AI Optimizer Tech Exercises

Homework Assignment

Objective

Fine-tune a Large Language Model to improve its performance on question-answering tasks. This assignment involves evaluating the model before and after fine-tuning, enhancing its ability to provide high-quality answers based on the provided dataset.

Dataset Description

The dataset to be used is the **SQuAD (Stanford Question Answering Dataset)** available on Hugging Face: SQuAD Dataset. The task is to fine-tune a small LLM to improve its accuracy in answering these questions.

Assignment Details

1. **Pre-training Evaluation**

- **Objective**: Assess the initial performance of the LLM before any fine-tuning.
- **Instructions**:
  - Select a small pre-trained language model suitable for running on a GPU.
  - Evaluate the model's performance on the SQuAD dataset by answering a subset of questions. Record the model's answers and compute baseline metrics of your choice.

2. **Fine-Tuning the Language Model**

- **Objective**: Fine-tune the selected language model on the SQuAD dataset to improve its question-answering capabilities.
- **Instructions**:
  - Preprocess the SQuAD dataset to convert it into a format suitable for fine-tuning
  - Fine-tune the language model using GPU resources.
    Note: you may use Google Colab, as it provides free access to GPU resources.
  - Monitor training metrics and adjust hyperparameters as needed to avoid overfitting.

3. **Post-training Evaluation**

- **Objective**: Evaluate the fine-tuned model to measure improvement in question-answering performance.
- **Instructions**:
  - Re-evaluate the fine-tuned model on the same subset of the SQuAD dataset used in the pre-training evaluation.
  - Compare the post-training results with the baseline results, focusing on improvements in metrics
  - Document the differences in performance, providing insights into the types of questions or answers that saw the most significant improvement.

Submission Requirements

- Provide a GitHub repository link containing all the required code, scripts, and documentation.
- A brief report in the repository that outlines your approach, results, key findings and any challenges encountered during the task.