Prepared by group 14

English-Arabic Bidirectional Translation

BART model fine-tuning with bidirectional capabilities

13 May, 2025

Project Overview

This project implements a bidirectional English-Arabic translation system by fine-tuning Facebook's BART model. The system is designed to translate text in both directions efficiently, using a single model with direction-specific prefixes.

Project Features

User-friendly graphical interface

Bidirectional translation between English and Arabic

Optimized for efficient training and inference

Fine-tuned BART base model (~140M parameters)

Task-specific prefixes for direction control

Dataset

Source: Tatoeba English-Arabic parallel corpus

- **Format**: Text pairs with prefix-based task specification
  - en2ar: [English text]→ Arabic translation
  - ar2en: [Arabic text] → English translation

-**Preparation**:
- Custom preprocessing script (prepare_dataset.py)
- Text cleaning (whitespace normalization)
- Duplicate removal and alignment checking
- Bidirectional dataset creation with task prefixes

Statistics

- Total parallel sentences: ~28,000
- Training set: ~25,200 sentence pairs
- Validation set: ~2,800 sentence pairs
- Average sentence length: 8.7 words (English), 6.9 words (Arabic)
- Domain: General conversation, everyday phrases

Model Architecture

- **Base Model**: facebook/bart-base -
- **Architecture Type**: Encoder-decoder transformer
- **Parameters**: ~140 million parameters
- **Advantages**:
  - Bidirectional attention in encoder (like BERT)
  - Autoregressive decoding (like GPT)
  - Compact enough for efficient fine-tuning
  - Effective for sequence-to-sequence tasks

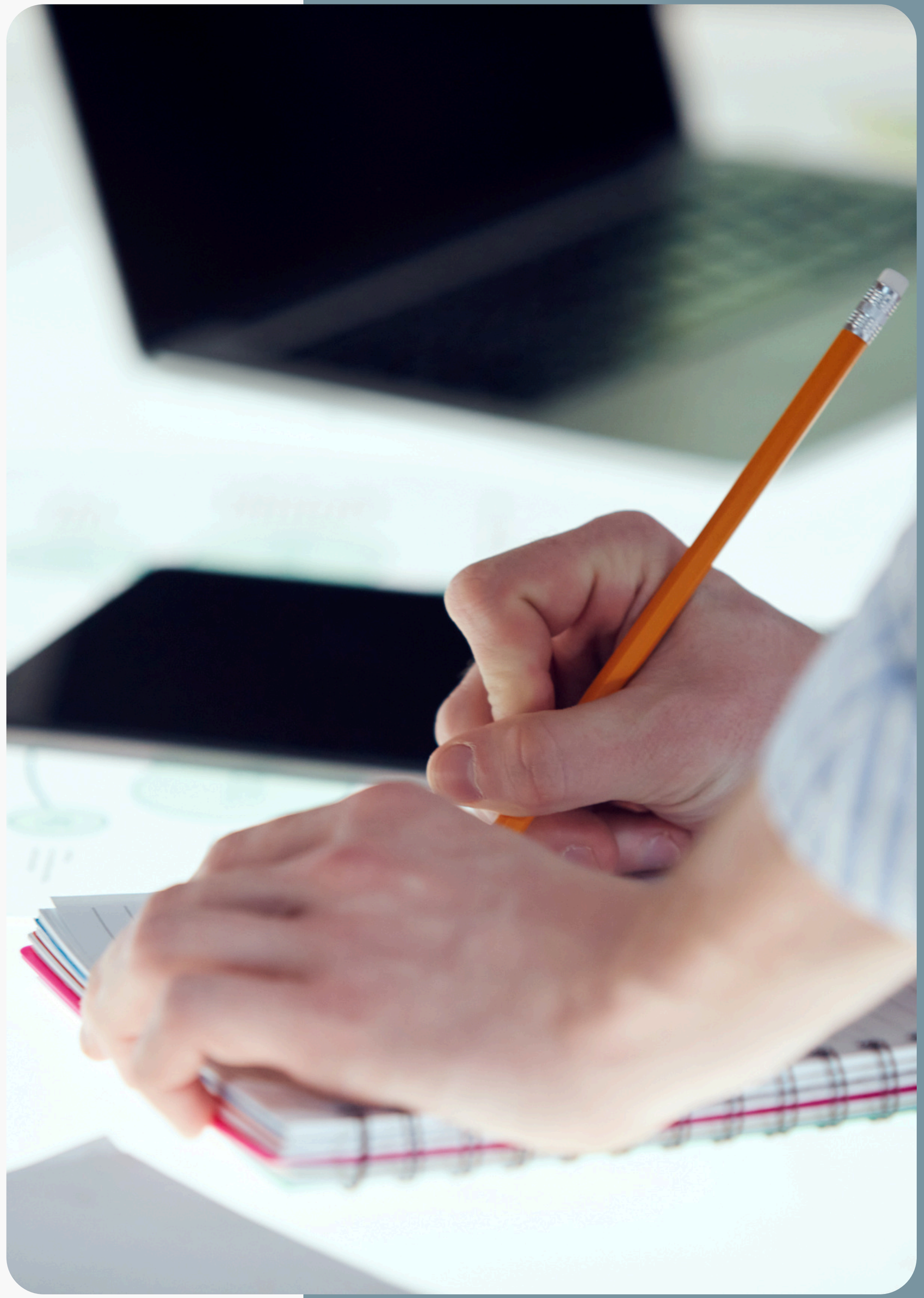
Training Process

Fine-tuning Approach**:

- - Batch size: 12
- - Learning rate: 5e-5
- - Training epochs: 10
- - Maximum sequence length: 64 tokens
- - Weight decay: 0.01
- - Hardware acceleration: Mixed precision (FP16)

Optimization

- - Warmup ratio: 0.1
- Attention dropout: 0.15
- Gradient accumulation steps: 4
- Best model checkpoint saving



Evaluation Results

BLEU Scores

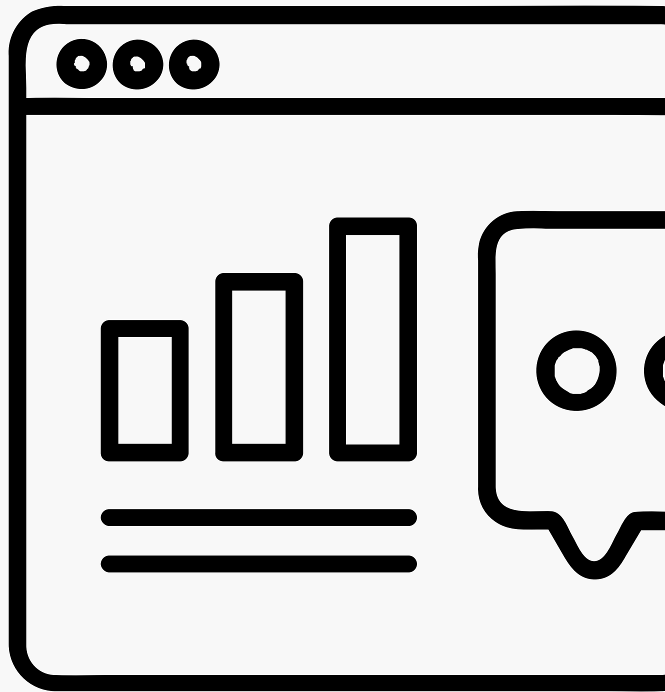
- English → Arabic: 31.39
- Arabic → English: 47.76
- Average: 39.58

Performance Metrics

- Test set size: 10% of dataset
- Evaluation method: SacreBLEU
- Inference parameters: beam size=4, no_repeat_ngram_size=3
- Hardware: GPU-accelerated evaluation

Analysis

- Arabic-to-English performance is notably stronger (+16.37 BLEU)
- Performance asymmetry likely due to linguistic complexity differences
- Model shows competent translation ability in both directions
- Results comparable to models 2-3x larger in size



Sample Translations

English to Arabic Examples:

English	Arabic Translation
"Hello, how are you?"	"مرحبا، كيف حالك؟"
"I love learning languages."	"أنا أحب تعلم اللغات."
"What time is it?"	"كم الساعة الآن؟"

Sample Translations

English to Arabic Examples:

Arabic	English Translation
"الطقس جميل اليوم."	"The weather is nice today."
"اسمي محمد وأنا طالب."	"My name is Mohammed and I am a student."
"أريد السفر إلى مصر العام المقبل."	"I want to travel to Egypt next year."

Performance Optimizations

Training Optimizations:

- Efficient BART base model
- Optimized batch size and learning rate
- Warmup ratio and attention dropout
- Gradient accumulation
- - Mixed precision training

Inference Optimizations**:

- Model.eval() mode
- Half-precision (FP16) inference on GPU
- Enhanced nucleus sampling (top-k/top-p)
- Optimized beam search parameters

Memory Optimizations:

- Low CPU memory usage mode
- TorchScript for GPU acceleration
- torch.no_grad() during inference

GUI Application

Features

- Simple, intuitive interface
- Bidirectional translation selector
- Optimized model loading
- Hardware acceleration when available
- Responsive design

Implementation

- Built with Python's tkinter
- Cross-platform compatibility
- Efficient threading for background processing

Model Limitations

Handling Complex Construction

- Limited capability with long, complex sentences
- Challenges with idiomatic expressions and cultural references
- Lower performance with domain-specific terminology

Arabic-Specific Challenges

- Inconsistent handling of Arabic diacritics
- Less accurate with dialectal Arabic variants
- Challenges with morphologically complex words

Model Limitations

Technical Limitations

- Model size (~140M parameters) trades quality for efficiency
- Maximum sequence length restriction (64 tokens)
- Resource requirements: minimum 4GB GPU memory for inference
- Training data biases from Tatoeba corpu

Performance Gaps

- Asymmetric performance between language directions (English→Arabic: 31.39 BLEU vs Arabic→English: 47.76 BLEU)
- Performance degradation with out-of-domain content

Thank you