

- **Provide a detailed description of the dataset you plan to use for fine-tuning? Please include the number of instances and the schema.**

An Arabic - English language pair, the dataset has 2 columns, source and target. Where the source is an English language text and target is the classical Arabic translation, the number of rows currently is around 90K but could be sampled down (did not explore that yet). <https://huggingface.co/datasets/Abdulmohsena/Classic-Arabic-English-Language-Pairs>

- **What specific type of fine-tuning do you plan to do (e.g., Full Fine-Tuning, LoRA, QLoRA, etc.)? Please explain your choice.**

Depends on the model's internal structure. But generally, we would use QLoRA to ensure the optimal translation time and effectiveness. We may also use knowledge distillation in order to ensure smaller model and quicker training/inference time.

- **How many GPU hours do you estimate will be required for your fine-tuning process?**

I would estimate it would take around 80 - 150 GPU hours to ensure we provide a complete product on a high-end GPU.

- **Describe your previous experience with fine-tuning? Have you worked with the specific method you're proposing before?**

Yes, I have fine-tuned facebook's NLLB200 model with 600M parameters, I fine-tuned it for the same goal we are proposing now, and it gave promising results

- **How do you plan to evaluate the effectiveness of the fine-tuning? What metrics or benchmarks will you use to determine success?**

BERTScore; as it is the main benchmark for contextual translation, which is our goal.

- **Is your dataset already pre-processed and ready for fine-tuning? what additional steps are required to prepare your data?**

the dataset is ready for training, the only preprocessing step remaining is tokenization.

- **Are there any specific constraints or limitations (e.g., time, computational resources) that we should be aware of regarding your fine-tuning request?**

we believe time is important because the LLM would take much time to train and evaluate.

