

The background is a light gray color with various abstract shapes and icons. In the top left, there are concentric brown circles. In the top center, there is a dark blue wavy shape. In the top right, there is a light gray wavy shape. In the bottom left, there is a light gray wavy shape. In the bottom right, there is a dark blue wavy shape. There are also several small brown speech bubbles: one with three dots in the top left, one with a question mark in the bottom right, and one with a question mark in the bottom right. There are also some small brown dots in the top left and some small brown lines in the top right.

Arabic Extractive Question Answering

Using AraBERTv2 Fine-Tuned on ArTrivia Dataset

Project Overview

- Goal: Build an Arabic extractive QA model to find answer spans from context.
- Approach: Fine-tune AraBERTv2 on ArTrivia dataset.
- Method: Predict start and end tokens for answers within the context.

Dataset – ArTrivia

- Language : Arabic
- Purpose : Extractive Question Answering
- Structure : (Question, Context, Answer)

Training Configuration

- Loss Function: Cross Entropy Loss
- Model: AraBERTv2 (110M parameters approx.)
- Fine-tuned on ArTrivia dataset

Model Details – AraBERTv2

- Model: BertForQuestionAnswering(AraBERTv2)
- Embedding size: 768 | Vocabulary size: 64,000
- Max sequence length: 512 | Dropout: 0.1
- 12 Transformer encoder layers
- Feedforward size: 3072 | Activation: GELU
- Output head: Linear(768 \rightarrow 2) for start/end logits

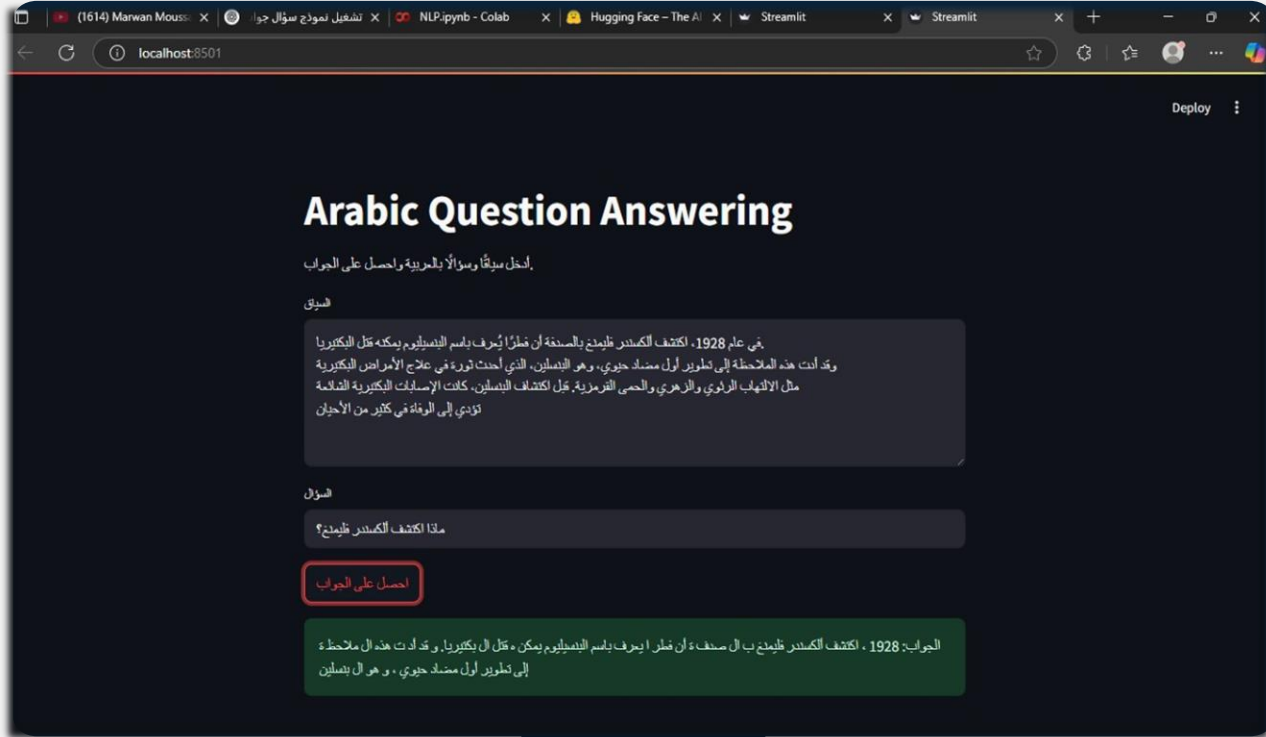
Evaluation (CROSSENTROPY)

- Training Loss: 0.5
- Test Loss: 1.0

Model Limitations

- Weak reasoning capability
- Fails with long or complex contexts
- Sensitive to paraphrased or indirect questions
- Not suitable for multi-hop or implicit answers

Example/Demo





Thank you

