Assignment: Fine-Tuning Qwen2.5 Model with LoRA on Alpaca-Cleaned Dataset in Google Colab

---

1. Exercise: Fine-Tuning Qwen2.5 Model with LoRA

Background:
A team working on a language model improvement project aims to enhance the performance of the Qwen2.5 model by fine-tuning it on instruction-following tasks. The team will use the Alpaca-Cleaned dataset to fine-tune the model, applying LoRA (Low-Rank Adaptation) for efficient parameter updates. This assignment will help you learn how to fine-tune a model in Google Colab using LoRA, providing you with the opportunity to experiment with a large-scale language model and optimize it for instruction-following tasks.

Type of Data:

- **Structured text** in tabular format containing instruction-output pairs from the Alpaca-Cleaned dataset.

Data Split:

- The dataset is randomly split into a **80% training** set and a **20% test** set. You will train the model on the training data and evaluate its performance using the unseen test data.

Link to Data:

- The Alpaca-Cleaned dataset is available for access on Hugging Face.

---

2. Expected Result:

A solution that demonstrates the fine-tuning of a Qwen2.5 model using LoRA to improve its ability to follow instructions and generate contextually relevant outputs.

---

3. Libraries to Use:

- **Hugging Face** pipelines
- **Langchain**
- **LlamaIndex**
- **LoRA** (Low-Rank Adaptation for model fine-tuning)

- **Unsloth** (For efficient model loading and fine-tuning)
- **PyTorch** and **Transformers** libraries

Table Extraction Packages:

- **Tabula** or **Camelot** (in case table data extraction is needed)

Open Source Models:

- **Qwen2.5**, **Dolly**, **Falcon**, **Llama**, **T5**, **BART**, **Flan-T5**, **Pegasus**

---

4. Deliverables:

- **Google Colab Notebook**: Upload the complete notebook containing all the code and instructions for the assignment.
- **Report**: A text or doc file containing:
    - **Section 1: Model Details**
        - Model Name: Qwen2.5
        - Number of Parameters: (example: 0.5B)
        - Fine-Tuning Method: LoRA (Low-Rank Adaptation)
    - **Section 2: Experiment Results**
        - Prompt used for generation
        - Example Results (Five examples of model outputs)
        - Metrics (BLEU, ROUGE, Semantic Match Scores)

Report Example:

- **Model Details**:
    - Model: Qwen2.5
    - Parameters: 0.5B
    - Fine-Tuning Method: LoRA
    - Libraries Used: Hugging Face, Langchain, LlamaIndex
- **Experiment Results**:
    - **Prompt Used**: "### Instruction: {instruction} ### Input: {input} ### Response:"
    - Example Results:

1. **Instruction**: "Translate the following text to French."
   Input: "Hello, how are you?"
   Output: "Bonjour, comment ça va?"
2. (Additional examples here)

- Metrics to Meet:
  - **BLEU**: 0.75
  - **ROUGE**: 0.80
  - **Semantic Match**: 0.85

- **Upload Instructions**:
  - Upload the Colab notebook and the report to GitHub in the folder **"Assignment 5"**.