



Langscape 3.0 Prompt-tuning

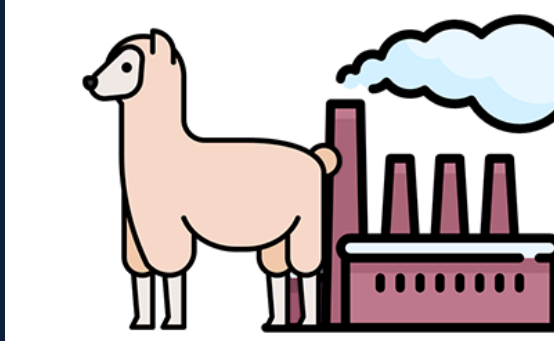
Student: Alexis Loriga

Mentor: *Steven Luis (Product Owner)*

Instructor/Faculty: Dr. Masoud Sadjadi, Florida International University

PROMPT-TUNING TOOLKIT

Mixtral 8x7B by hiyouga
with apache license 2.0



LLaMA-Factory
Easy and Efficient LLM Fine-Tuning

SUMMARY OF WORK

With the updated Langscape 3.0 changed from GPT 4 to Mixtral 7B, prompt tuning is necessary to provide a more accurate and friendlier response to the user. Thus, allowing the user to more easily interact with the AI in a way that is more welcoming and easier to do.

Stage The stage to perform in training. Supervised Fine ▾	Data dir Path to the data directory. data	Dataset 	Preview dataset	
Learning rate Initial learning rate for AdamW. 5e-5	Epochs Total number of training epochs to perform. 3.0	Maximum gradient norm Norm for gradient clipping. 1.0	Max samples Maximum samples per dataset. 100000	Compute type Whether to use mixed precision training. fp16 ▾
Cutoff length Max tokens in input sequence. 1024	Batch size Number of samples processed on each GPU. 2	Gradient accumulation Number of steps for gradient accumulation. 8	Val size Proportion of data in the dev set. 0	LR scheduler Name of the learning rate scheduler. cosine ▾
Extra configurations				

FUTURE OBJECTIVES

- More fine-tuning is possibly with different LLMS or prompt-tuning toolkits
- More filtering out bad results

FINE TUNING MODEL TEST

- Utilized hyperparameters for filtering bad results
- Model is fine-tuned until results are more accurate
- Best version is implemented to being tested in the Langscape 3.0 environment

ACKNOWLEDGEMENT

The material presented in this poster is based upon the work supported by Steven. We/I thank Steven Luis for his assistance and mentorship that I/we received throughout the senior design project.