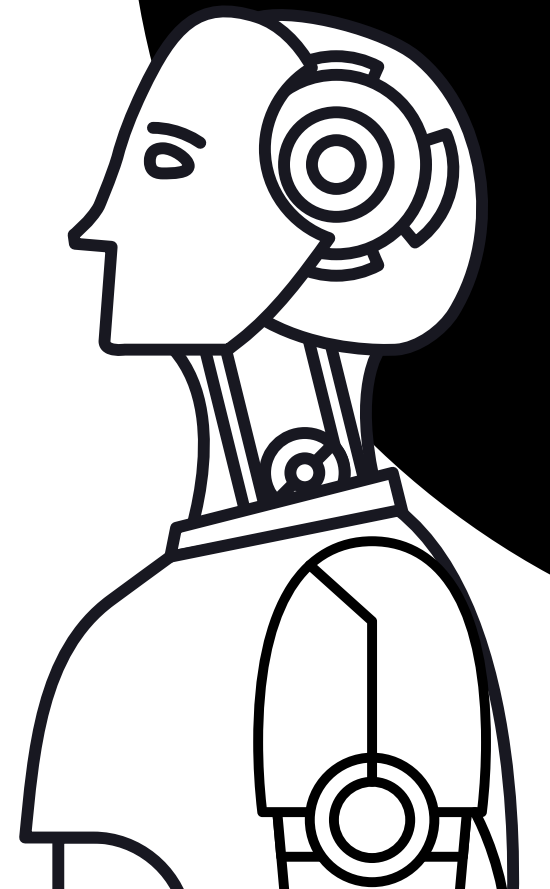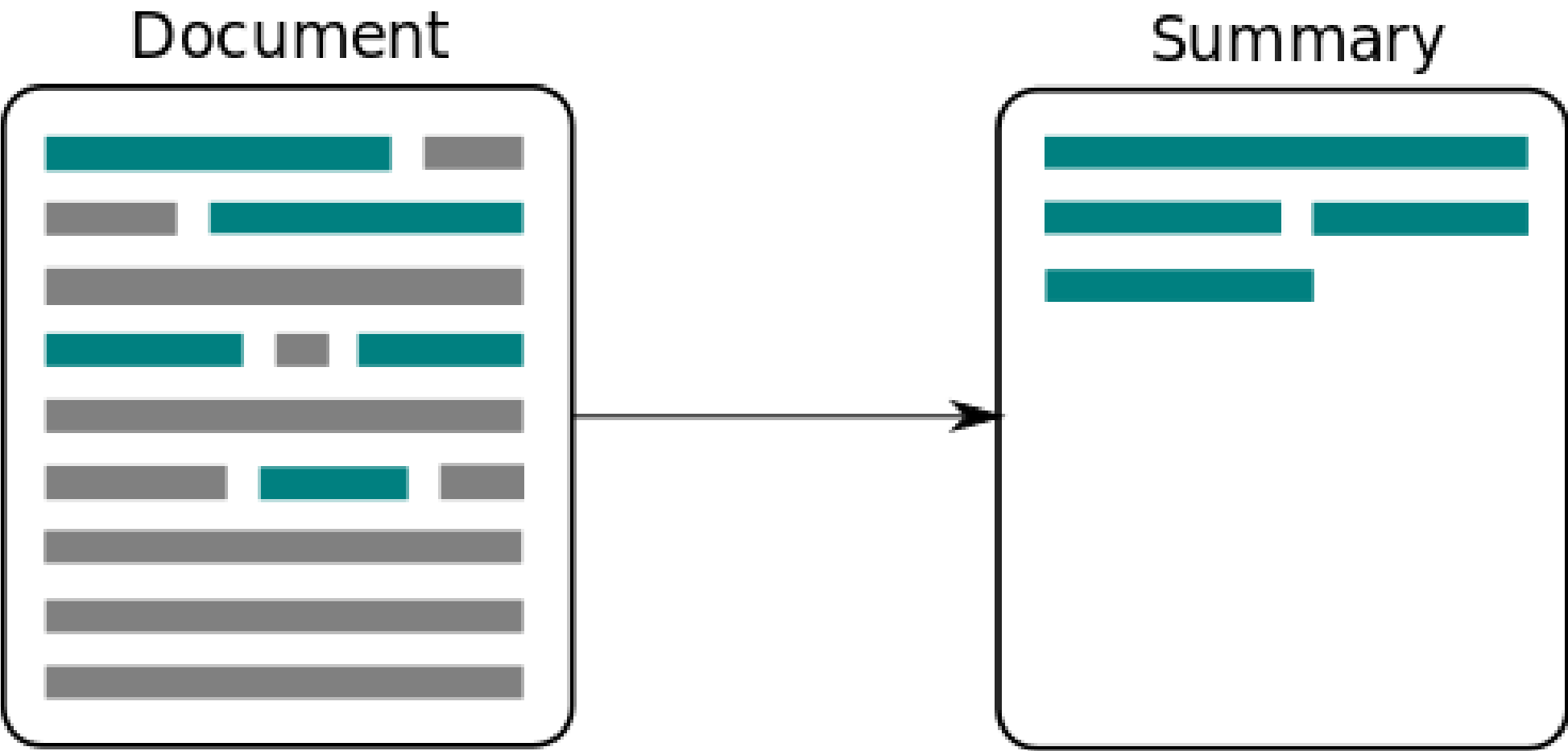# TASK SUMMARIZATION

## PRESENTATION

# WHAT IS TASK SUMMARIZATION ?

Task summarization is the process of generating a brief, clear description of a task or action from a longer input like text, dialogue, or activity logs. It helps identify the key objective or instruction.
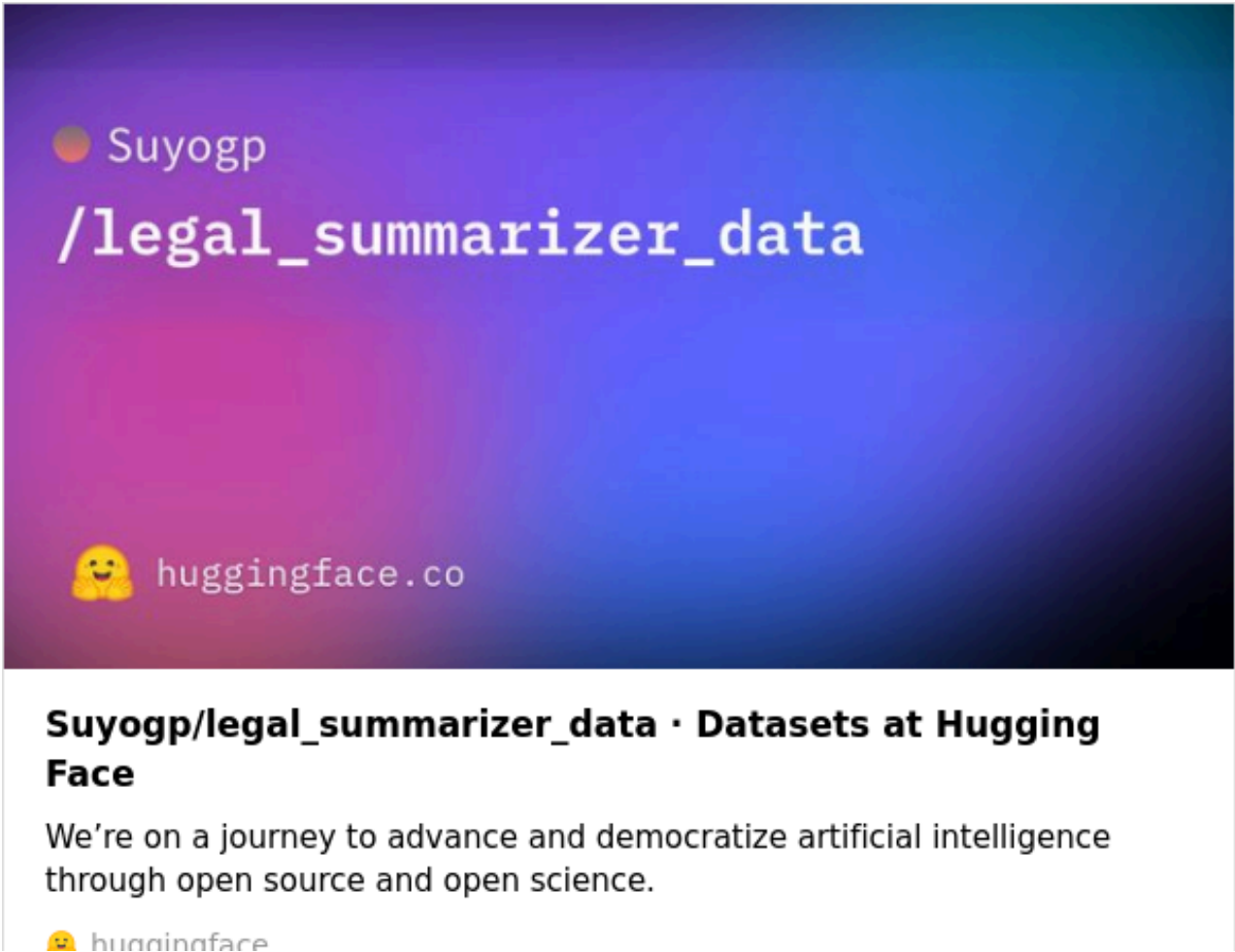


Document

Summary

# WHAT IS OUR DATA SET ?

## LEGAL SUMMARIZER DATASET (SUYOGP/LEGAL_SUMMARIZER_DATA)

- **Total Entries: Approximately 7,026 legal documents**
- **Splits:**
- **Training Set: 5,620 documents**
- **Validation Set: 703 documents**
- **Test Set: 703 documents**
- **Format: Each entry comprises a full legal document paired with its corresponding summary.Language: English**

- **Source: Hugging Face**
- **Created by: Suyog Poudel**
- **Entries: ~7,000 legal documents with summaries**
- **Language: English**
- **Jurisdiction: Indian courts (High Courts & Supreme Court)**
- **Use case: Legal text summarization**

*EACH ENTRY INCLUDES A FULL LEGAL DOCUMENT AND A HUMAN-WRITTEN SUMMARY HIGHLIGHTING KEY LEGAL POINTS AND DECISIONS.*



Suyogp
/legal_summarizer_data

huggingface.co

**Suyogp/legal_summarizer_data · Datasets at Hugging Face**

We're on a journey to advance and democratize artificial intelligence through open source and open science.

huggingface

### EXAMPLE

**01**
- Text: Legal appeal about income tax deductions
- Summary: Focuses on whether certain taxes are deductible under Indian tax law

**02**
Applications: AI legal assistants, legal search tools, legal education, case brief generation.

# DATA PREPROCESSING

```
# Load the dataset
ds = load_dataset("Suyogp/legal_summarizer_data")
train_dataset = ds["train"].filter(lambda x: x["summary"].strip() != "").remove_columns(["id"])
val_dataset   = ds["validation"].filter(lambda x: x["summary"].strip() != "").remove_columns(["id"])
test_dataset  = ds["test"].filter(lambda x: x["summary"].strip() != "").remove_columns(["id"])
```

**Filter and Clean:**

For each split (train, validation, test):

- Removes entries where the "summary" field is empty ("") after stripping whitespace.
- Removes the "id" column since it's likely not needed for training.

**THIS ENSURES ALL DATASETS CONTAIN ONLY MEANINGFUL SUMMARIES AND NO UNNECESSARY ID FIELDS.**



```
ef preprocess_data(examples):
    # Prompt format
    inputs = [f"Summarize the following legal case dialogue:\n{dialogue}" for dialogue in examples["dialogue"]]
    model_inputs = tokenizer(inputs, max_length=512, padding="max_length", truncation=True)

    # Tokenize targets
    with tokenizer.as_target_tokenizer():
        labels = tokenizer(examples["summary"], max_length=128, padding="max_length", truncation=True)

    # Mask pad tokens in the labels
    model_inputs["labels"] = [
        [(token if token != tokenizer.pad_token_id else -100) for token in label]
        for label in labels["input_ids"]
    ]

    return model_inputs
```

**Step 1: Create Prompts**

We format each example into a prompt:

"Summarize the following legal case dialogue: [dialogue text]"

This guides the model to understand the task during training.

**Step 2: Tokenize the Input**

We tokenize the dialogue prompts using a tokenizer:

- Max length: 512 tokens
- Padding: to the maximum length
- Truncation: enabled

This prepares consistent-length input sequences for the mod

**Step 3: Tokenize the Targets (Summaries)**

The target summaries are also tokenized:

- Max length: 128 tokens
- Same padding/truncation settings
- Done in target mode so the tokenizer knows this is the expected output

**Step 4: Mask Padding Tokens**

In the target (labels), padding tokens are replaced with -100:

- This tells the model to ignore them during loss calculation
- Prevents penalizing the model for meaningless padded outputs

# OUR BASE MODEL EXPLANAITON(FLAN T5 BASE MODEL)

🤖 **What is FLAN-T5 Base?**

FLAN-T5 is an enhanced version of the original T5 model, developed by Google.
It stands for Fine-tuned LAnguage Net (FLAN) and is designed to follow instructions better than standard T5.
.

📦 **FLAN-T5 Base at a Glance**

- Model size: ~250 million parameters
- Architecture: Encoder-Decoder (like T5)
- Backbone: Same as T5, but with better instruction-following capabilities
- Pretrained on: C4 (like T5), then fine-tuned on over 1,000 instruction-based tasks

🚀 **What Makes FLAN-T5 Special?**

🧠 It's not just trained to complete tasks — it's trained to understand and follow human instructions.
This makes it highly effective for tasks like:

- ✍️ Text summarization
- ❓ Question answering
- 🔄 Translation
- ✅ Classification

💡 **How FLAN-T5 Works**

Like T5, it treats every NLP task as a text-to-text problem:
Input:  "Summarize: The court said the defendant..."Output:  "The defendant was found guilty."
It uses a prompt + task format and generates the desired output.

✅ **Why Choose FLAN-T5 Base?**

- More accurate and generalizable than vanilla T5
- Better at zero-shot and few-shot learning
- Strong performance even on domain-specific tasks like legal summarization

# 🧠 FLAN-T5 ARCHITECTURE OVERVIEW

**FLAN-T5 IS BASED ON THE ENCODER-DECODER TRANSFORMER ARCHITECTURE, SIMILAR TO THE ORIGINAL T5. IT PROCESSES AND GENERATES TEXT USING TWO MAIN COMPONENTS:**

**◆ 1. ENCODER (LEFT SIDE)**

TAKES IN THE INPUT TEXT (E.G., "SUMMARIZE: THE LEGAL DIALOGUE...")

PASSES IT THROUGH MULTIPLE TRANSFORMER LAYERS

EACH LAYER INCLUDES:

SELF-ATTENTION: HELPS THE MODEL UNDERSTAND RELATIONSHIPS WITHIN THE INPUT

FEED-FORWARD NETWORKS: ADDS DEPTH AND LEARNING POWER

PRODUCES A RICH ENCODED REPRESENTATION OF THE INPUT

**◆ 2. DECODER (RIGHT SIDE)**

GENERATES THE OUTPUT TEXT STEP-BY-STEP (E.G., "THE DEFENDANT WAS FOUND GUILTY.")

USES:

MASKED SELF-ATTENTION: LOOKS ONLY AT PREVIOUSLY GENERATED TOKENS (AUTO-REGRESSIVE)

CROSS-ATTENTION: ATTENDS TO ENCODER'S OUTPUT TO STAY RELEVANT TO THE INPUT

FEED-FORWARD NETWORKS: TRANSFORMS REPRESENTATIONS INTO MEANINGFUL TOKENS

PRODUCES ONE TOKEN AT A TIME UNTIL THE FULL SUMMARY IS COMPLETE

**🔁 TRAINING OBJECTIVE**

THE MODEL IS TRAINED TO PREDICT THE NEXT TOKEN IN THE OUTPUT SEQUENCE, GIVEN THE INPUT AND PREVIOUSLY GENERATED TOKENS.



FLAN-T5

Encoder — Decoder

**Input:** "Summarize: The court said thfendant..."

**Output:** "The defendant was found guilty."
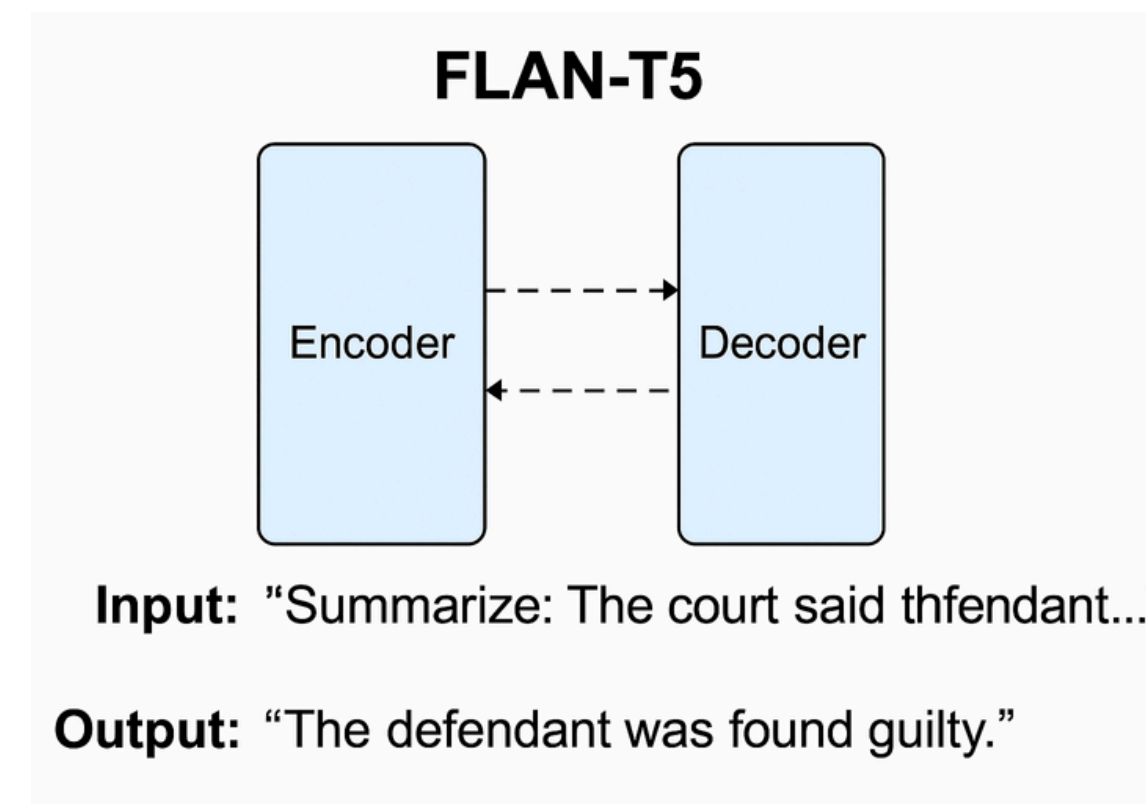
# ✦ WHAT IS LORA (LOW-RANK ADAPTATION)?

LORA IS A LIGHTWEIGHT FINE-TUNING TECHNIQUE THAT ALLOWS US TO ADAPT LARGE LANGUAGE MODELS WITHOUT UPDATING ALL OF THEIR PARAMETERS.

## 💡 WHY USE LORA?

- FULL FINE-TUNING OF MODELS LIKE T5 IS COMPUTATIONALLY EXPENSIVE
- LORA REDUCES COST BY ADDING SMALL TRAINABLE LAYERS INTO THE MODEL
- THE ORIGINAL MODEL WEIGHTS ARE FROZEN, AND ONLY A FEW NEW PARAMETERS ARE TRAINED

## ⚙️ HOW IT WORKS (HIGH-LEVEL)

DECOMPOSITION: LORA INSERTS LOW-RANK MATRICES (A AND B) INTO THE ATTENTION LAYERS

INSTEAD OF UPDATING THE FULL WEIGHT MATRIX, WE LEARN:
$\Delta W = A \times B$ (WHERE A AND B ARE MUCH SMALLER THAN W)

ADAPTATION: DURING TRAINING, THE MODEL LEARNS THESE SMALL MATRICES
MUCH FASTER AND USES LESS MEMORY

INFERENCE: THE BASE MODEL + LEARNED LORA LAYERS = ADAPTED OUTPUT
NO NEED TO RETRAIN THE FULL MODEL AGAIN

## 🧠 HOW WE APPLIED LORA

WE USED FLAN-T5 BASE AS THE FROZEN BACKBONE
LORA WAS APPLIED TO THE ATTENTION LAYERS
ONLY A SMALL NUMBER OF PARAMETERS WERE TRAINED
ACHIEVED EFFECTIVE SUMMARIZATION PERFORMANCE WITH LOW RESOURCE USAGE

## ✅ BENEFITS OF LORA

| Feature | Benefit |
| --- | --- |
| Memory-efficient | Trains fewer parameters |
| Faster fine-tuning | Shorter training time |
| Plug-and-play | Can reuse same base model |
| Cost-effective | Ideal for small-scale |

# 📊 EVALUATION METHODS:

## WE USED TWO EVALUATION METHODS:

🟥 **1. ROUGE SCORE (RECALL-ORIENTED UNDERSTUDY FOR GISTING EVALUATION)**

COMPARES THE OVERLAP OF WORDS OR PHRASES BETWEEN THE GENERATED SUMMARY AND THE REFERENCE (HUMAN-WRITTEN) SUMMARY

**WE USED:**
**ROUGE-1: OVERLAP OF UNIGRAMS (SINGLE WORDS)**
**ROUGE-2: OVERLAP OF BIGRAMS (2-WORD PAIRS)**
**ROUGE-L: LONGEST COMMON SUBSEQUENCE (SENTENCE-LEVEL STRUCTURE)**
✅ **GOOD FOR:** MEASURING HOW MUCH CONTENT FROM THE REFERENCE IS PRESERVED
🚫**LIMITATION:** IGNORES WORD MEANING OR PARAPHRASING

🤖 **2. BERTSCORE**

USES BERT EMBEDDINGS TO EVALUATE THE SEMANTIC SIMILARITY BETWEEN GENERATED AND REFERENCE SUMMARIES COMPUTES PRECISION, RECALL, AND F1 USING CONTEXTUAL MEANING, NOT JUST WORD OVERLAP
✅ **GOOD FOR:** CAPTURING PARAPHRASED CONTENT AND SEMANTIC RELEVANCE
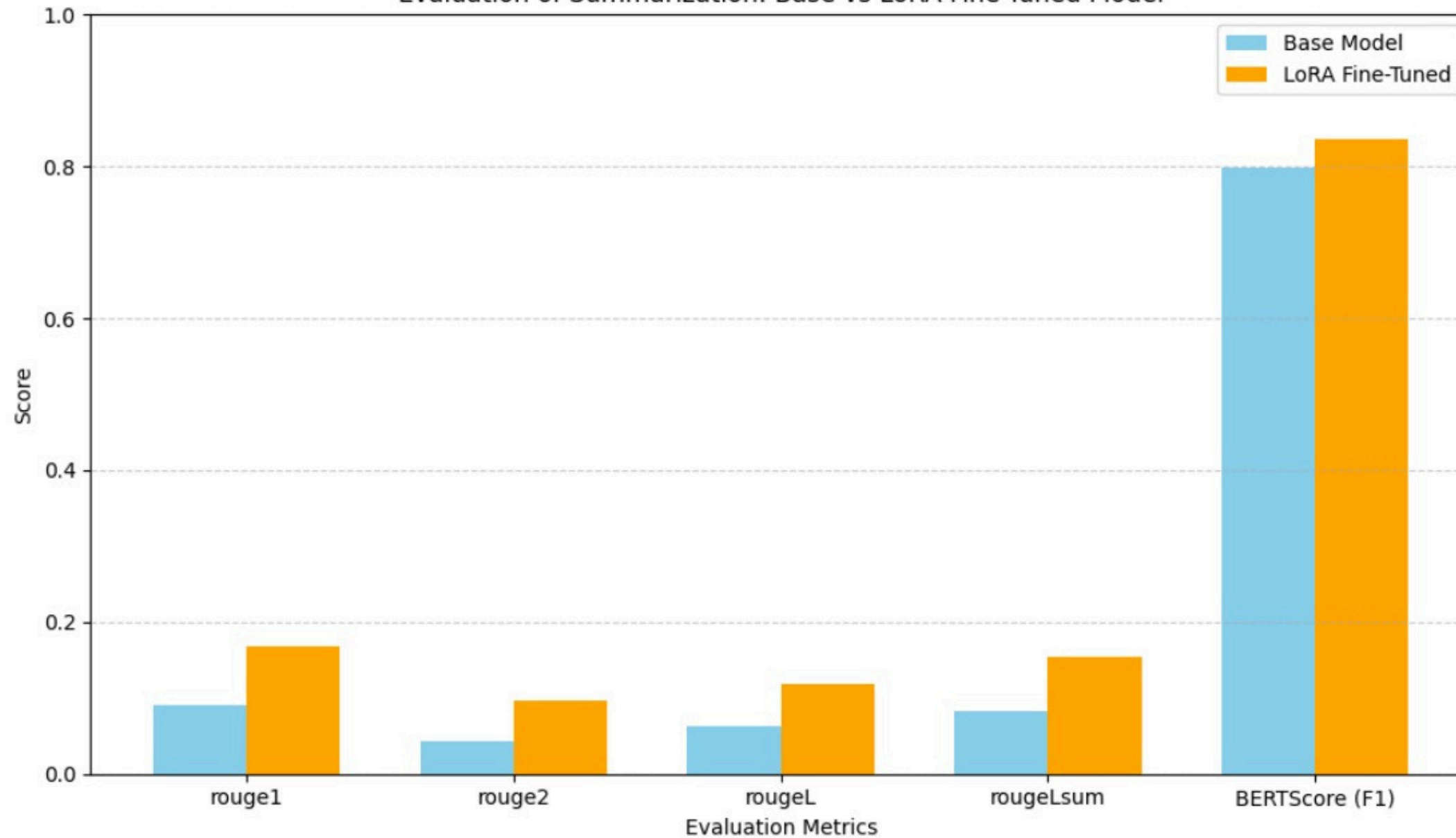🚫 **LIMITATION:** SLOWER TO COMPUTE THAN ROUGE

| Metric | Measures | Strength |
|---|---|---|
| **ROUGE** | Token overlap | Simple and fast |
| **BERTScore** | Semantic similarity | Understands meaning and context |

# PROJECT RESULTS AND PLOTS:



Evaluation of Summarization: Base vs LoRA Fine-Tuned Model

**BASE MODEL BERTSCORE (F1): 0.7988**

**LORA FINE-TUNED BERTSCORE (F1): 0.8352**

## 📈 ROUGE Score Comparison

| Metric | Base Model | Fine-Tuned Model |
|---|---|---|
| ROUGE-1 | 0.0904 | 0.1689 |
| ROUGE-2 | 0.0427 | 0.0958 |
| ROUGE-L | 0.0636 | 0.1183 |
| ROUGE-Lsum | 0.0820 | 0.1540 |

# 📊 STRUCTURED OUTPUT:

| Input | Real Summary |
|---|---|
| ivil Appeal No. 3002 of 1983. From the Judgment and Order dated 20.8.1982 of the Madras High Court in S.A. No. 83 of '1982. G. Ramaswamy (N.P.), Mrs. Anjani and K. Ramkumar for the Appellant. P.S. Poti and K.V. Sreekumar for the Respondent. The Judgment of the Court was delivered by SHARMA, J. This ... | In pursuance of a Housing Scheme the Tamil Nadu Housing Board, Madras had allotted residential plots over the land acquired under the Land Acquisition Act, to different groups of applicants including the low income group on terms and conditions stipulated in the lease deed Exh. B 3 sometime in the y... |
| vil Appeal Nos.1177 to 1184 (NT) of 1990. From the Judgments and Order dated 5.3.85, 21.1.85, 25.2.85, 11.2.85, 14.10.85, 11.2.85 and 20.10.86 of the Madras High Court in T.C. Nos.694/82,565/80, 1404/80, 637/81,638/81,521/81,429/83 and 572/83. T.A. Ramachandran and Mrs. Janki Ramachandran for the Ap... | Section 40(b) of the Income Tax Act, 1961, as it stood at the relevant time, prohibited deduction of interest, salary, bonus, commission or remuneration paid by the firm to the partner. Explanation 1 introduced thereto by the Taxation Laws (Amendment) Act, 1984, which took effect from 1st April, 198... |
| Criminal Appeal Nos. 11 & 12 of 1990. From the Judgment and Order dated 7.9.1989 of the Delhi High Court in Criminal Writ No. 591 & 591 A of 1988. Harjinder Singh, R.N. Joshi and Latha Krishnamurthy for the Appellants. Soli J. Sorabjee, Attorney General, P. Parameswaran, B. Parthasarthy, N.N. Johari... | The appellants were arrested for offences punishable under the provisions of the Narcotic Drugs and Psychotropic Substances Act, 1988. They were remanded to judicial custody till October 13, 1988. On October 11, 1988 orders were passed under Section 3(1) of the Prevention of illicit Traffic in Narco... |
| ivil Appeal No. 2422 of 1989. From the Judgment and Order dated 3.9.1988 of the Bombay High Court in Appeal from Order No. 707 of 1987. Soli J. Sorabjee, R.F. Nariman, Raian Karanjawala, Ms. Meenakshi Arora, Ms. Nandini Gore and Ms. Manik Karanjawala for the Appellant. Anil Diwan, Harish N. Salve, M... | The appellant along with his father and mother, were the joint owners of the suit property. After the death of the appellant 's mother, he and his father executed an agreement dated 23rd August, 1951 by which they severed their status as joint owners and agreed to hold the property as tenants in com... |
| No. 975 of 1986. (Under Article 32 of the Constitution of India). D.D. Thakur, V.C. Mahajan, section Markandaya, G.S. Rao, Sreepal Singh and Ms. Kusum Chowdhary for the Petitioners. R N. Trivedi, S.C. Batra and Raju Ramachandran for the Respondents. The Judgment of the Court was delivered by RANGANA... | Certain cooperative housing societies comprising of the petitioners and others had acquired lands in the trans Jamuna area of Uttar Pradesh prior to the setting up of the New Okhla Industrial Development Authority in 1976. When the said lands came to be notified for the Development Authority writ pe... |

| Base Model Summary | LoRA Fine-Tuned Summary |
|---|---|
| a fixed price. | The appellant Board settled a large number of residential plots to different groups of applicants including one described as low income group. The appellant Board proceeded to settle a large number of residential plots to different groups of applicants including one described as low income group. A number of allottees, including the plaintiff respondent, were selected and settlement in their favour was made in 1963. A copy of the document executed separately in respect to the plots is on the record of this case as Exh. B 3, setting out the terms and conditions of the lease. The term as mentioned in the 15th clause |
| A sharp divergence of juris cial opinion in the High Courts. | In Tax Case Nos. 694 of 1982, 565 of 1980, 1404 of 1980, 637 and 638 of 1981, 521 of 1981, 429 of 1983 and 572 of 1983, the High Court answered the question of law, similar in all the cases, in favour of the revenue. The question was whether in making a disallowance for the inter est paid by a partnership firm to a partner under Section 40(b) of the Act the interest, in turn, paid by the partner on his borrowings from the firm should be taken account of and deducted and only |
| The writ petitions filed under Article 226 of the Constitution to challenge the legality of the orders dated October 11, 1988 passed under Section 3(1) of the Prevention of Illicit Traffic in Narcotic Drugs and Psychotropic Substances Act, 1988 (hereinafter referred to as 'the Act ') for the detention of the appellants have been dismissed. This Court by the order dated January 11, 1990 allowed the appeals and after setting aside the orders of detention dated October 11, 1988 directed that the appellants be set at liberty forthwith and that reasone | The writ petitions filed under Article 226 of the Constitution to challenge the legality of the orders dated October 11, 1988 passed under Section 3(1) of the Prevention of Illicit Traffic in Narcotic Drugs and Psychotropic Substances Act, 1988 (hereinafter referred to as 'the Act ') for the deten tion of the appellants have been dismissed. This Court by the order dated January 11, 1990 allowed the appeals and after setting aside the orders of detention dated October 11, 1988 directed that the appellants be set at liberty forthwith and that |
| ivil Appeal No. 2422 of 1989. From the Judgment and Order dated 3.9.1988 of the Bombay High Court in Appeal from Order No. 707 of 1987. Soli J. Sorabjee, R.F. Nariman, Raian Karanjawala, Ms. Meenakshi Arora, Ms. Nandini Gore and Ms. Manik Karanjawala for the Appellant. Anil Diwan, Harish N. Salve, Ms. Indu Malhotra, Mrs | The appellant is the owner of an undivided half share in the suit property. The suit property was purchased originally under a deed dated 12th January, 1934 by Cawasji Dorabji Warden, Banubai Warden and the appellant as joint owners. Cawasji Dorabji Warden and Banubai Warden are respectively the father and mother of the appellant. It appears that the super structure on the land was constructed subsequent to the purchase. At the time when the property was purchase the appellant was a minor. By |