MACHINE TRANSLATION



PROJECT GOAL

- Build an English-to-Arabic translation model.
- Start from a pretrained Transformer and fine-tune it on parallel text.
- Evaluate quality using BLEU and show real translation examples.



TOOLS

- Hugging Face: Transformers, Datasets, Evaluate
- Metric: SacreBLEU
- Demo UI: Gradio



HUGGING FACE

DATASET

Dataset Used:

- OPUS-100 English/Arabic parallel corpus
- Source: MagedSaeed/opus-100_ar_en_experimental

Split Strategy:

- Train set split into 90% train / 10% validation
- Separate test set used for final evaluation



DATA PREPROCESSING

Tokenization Pipeline

- Input: English text
- Target: Arabic text
- Uses text_target= for modern seq2seq tokenization

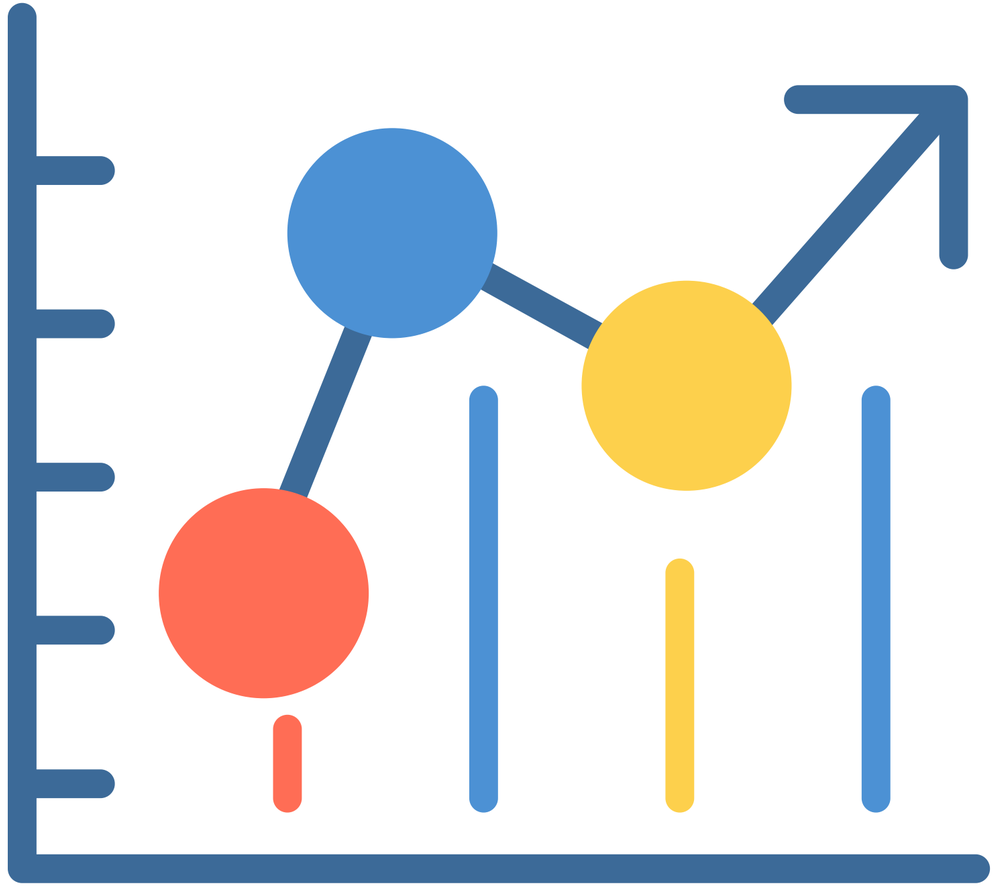
Output

- Tokenized train, validation, and test datasets

TRAINING CONFIGURATION

Trainer Setup (Seq2SeqTrainer)

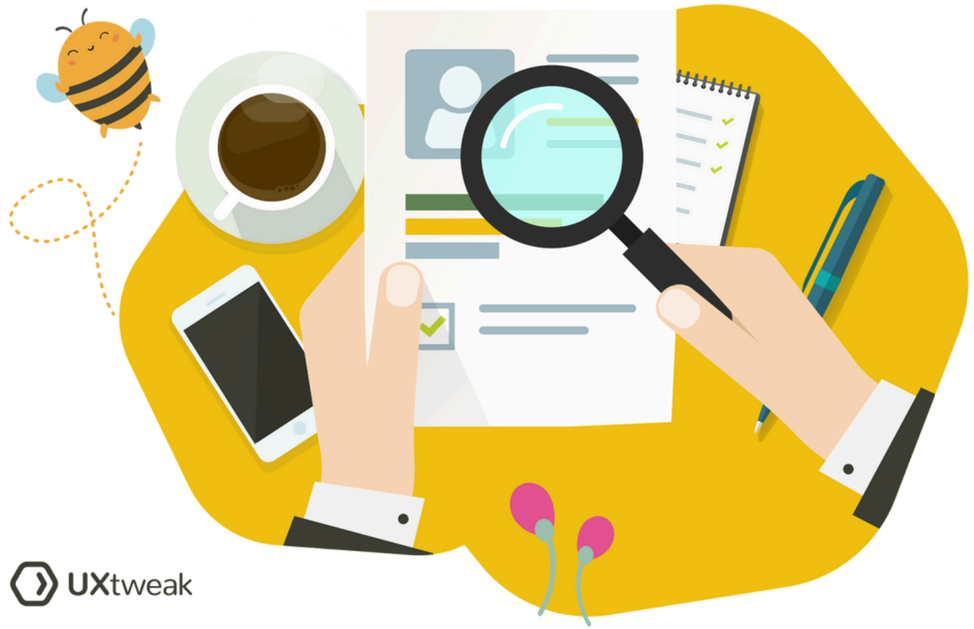
- Epochs: 3
- Learning rate: 1e-5
- Batch size: 16 (train) / 16 (eval)
- Optimizer: Adafactor



EVALUATION

Metric: SacreBLEU

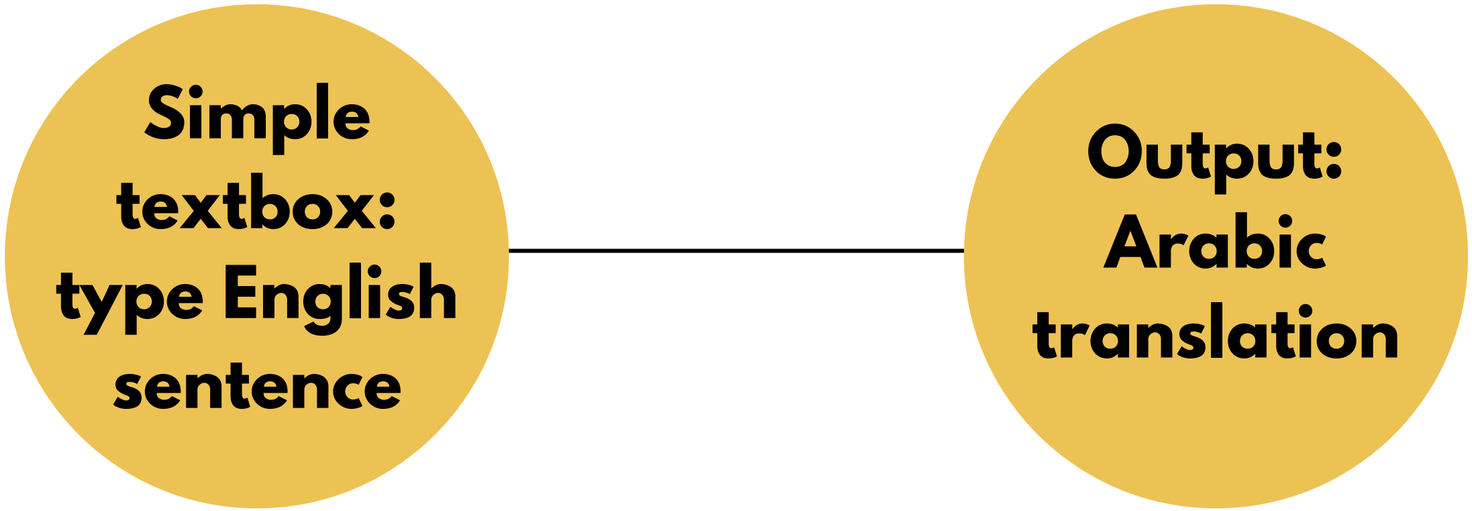
- Measures translation similarity vs reference
- Used during evaluation and for selecting best model

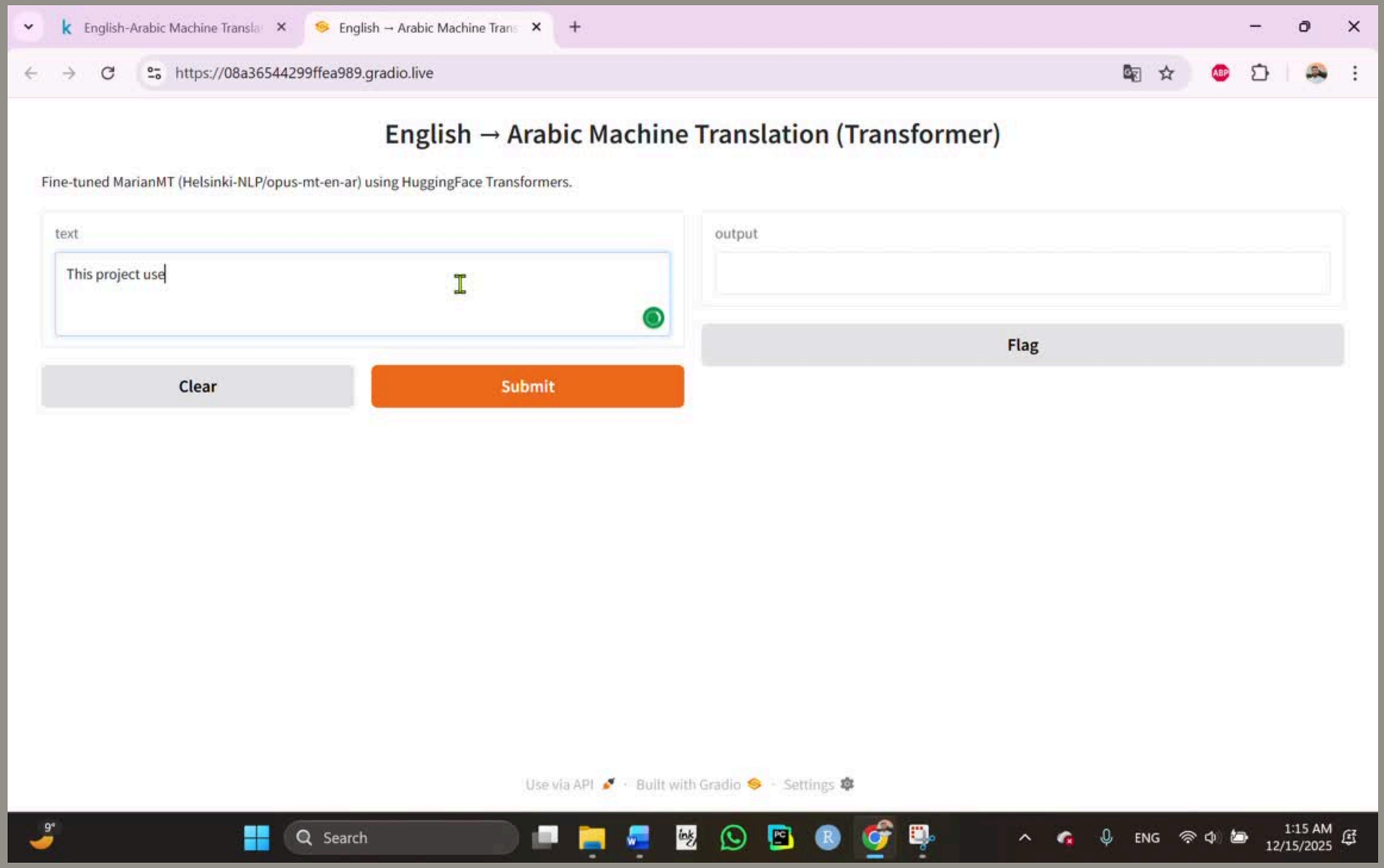


Two Types of Evaluation

- Quantitative: BLEU score on the test set
- Qualitative: random examples (EN → AR prediction vs AR reference)

INTERACTIVE DEMO (GRADIO)




English-Arabic Machine Translat
English → Arabic Machine Trans
https://08a36544299ffea989.gradio.live
English → Arabic Machine Translation (Transformer)
Fine-tuned MarianMT (Helsinki-NLP/opus-mt-en-ar) using HuggingFace Transformers.
text
This project use
output
Flag
Clear
Submit
Use via API · Built with Gradio · Settings
Search
ENG
1:15 AM
12/15/2025

KAGGLE NOTEBOOK CODE



English-Arabic Machine Translation (Transformer)

Explore and run machine learning code with Kaggle Notebooks | Using data from No attached data sources

Kaggle / Dec 15, 2025

THANK YOU

Presented to Dr. Suhaila

+962 78 868 9285 moustafaayoub2005@gmail.com