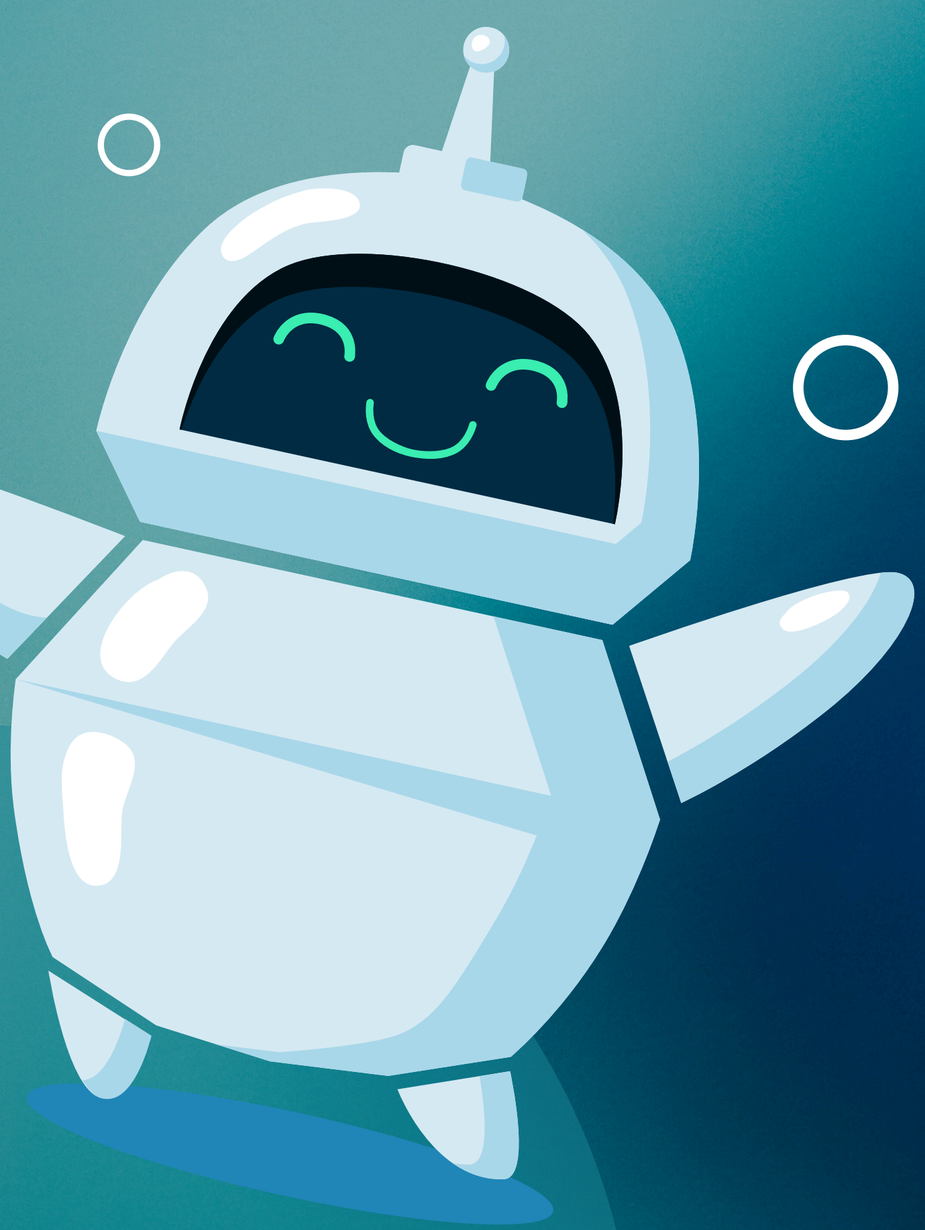
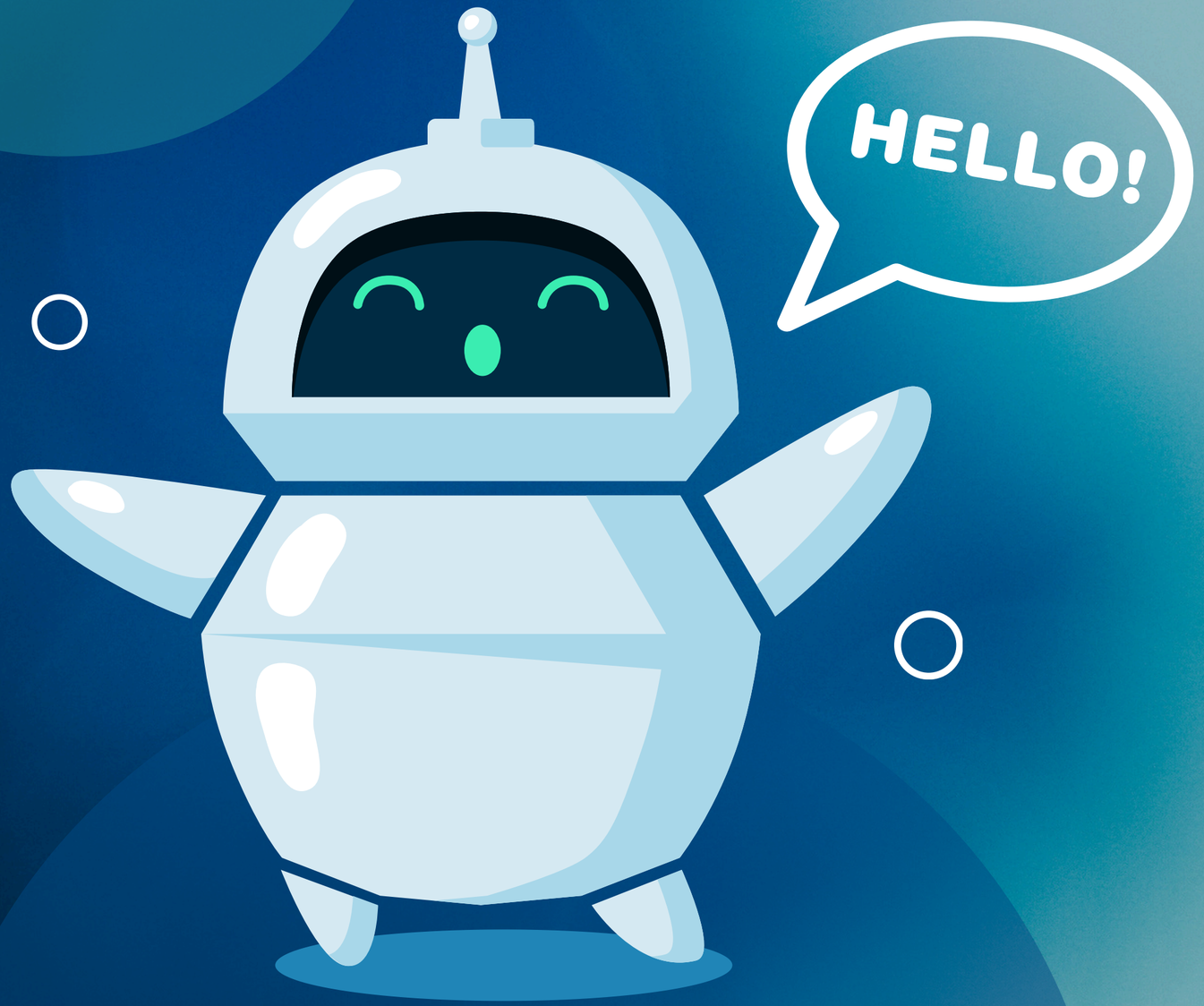


# Image Captioning in Arabic (Multilingual Seq2Seq)

## Deep Learning

# Team Work

Hagar Khaled Ibrahim
Nada Maher Mohamed Elhady
Menna Allah Elsaeed Abdelfattah
Hanaa Mahmoud Ahmed



# Over view :

- Problem statement: Automatically generate meaningful captions for images:
  - Captions must describe relevant objects and actions.
  - Simple models can produce incomplete captions.
- The project focuse on descripe captions for images and translate them into Arabic.
- It combines computer vision and natural language processing to generate.

# Dataset

**The Flickr8k dataset : flickr8k dataset**

- 8,000 images, each with 5 descriptive captions.
- Captions describe main objects and events.
- Used for image captioning and text-to-image search.

**The Arabic to English Translation Sentence: arabic to english translation sentence**

- Contains parallel sentences in Arabic and English.
- Helps train models for sentence translation.
- Useful for English to Arabic caption translation.

# Goal

- Develop a fully integrated system for generating image captions in both English and Arabic.
- Generate high-quality English captions from images using the fine-tuned BLIP model.
- Improve system speed and efficiency by reducing preprocessing steps.

# Goal

- Translate English captions into Arabic with the MBART50 NMT “neural machine translation” model, ensuring fluency and semantic accuracy.
- Produce accurate and context-aware captions that reflect the content of the image.
- Combine computer vision, natural language generation, and machine translation into a single, seamless workflow.

# Model Overview

- From-Scratch Model (ResNet “encoder” + LSTM “decoder” + Attention).
- Bootstrapping Language Image Pretraining model (BLIP) “image to english caption model”.
- Multilingual Bidirectional Auto-Regressive Transformers 50 (mBART50) “English to Arabic Translation Model”.

# Results

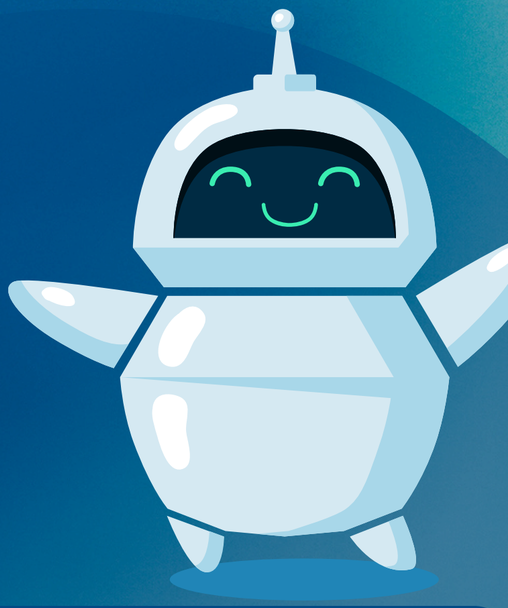
- Accurate English captions generated using BLIP model
- Improved caption relevance after fine-tuning.
- Fluent Arabic translations using mBART50 model
- Successful end-to-end bilingual image captioning system
- The system produces visually relevant and linguistically coherent captions

# Performance of From-Scratch Encoder–Decoder (ResNet50 + LSTM + Attention) :

- Final training loss (epoch 29): 0.6489
- Final validation loss (epoch 29): 0.9345
- Early stopping triggered at: epoch 29
- The best model checkpoint was saved when val_loss = 0.9313 at epoch 24

```
Epoch 24/60
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 0s 329ms/step - loss: 0.6952
Epoch 24: val_loss improved from 0.93239 to 0.93133, saving model to best_model.weights.h5
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 360s 356ms/step - loss: 0.6952 - val_loss: 0.9313 - learning_rate: 1.0000e-
Epoch 25/60
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 0s 328ms/step - loss: 0.6859
Epoch 25: val_loss did not improve from 0.93133
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 357s 353ms/step - loss: 0.6859 - val_loss: 0.9328 - learning_rate: 1.0000e-
Epoch 26/60
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 0s 330ms/step - loss: 0.6782
Epoch 26: val_loss did not improve from 0.93133
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 359s 355ms/step - loss: 0.6782 - val_loss: 0.9367 - learning_rate: 1.0000e-
Epoch 27/60
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 0s 330ms/step - loss: 0.6716
Epoch 27: ReduceLROnPlateau reducing learning rate to 1.9999999494757503e-05.

Epoch 27: val_loss did not improve from 0.93133
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 359s 355ms/step - loss: 0.6716 - val_loss: 0.9322 - learning_rate: 1.0000e-
Epoch 28/60
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 0s 331ms/step - loss: 0.6600
Epoch 28: val_loss did not improve from 0.93133
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 359s 356ms/step - loss: 0.6600 - val_loss: 0.9362 - learning_rate: 2.0000e-
Epoch 29/60
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 0s 330ms/step - loss: 0.6489
Epoch 29: val_loss did not improve from 0.93133
1011/1011 ━━━━━━━━━━━━━━━━━━━━ 359s 355ms/step - loss: 0.6489 - val_loss: 0.9345 - learning_rate: 2.0000e-
```



# Performance of Bootstrapping Language Image Pretraining model (BLIP):

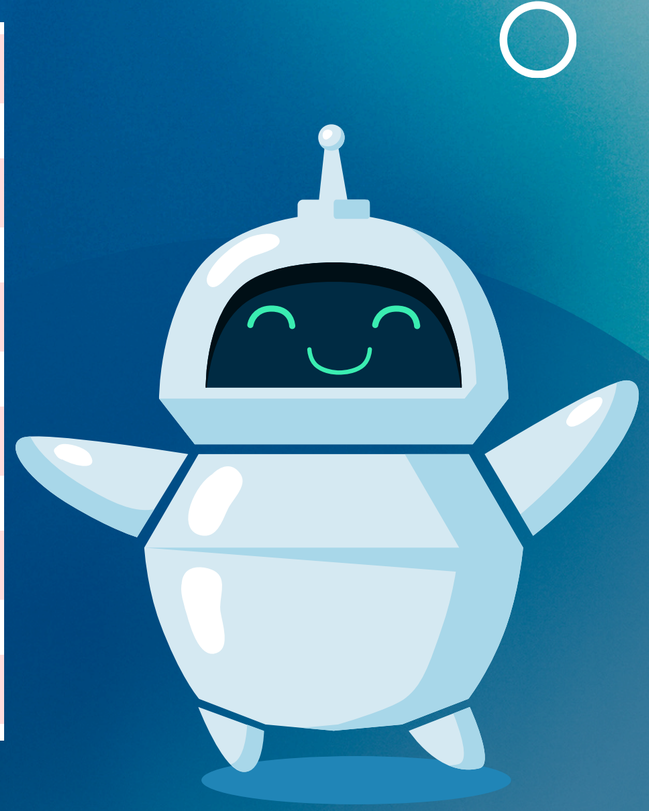
- During training, the model showed stable convergence, with the loss steadily decreasing across epochs.
- Total training epochs: 7
- Final training loss: 0.3868

```
Epoch 1/7: 100%|██████████| 5057/5057 [55:15<00:00,  1.53it/s]
Epoch 1/7 - Average Loss: 1.9890
Epoch 2/7: 100%|██████████| 5057/5057 [55:18<00:00,  1.52it/s]
Epoch 2/7 - Average Loss: 1.5720
Epoch 3/7: 100%|██████████| 5057/5057 [55:18<00:00,  1.52it/s]
Epoch 3/7 - Average Loss: 1.2803
Epoch 4/7: 100%|██████████| 5057/5057 [55:20<00:00,  1.52it/s]
Epoch 4/7 - Average Loss: 1.0073
Epoch 5/7: 100%|██████████| 5057/5057 [55:18<00:00,  1.52it/s]
Epoch 5/7 - Average Loss: 0.7538
Epoch 6/7: 100%|██████████| 5057/5057 [55:17<00:00,  1.52it/s]
Epoch 6/7 - Average Loss: 0.5420
Epoch 7/7: 100%|██████████| 5057/5057 [55:20<00:00,  1.52it/s]
Epoch 7/7 - Average Loss: 0.3868
```

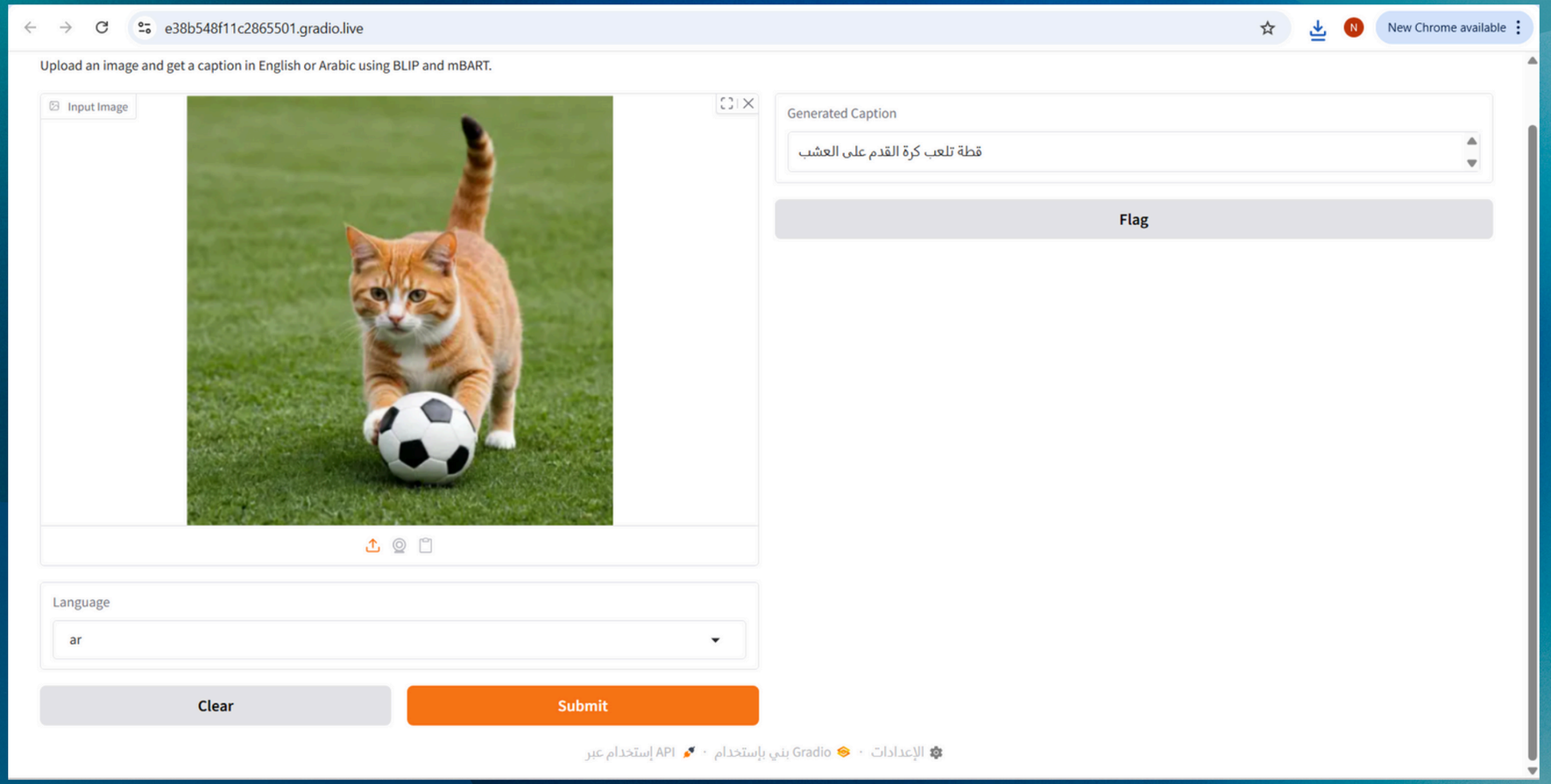
# Performance Of Pretrained Model mBART50 for English to Arabic Translation:

- Total training epochs: 5
- Final training loss: 0.6618

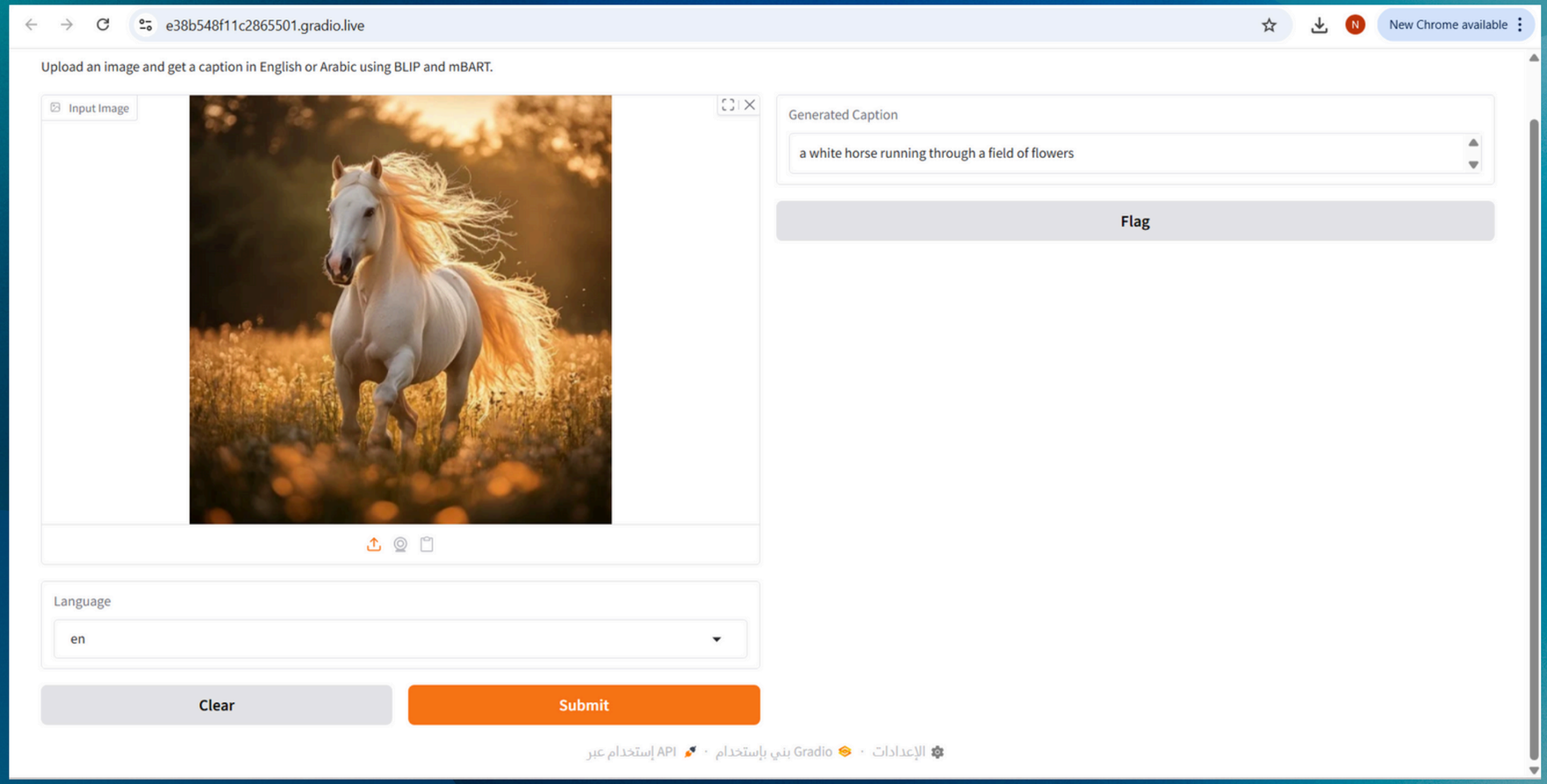
```
Epoch 1: 100%|██████████| 6160/6160 [47:18<00:00,  2.17it/s]
Epoch 1 Loss = 2.20207907373642
Epoch 2: 100%|██████████| 6160/6160 [47:20<00:00,  2.17it/s]
Epoch 2 Loss = 1.5835523873180537
Epoch 3: 100%|██████████| 6160/6160 [47:18<00:00,  2.17it/s]
Epoch 3 Loss = 1.2050856505315024
Epoch 4: 100%|██████████| 6160/6160 [47:19<00:00,  2.17it/s]
Epoch 4 Loss = 0.8990478008163053
Epoch 5: 100%|██████████| 6160/6160 [47:20<00:00,  2.17it/s]
Epoch 5 Loss = 0.6618388353314782
```



# Deployment


e38b548f11c2865501.gradio.live
New Chrome available
Upload an image and get a caption in English or Arabic using BLIP and mBART.
Input Image
Generated Caption
قطة تلعب كرة القدم على العشب
Flag
Language
ar
Clear
Submit
الإعدادات · Gradio بني بإستخدام · API إستخدام عبر

# Deployment


e38b548f11c2865501.gradio.live
New Chrome available
Upload an image and get a caption in English or Arabic using BLIP and mBART.
Input Image
Generated Caption
a white horse running through a field of flowers
Flag
Language
en
Clear
Submit
الإعدادات · Gradio بني بإستخدام · API إستخدام عبر

# Any Question ?

# Thank You