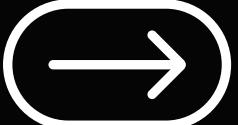




# LEGAL DOCUMENT SUMMARIZATION

BY

- 1-MARTINA MOUSA
- 2-FATMA ELZAHRAA AHMED
- 3-ABDELRAHMAN TAREK
- 4-FASIAL ABDELNASER
- 5-ABDELRAHMAN SALM



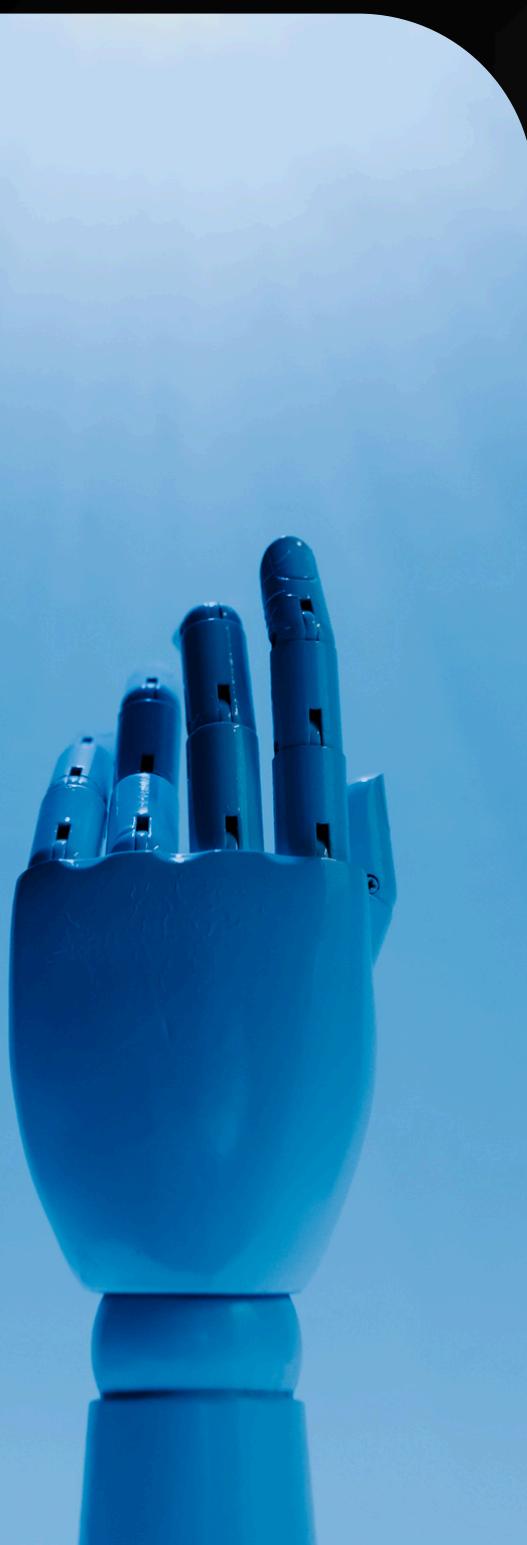
---

# PROJECT OVERVIEW

- This project focuses on enhancing a pre-trained Large Language Model (LLM) for the task of text summarization, specifically in the legal domain.
  - The objective is to produce accurate, concise summaries from lengthy legal texts by fine-tuning the model using Low-Rank Adaptation (LoRA) for efficiency and performance.
-

# DATASET INFORMATION

- SOURCE: LEGAL TEXT DATASET WITH DOCUMENT-SUMMARY PAIRS FROM KAGGLE
- SIZE: [TRAIN:639 , TEST:100]
- STRUCTURE:
  - INPUT: FULL LEGAL DOCUMENT
  - TARGET: HUMAN-WRITTEN SUMMARY
- PREPROCESSING STEPS:
  - TEXT CLEANING (REMOVAL OF HEADERS, SYMBOLS)
  - TRUNCATION FOR LONG DOCUMENTS
  - CONVERSION TO FORMAT: INPUT\_TEXT → TARGET\_TEXT



---

# BASE LLM OVERVIEW

Model Name: facebook/bart-base

Developed By: Facebook AI (Meta AI)

## Architecture Details:

Model Type: Transformer-based Seq2Seq (Encoder-Decoder)

Layers:

6 encoder layers

6 decoder layers

Hidden Size: 768

Attention Heads: 12

Total Parameters: ~139 million

Tokenizer: Byte-Pair Encoding (BPE) from RoBERTa

## Key Features & Capabilities:

Pre-trained on denoising autoencoding tasks (masking and permuted spans).

Strong generalization in text generation tasks.

Performs particularly well on:

Summarization

Machine Translation

Question Answering

Dialogue Systems

Supports long input sequences (up to 1024 tokens)

## Training Details:

Pre-training Corpus: Large-scale datasets (Books, Wikipedia, News)

Pre-training Objective: Reconstruct original text from corrupted input (like a noisy autoencoder)

## Why BART is a Good Base Model:

Combines the strengths of BERT (understanding) and GPT (generation).

Encoder understands context; decoder generates fluent, high-quality summaries.

Suitable for downstream fine-tuning with limited data, especially using techniques like LoRA.



# FINE-TUNING WITH LORA

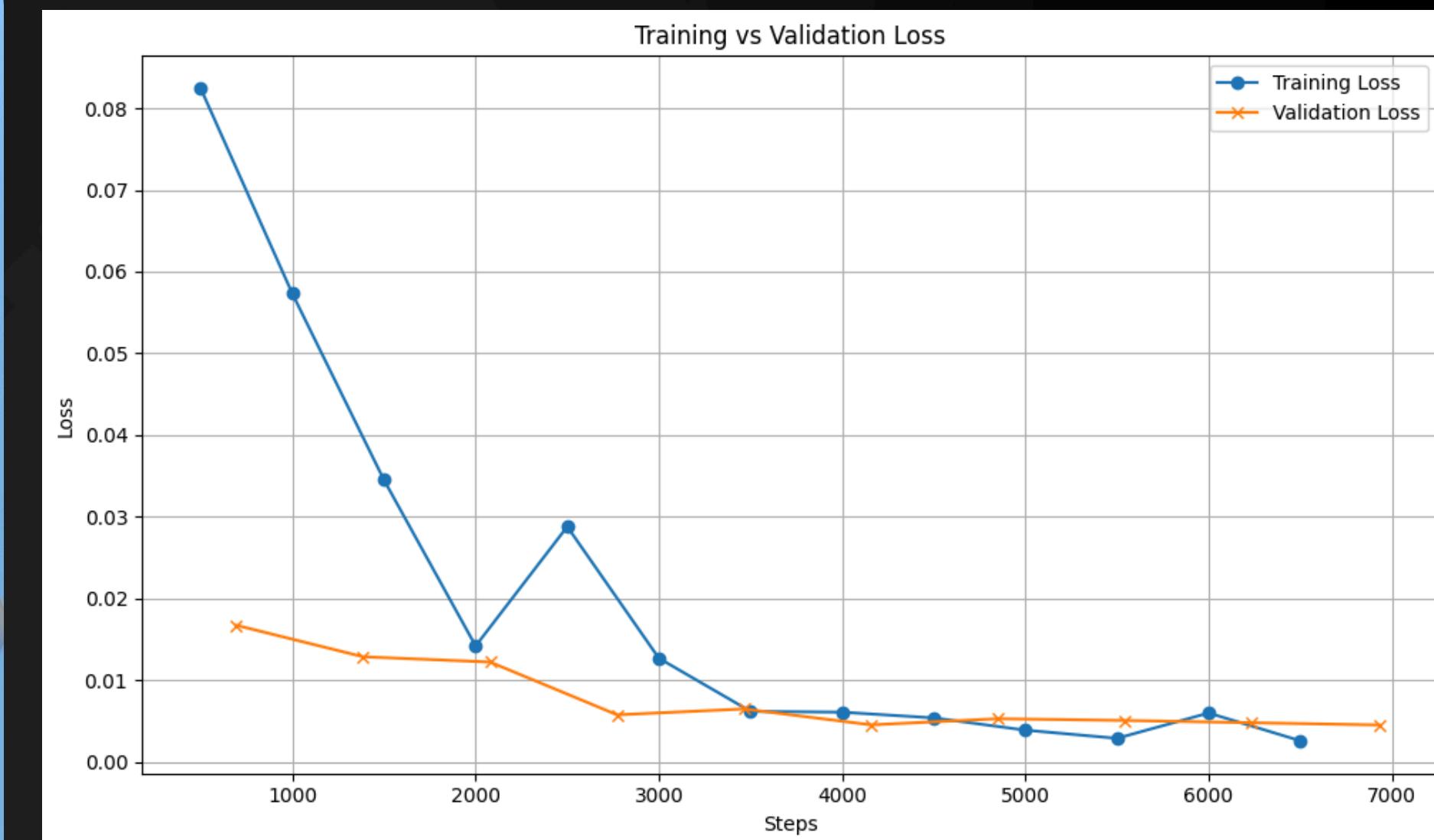
- WHAT IS LORA?
- A PARAMETER-EFFICIENT FINE-TUNING METHOD THAT ADDS LOW-RANK MATRICES TO ATTENTION LAYERS.
- BENEFITS:
  - REDUCES TRAINING MEMORY AND COMPUTE REQUIREMENTS
  - MAINTAINS PERFORMANCE CLOSE TO FULL FINE-TUNING
- IMPLEMENTATION:
- USED HUGGING FACE PEFT LIBRARY TO APPLY LORA TO ATTENTION MODULES WITHIN THE BART MODEL.

# EVALUATION METRICS

- Main Metrics:
  - ROUGE-1, ROUGE-2, ROUGE-L for summary quality
  - Training and validation loss curves
  - BERTSCORE

# PROJECT RESULTS AND PLOTS

## 1- TRAINING LOSS PLOT:



## 2- ROUGE SCORE COMPARISON:

```
Base Model ROUGE Scores:
```

```
{'rouge1': np.float64(0.22036732706317513), 'rouge2': np.float64(0.10629193724448305), 'rougeL': np.float64(0.1320037134002484),
```

```
Fine-Tuned Model ROUGE Scores:
```

```
{'rouge1': np.float64(0.22163033853413955), 'rouge2': np.float64(0.1128370728840656), 'rougeL': np.float64(0.13888007003466285),
```

```
'rougeLsum': np.float64(0.13231020953213518)}
```

```
'rougeLsum': np.float64(0.13908888273398123)}
```

# 3- BERTSCORE



**Base Model BERTScore (P, R, F1):**

P: 0.8576

R: 0.8189

F1: 0.8377

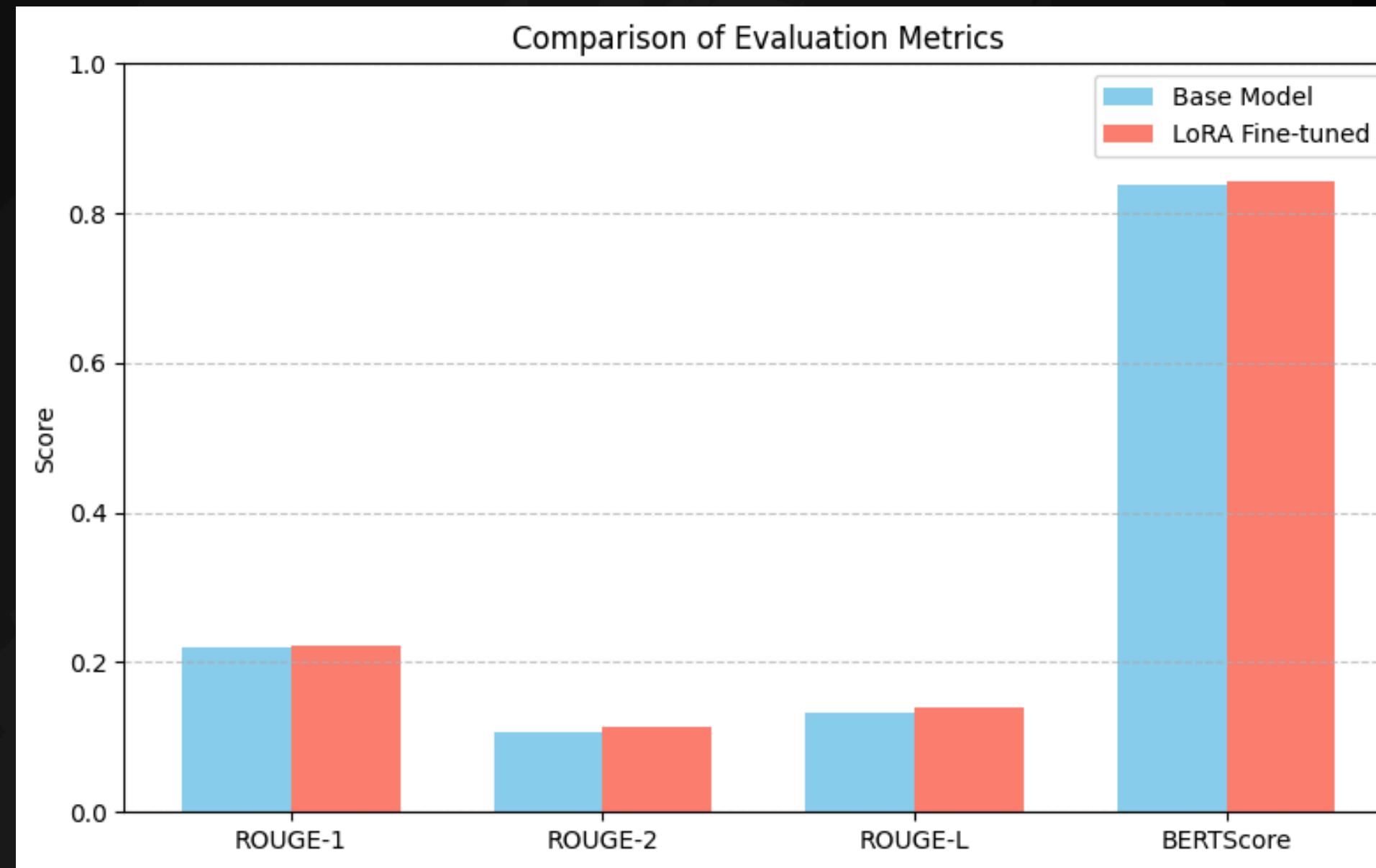
**Fine-Tuned Model BERTScore (P, R, F1):**

P: 0.8628

R: 0.8218

F1: 0.8417

# PLOT THE COMPARISION BETWEEN BASE MODEL AND MODEL FINETUNING



# STRUCTURED OUTPUT SHOWCASE

## Legal Text Summarizer

Provide a legal text to generate a structured summary with each sentence in a new line, along with the time taken for response.

legal\_text

Mr Martin Fowler is a qualified diver, resident in the Republic of South Africa. During the 2011/12 and 2012/13 tax years he undertook diving engagements in the waters of the UK Continental Shelf. Although his status has yet to be determined, the preliminary issue which is the subject of this appeal requires it to be assumed that he undertook those engagements as an employee, rather than as a self employed contractor. HMRC claim, but Mr Fowler denies, that the income which he earned from those diving engagements is subject to UK taxation. That depends on how the double taxation treaty between the UK and South Africa (the Treaty) applies to a person in his position. In a nutshell, the Treaty provides for employment income to be taxed in the place where it is earned, in the present case in the UK, but for the earnings of self employed persons to be taxed only where they are resident, in Mr Fowlers case in South Africa. Thus far the answer might appear to be simple. If (as is to be assumed) Mr Fowler was an employee, then he should be taxable only in the UK. But the matter is complicated by two factors. The first is that employed divers doing the particular kind of diving work in UK waters which Mr Fowler did are, under UK tax law, to be treated as if they were self employed for income tax purposes. The second is that terms used in the Treaty, if not defined in the Treaty itself, are to be given the meaning which they have in the tax law, or the general law, of the state seeking to recover tax, here the UK. Thus, if the effect of the UK tax laws requirement to treat Mr Fowler as if he was self employed is to govern the meaning of relevant terms in the Treaty, the outcome might be that he was to be treated as self employed under the Treaty, and therefore taxable, if at all, in South Africa. This was the conclusion of the majority in the Court of Appeal, from which HMRC appeals to this court. In fact, such an outcome could mean that Mr Fowler was not taxable in either country, because the question whether he was taxable in South Africa would not be governed by the meaning of Treaty terms established by reference to UK tax law. He would probably be treated in South Africa as an employee. To the extent that domestic South African tax legislation did not tax the earnings of residents employed abroad he would not be taxable there or in the UK. There is no general provision in this Treaty as to how to deal with

Mr Martin Fowler is a qualified diver, resident in the Republic of South Africa.

During the 2011/12 and 2012/13 tax years he undertook diving engagements in the waters of the UK Continental Shelf. HMRC claim, but Mr Fowler denies, that the income which he earned from those diving engagements is subject to UK taxation. Although his status has yet to be determined, the preliminary issue which is the subject of this appeal requires it to be assumed that he undertook those engagements as an employee, rather than as a self employed contractor. That depends on how the double taxation treaty between the UK and South Africa (the Treaty) applies.

Time taken for response: 50.97 seconds

Flag

Clear

Submit



**THANKS**