# **Fine-Tuning Documentation for "01-ai/Yi-1.5-9B-Chat" Model on "mlabonne/guanaco-llama2-1k" Dataset**

## **Overview**

This document details the fine-tuning process for the "01-ai/Yi-1.5-9B-Chat" model using the "mlabonne/guanaco-llama2-1k" dataset. The process includes the configuration of the model, dataset description, parameters used, hardware specifications, and training metrics.

## **Model Description**

### **Model: 01-ai/Yi-1.5-9B-Chat**

* **Type**: Causal Language Model
* **Parameters**: 1.5 Billion
* **Initial Configuration**:
  + **Quantization**: 4-bit precision
  + **LoRA attention dimension (r)**: 64
  + **LoRA alpha**: 16
  + **LoRA dropout**: 0.1

### **Dataset: mlabonne/guanaco-llama2-1k**

* **Source**: Hugging Face Dataset
* **Description**: A dataset specifically curated for fine-tuning language models with instructions related to LLaMA (Large Language Model Meta AI) versions.
* **Split Used**: Training split

## **Hardware Specifications**

* **GPU**: L4
* **GPU Memory**: 22 GB
* **Disk Space**: 50 GB

## **Training Configuration**

### **Parameters**

* **Number of Training Epochs**: 1
* **Batch Size per GPU for Training**: 4
* **Batch Size per GPU for Evaluation**: 4
* **Gradient Accumulation Steps**: 1
* **Gradient Checkpointing**: Enabled
* **Maximum Gradient Norm**: 0.3
* **Learning Rate (AdamW optimizer)**: 2e-4
* **Weight Decay**: 0.001
* **Optimizer**: Paged AdamW 32-bit
* **Learning Rate Scheduler**: Constant
* **Maximum Steps**: -1 (use number of epochs)
* **Warmup Ratio**: 0.03
* **Group by Length**: True
* **Save Steps**: Every 25 steps
* **Log Steps**: Every 25 steps
* **Reporting**: Tensorboard

### **Tokenizer Configuration**

* **Special Tokens**:
  + Added [PAD] token
  + Set padding token to EOS token
  + Padding side set to right

### **LoRA Configuration**

* **LoRA Alpha**: 16
* **LoRA Dropout**: 0.1
* **Attention Dimension (r)**: 64
* **Bias**: None
* **Task Type**: CAUSAL\_LM

## **Training Metrics**

* **Total Training Time**: 1257.13 seconds (approximately 0.35 hours)
* **Training Loss at Various Steps**:
  + **Step 25**: 1.271400
  + **Step 50**: 1.499100
  + **Step 75**: 1.210700
  + **Step 100**: 1.304300
  + **Step 125**: 1.139700
  + **Step 150**: 1.321200
  + **Step 175**: 1.110600
  + **Step 200**: 1.382400
  + **Step 225**: 1.122700
  + **Step 250**: 1.399000