RLHF Training / Modeling

A Text-based Architectural Flow

And Comprehensive Outline

┌──────────────────────────────────────────────────┐

│ Data Preparation │

│ - Tokenization & Padding │

│ - Truncation for Consistency │

│ - Dataset Construction │

└──────────────────────────────────────────────────┘

│

▼

┌──────────────────────────────────────────────────┐

│ LoRA Rank Determination │

│ - SVD on Hidden State Representations │

│ - Compute Effective Rank │

│ - Select Adaptive Low-Rank Matrices │

└──────────────────────────────────────────────────┘

│

▼

┌──────────────────────────────────────────────────┐

│ Model Preparation │

│ - Configure LoRA Adapter │

│ - Enable PEFT Model Wrapping │

│ - Adjust Model Configuration for Compatibility │

│ - Handle JSON Serialization Issues │

└──────────────────────────────────────────────────┘

│

▼

┌──────────────────────────────────────────────────┐

│ Training Phase │

│ - Define Training Arguments │

│ - Select Reward Model or Traditional Model │

│ - Initialize Trainer (RewardTrainer / Trainer) │

│ - Fine-tune with LoRA │

└──────────────────────────────────────────────────┘

│

▼

┌──────────────────────────────────────────────────┐

│ Reinforcement Learning Integration │

│ - Train PPO-Based Reward Model │

│ - Fine-tune Using Reward Feedback │

│ - Optimize Policy Learning │

└──────────────────────────────────────────────────┘

│

▼

┌──────────────────────────────────────────────────┐

│ Model Upload & Deployment │

│ - Save Model Locally │

│ - Upload Model to Hugging Face Hub │

│ - Ensure Compatibility for Future Inference │

└──────────────────────────────────────────────────┘

## **Detailed Component Breakdown**

### **1. Data Preparation**

* **Tokenizer Function** processes dataset entries for **input\_ids, attention\_mask, chosen/rejected responses**.
* Applies **padding & truncation** to ensure **uniform sequence lengths** (mitigates memory & indexing issues).
* Constructs tokenized dataset for structured processing.

### **2. LoRA Rank Determination**

* Uses **Singular Value Decomposition (SVD)** on hidden state matrices to compute **LoRA rank**.
* Determines the optimal **low-rank representation**, adjusting model compression efficiently.
* Computes energy variance and selects the smallest **number of singular values** exceeding a predefined variance threshold.

### **3. Model Preparation**

* Wraps the base model with **PEFT LoRA configuration** (get\_peft\_model).
* **Handles serialization issues**, preventing JSON encoding errors.
* **Adjusts model configuration** to ensure compatibility with fine-tuning.

### **4. Training Phase**

* **Defines training arguments** (epochs, batch size, warmup steps, weight decay).
* Selects **RewardTrainer for reinforcement learning models** or standard Trainer for traditional fine-tuning.
* Executes **fine-tuning based on LoRA-enhanced model architecture**.

### **5. Reinforcement Learning Integration**

* Initializes **PPOTrainer** for reinforcement learning.
* Uses **reward-based feedback** during fine-tuning to optimize model behavior.
* Enhances **policy learning by adjusting reward configurations**.

### **6. Usage**

* from Modeling import Modeling
* Modeling = Modeling(<training data file location>, True|False)
* Training data is a list of lists where each inner list consists of 3 elements.
  + A prompt, e.g. “Tell me about machine learning.”
  + The user’s chosen response for the input prompt.
  + The user’s rejected response for the input prompt.

### **7. Example Training Data (stored in a file)**

[

[

"A Study of Atezolizumab (MPDL3280A) Compared With a Platinum Agent (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Participants With Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research on Atezolizumab (MPDL3280A) Versus Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC Patients [IMpower110]",

"Comparison of Atezolizumab (MPDL3280A) and Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"A Study of Atezolizumab (MPDL3280A) Versus Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research Comparing Atezolizumab (MPDL3280A) and Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study on Atezolizumab (MPDL3280A) Compared to Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"Atezolizumab (MPDL3280A) Study Compared With Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study of Atezolizumab (MPDL3280A) Versus Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research on Atezolizumab (MPDL3280A) Compared to Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"Comparison Study of Atezolizumab (MPDL3280A) and Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Atezolizumab (MPDL3280A) Versus Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study Comparing Atezolizumab (MPDL3280A) and Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"Atezolizumab (MPDL3280A) Study Versus Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research Comparing Atezolizumab (MPDL3280A) and Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study of Atezolizumab (MPDL3280A) Compared With Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"A Study of Atezolizumab (MPDL3280A) Versus Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research on Atezolizumab (MPDL3280A) Compared to Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Comparison of Atezolizumab (MPDL3280A) and Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"Atezolizumab (MPDL3280A) Study Compared With Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study of Atezolizumab (MPDL3280A) Versus Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research on Atezolizumab (MPDL3280A) Compared to Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"Comparison Study of Atezolizumab (MPDL3280A) and Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Atezolizumab (MPDL3280A) Versus Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study Comparing Atezolizumab (MPDL3280A) and Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

],

[

"Atezolizumab (MPDL3280A) Study Versus Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Research Comparing Atezolizumab (MPDL3280A) and Platinum-Based Chemotherapy (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]",

"Study of Atezolizumab (MPDL3280A) Compared With Platinum Agents (Cisplatin or Carboplatin) + (Pemetrexed or Gemcitabine) in Stage IV Non-Squamous or Squamous NSCLC [IMpower110]"

]

]