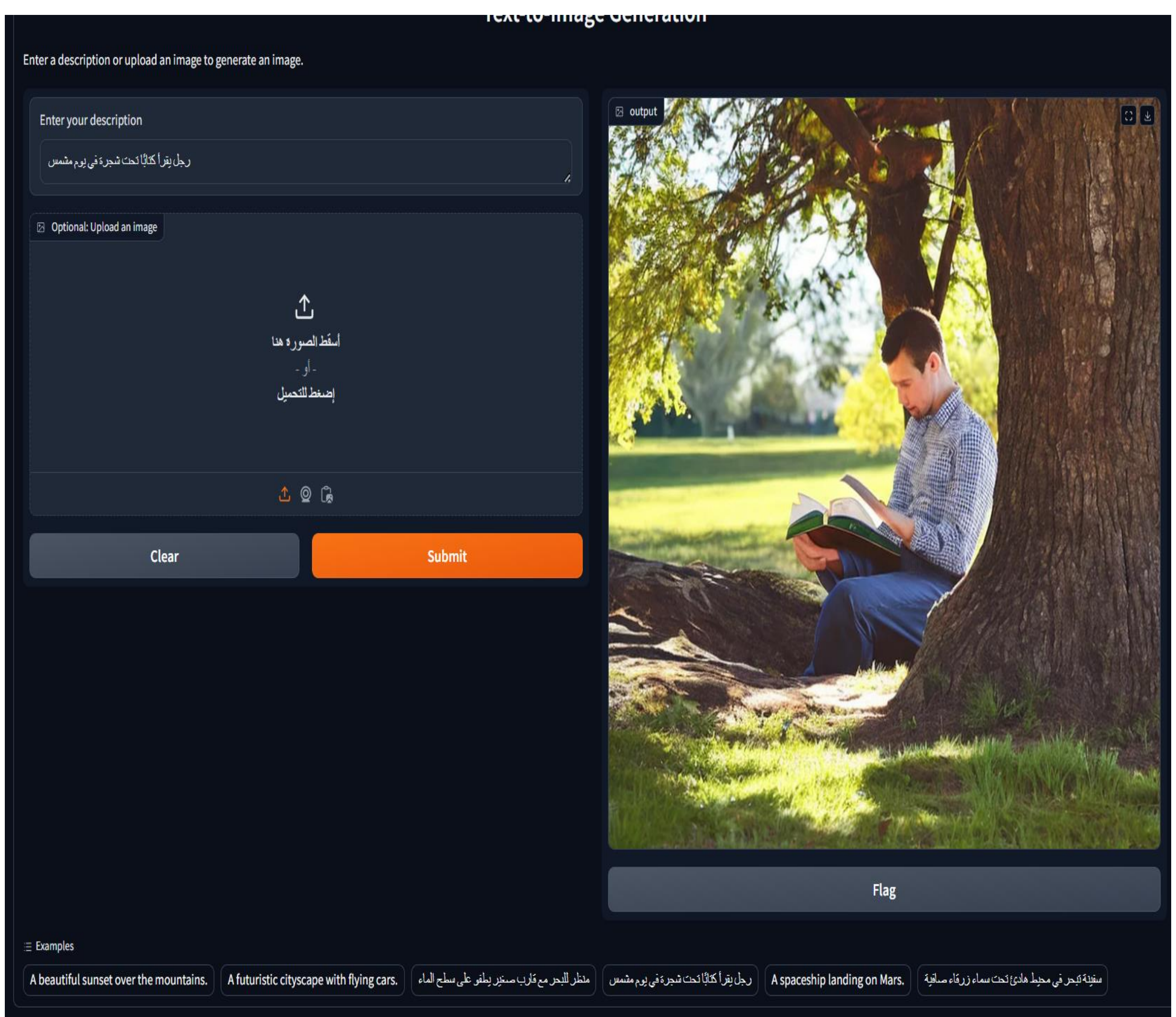
Enter a description or upload an image to generate an image.

Enter your description

A spaceship landing on Mars.

Examples

A beautiful sunset over the mountains. | A futuristic cityscape with flying cars. | منظر للبحر مع قارب صغير يطفو على سطح الماء | رجل يقرأ كتابًا تحت شجرة في يوم مشمس | A spaceship landing on Mars. | سفينة تبحر في محيط هادئ تحت سماء زرقاء صافية



Text-to-Image Generation

Enter a description or upload an image to generate an image.

Enter your description

رجل يقرأ كتابًا تحت شجرة في يوم مشمس

Examples

A beautiful sunset over the mountains. | A futuristic cityscape with flying cars. | منظر للبحر مع قارب صغير يطفو على سطح الماء | رجل يقرأ كتابًا تحت شجرة في يوم مشمس | A spaceship landing on Mars. | سفينة تبحر في محيط هادئ تحت سماء زرقاء صافية

Previous Work: Text Translation Project

- **Project Overview**:

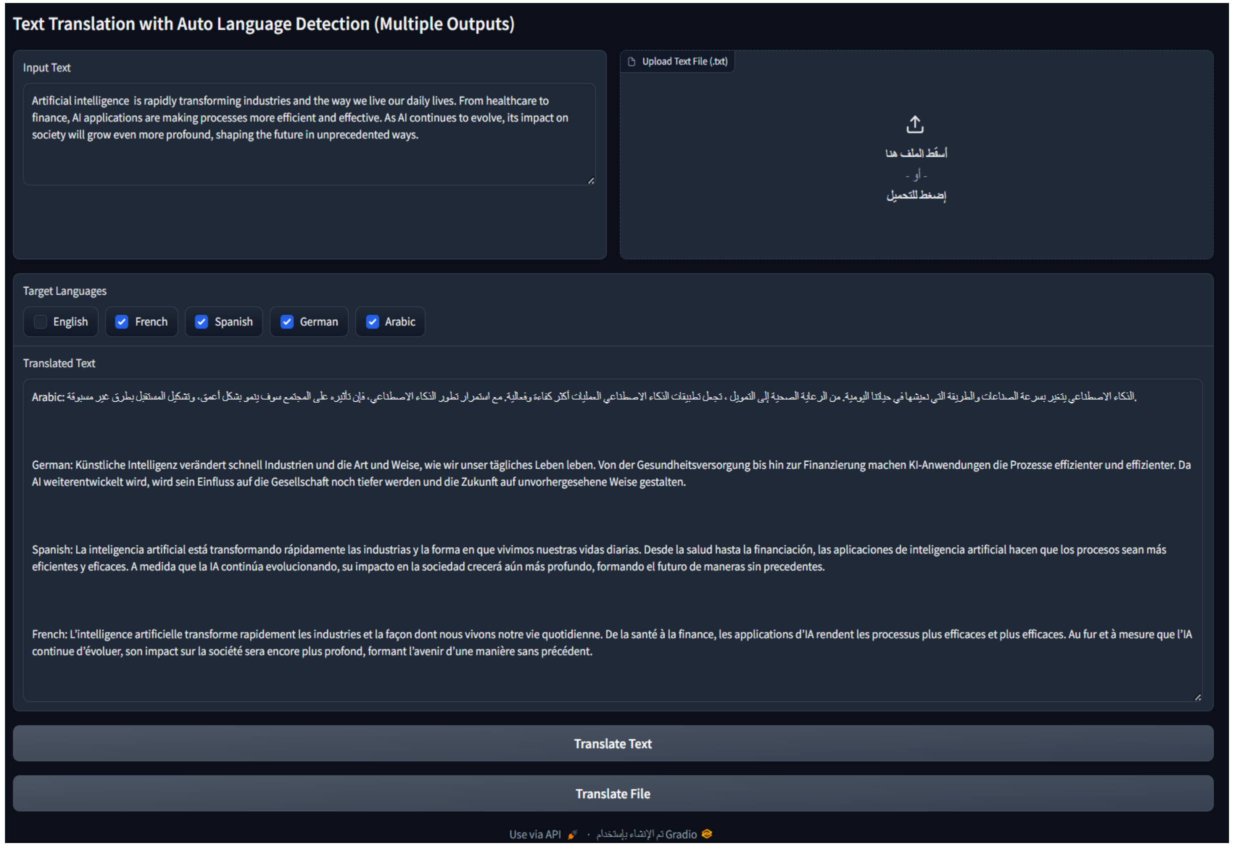
This project was focused on translating text using the **M2M100** (**418M**) model by Facebook AI, with automatic detection of the input language.

Features:

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

Easy-to-use web interface using Gradio.

Text-Translation Results



Text Translation with Auto Language Detection (Multiple Outputs)

Input Text

Artificial intelligence is rapidly transforming industries and the way we live our daily lives. From healthcare to finance, AI applications are making processes more efficient and effective. As AI continues to evolve, its impact on society will grow even more profound, shaping the future in unprecedented ways.

Upload Text File (.txt)

أسقط الملف هنا
- أو -
إضغط للتحميل

Target Languages

☐ English ☑ French ☑ Spanish ☑ German ☑ Arabic

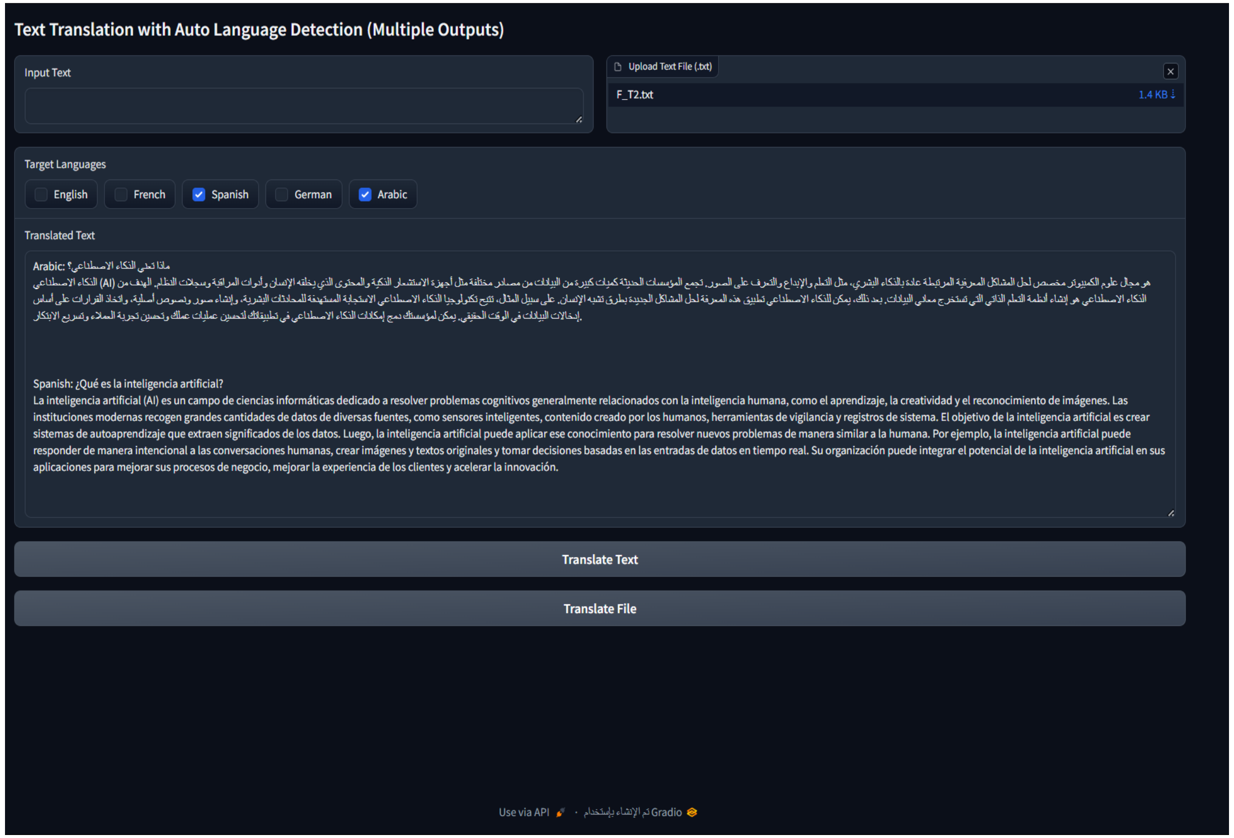
Translated Text

Arabic: الذكاء الاصطناعي يتغير بسرعة الصناعات والطريقة التي نعيشها في حياتنا اليومية. من الرعاية الصحية إلى التمويل ، تجعل تطبيقات الذكاء الاصطناعي العمليات أكثر كفاءة وفعالية. مع استمرار تطور الذكاء الاصطناعي، فإن تأثيره على المجتمع سوف ينمو بشكل أعمق، وتشكيل المستقبل بطرق غير مسبوقة.

German: Künstliche Intelligenz verändert schnell Industrien und die Art und Weise, wie wir unser tägliches Leben leben. Von der Gesundheitsversorgung bis hin zur Finanzierung machen KI-Anwendungen die Prozesse effizienter und effizienter. Da AI weiterentwickelt wird, wird sein Einfluss auf die Gesellschaft noch tiefer werden und die Zukunft auf unvorhergesehene Weise gestalten.

Spanish: La inteligencia artificial está transformando rápidamente las industrias y la forma en que vivimos nuestras vidas diarias. Desde la salud hasta la financiación, las aplicaciones de inteligencia artificial hacen que los procesos sean más eficientes y eficaces. A medida que la IA continúa evolucionando, su impacto en la sociedad crecerá aún más profundo, formando el futuro de maneras sin precedentes.

French: L'intelligence artificielle transforme rapidement les industries et la façon dont nous vivons notre vie quotidienne. De la santé à la finance, les applications d'IA rendent les processus plus efficaces et plus efficaces. Au fur et à mesure que l'IA continue d'évoluer, son impact sur la société sera encore plus profond, formant l'avenir d'une manière sans précédent.



Text Translation with Auto Language Detection (Multiple Outputs)

Input Text

Upload Text File (.txt)

F_T2.txt 1.4 KB ↓

Target Languages

☐ English ☐ French ☑ Spanish ☐ German ☑ Arabic

Translated Text

Arabic: ماذا تعني الذكاء الاصطناعي؟
الذكاء الاصطناعي (AI) هو مجال علوم الكمبيوتر مخصص لحل المشاكل المعرفية المرتبطة عادة بالذكاء البشري، مثل التعلم والإبداع والتعرف على الصور. تجمع المؤسسات الحديثة كميات كبيرة من البيانات من مصادر مختلفة مثل أجهزة الاستشعار الذكية والمحتوى الذي يخلقه الإنسان وأدوات المراقبة وسجلات النظام. الهدف من الذكاء الاصطناعي هو إنشاء أنظمة التعلم الذاتي التي تستخرج معاني البيانات. بعد ذلك، يمكن للذكاء الاصطناعي تطبيق هذه المعرفة لحل المشاكل الجديدة بطرق تشبه الإنسان. على سبيل المثال، تتيح تكنولوجيا الذكاء الاصطناعي الاستجابة المستهدفة للمحادثات البشرية، وإنشاء صور ونصوص أصلية، واتخاذ القرارات على أساس إدخالات البيانات في الوقت الحقيقي. يمكن لمؤسستك دمج إمكانات الذكاء الاصطناعي في تطبيقاتك لتحسين عمليات عملك وتحسين تجربة العملاء وتسريع الابتكار.

Spanish: ¿Qué es la inteligencia artificial?
La inteligencia artificial (AI) es un campo de ciencias informáticas dedicado a resolver problemas cognitivos generalmente relacionados con la inteligencia humana, como el aprendizaje, la creatividad y el reconocimiento de imágenes. Las instituciones modernas recogen grandes cantidades de datos de diversas fuentes, como sensores inteligentes, contenido creado por los humanos, herramientas de vigilancia y registros de sistema. El objetivo de la inteligencia artificial es crear sistemas de autoaprendizaje que extraen significados de los datos. Luego, la inteligencia artificial puede aplicar ese conocimiento para resolver nuevos problemas de manera similar a la humana. Por ejemplo, la inteligencia artificial puede responder de manera intencional a las conversaciones humanas, crear imágenes y textos originales y tomar decisiones basadas en las entradas de datos en tiempo real. Su organización puede integrar el potencial de la inteligencia artificial en sus aplicaciones para mejorar sus procesos de negocio, mejorar la experiencia de los clientes y acelerar la innovación.

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