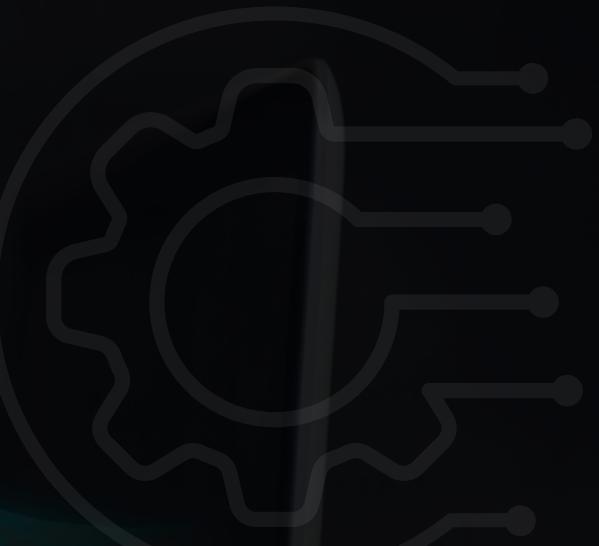


# IMAGE CAPTIONING AND TEXT TO IMAGE

BY: YAZEED ABDULLAH



# OBJECTIVES

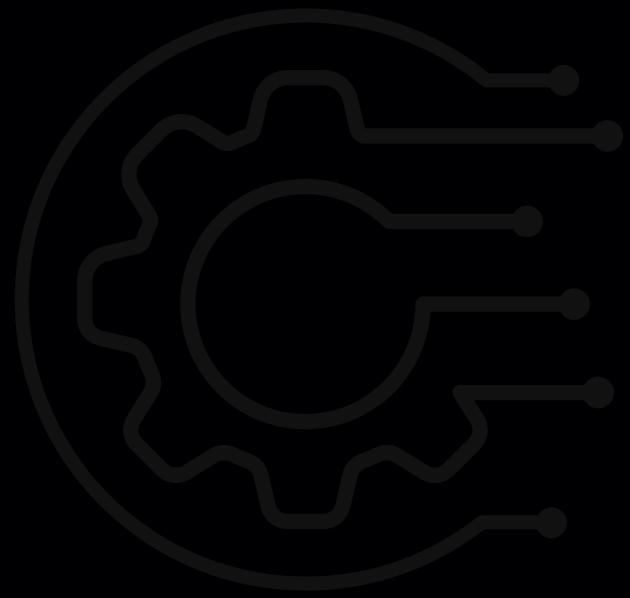
Enhance Visual Understanding through Automated Image Descriptions.

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

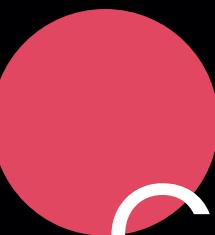
**Image Input:** The user can upload an image to enhance the output.

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



# Model Justification

# Translation Models



# 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\_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

```
# 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 with Speech Output

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

 image

a woman with a backpack walking through a field of flowers

Select Output Language

English

Clear

Submit

output 0

a woman with a backpack walking through a field of flowers

output 1

0:04

Flag

# Image Captioning Results

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.

image



output0

مجموعة من الأشخاص يملون في غرفة زجاجية

output1

0:05 0:05

Flag

Select Output Language

Arabic

Clear Submit

This interface displays the results of an image captioning task. At the top, there are two tabs: "Image Captioning with Speech" (selected) and "Text-to-Image". Below the tabs, the title "Image Captioning with Speech Output" is centered. A sub-instruction "Upload an image and get a caption in the selected language with audio output." is present. On the left, there is a file input field labeled "image" with a checked checkbox, containing a thumbnail of a sunset scene viewed through a large glass window, showing silhouettes of several people. Below the image input is a "Select Output Language" dropdown menu set to "Arabic". At the bottom of the left column are two buttons: "Clear" (gray) and "Submit" (orange). To the right of the image input, there are two output sections. The first section, "output0", shows the caption "مجموعة من الأشخاص يملون في غرفة زجاجية" (A group of people are standing in a glass room). The second section, "output1", shows an audio waveform and a play bar indicating a duration of 0:05, with playback controls for previous, next, and double-speed. A "Flag" button is located at the bottom right of the output area.

# Text-to-Image Generation Results

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

Enter your description  
A spaceship landing on Mars.

Optional: Upload an image

↑  
أسقط الصورة هنا  
أو -  
إضغط للتحميل

Clear Submit

output



Flag

Examples

A beautiful sunset over the mountains. مغارة تحيط بهلبي تحت سماء زرقاء صافية

A futuristic cityscape with flying cars. منظر لليحر مع قارب صغير يطفو على سطح الماء

منظر لليحر مع قارب صغير يطفو على سطح الماء

رجل يقرأ كتاباً تحت شجرة في يوم مشمس

A spaceship landing on Mars.

# 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



## 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.
```

أسقط الملف هنا  
أو -  
[اضغط للتحميل]

Target Languages

English  French  Spanish  German  Arabic

Translated Text

الذكاء الاصطناعي يتغير بسرعة الصداعات والطريق الذي تعيشها في حياتنا اليومية، من الرعاية الصحية إلى التمويل. تحول تطبيقات الذكاء الاصطناعي العمليات أكثر كفاءة وفعالية، مع استمرار تطور الذكاء الاصطناعي، فإن تأثيره على المجتمع سوف يتمثل بشكل أكبر، وتحتكر المستقبل بطرق غير مسبوقة.

Arabic: ماذا تعني الكلمات الجديدة؟

Spanish: ¿Qué es la inteligencia artificial?

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.

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.

Translate Text

Translate File

## Text Translation with Auto Language Detection (Multiple Outputs)

Input Text

```
F_T2.txt
```

1.4 KB

Target Languages

English  French  Spanish  German  Arabic

Translated Text

Arabic: ماذا تعني الكلمات الجديدة؟

Spanish: ¿Qué es la inteligencia artificial?

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.

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.

Translate Text

Translate File



# GitHub & Hugging Face Links



## GitHub Repository:

<https://github.com/Yazl-e/Image-Captioning-Text2Image>

## Hugging Face Space:

<https://huggingface.co/spaces/Yazl-e/Image-Captioning-Text2Image>



# THANK YOU

