

CS 6120 NLP

Parameter Efficient Fine-Tuning of LLMs towards ASAG

...

Team #6
Gugan
Anson
Aditya
Prabhleen



Write me a short **Introduction**



NU-GPT

Sure! Here's a brief introduction for your project...

- This project harnesses Instruction - Parameter Efficient Fine Tuning (PEFT) for LLMs to create a packaged solution for Automated Short Answer Grading (ASAG) on Computer Science assignments.

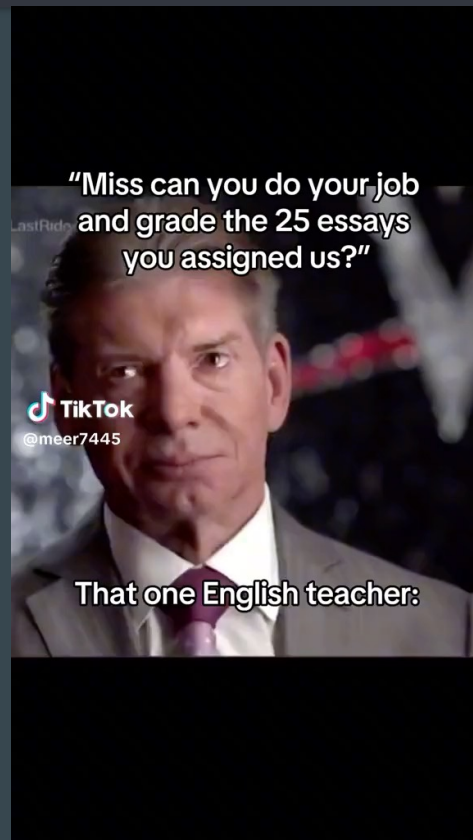


Why work on ASAG?



NU-GPT

- Efficient & Consistent grading
- Instant objective Feedback
- Time-saving for educators
- Scalability for large classes
 - Less TA's to hire (Sorry)



Why fine tune LLMs?



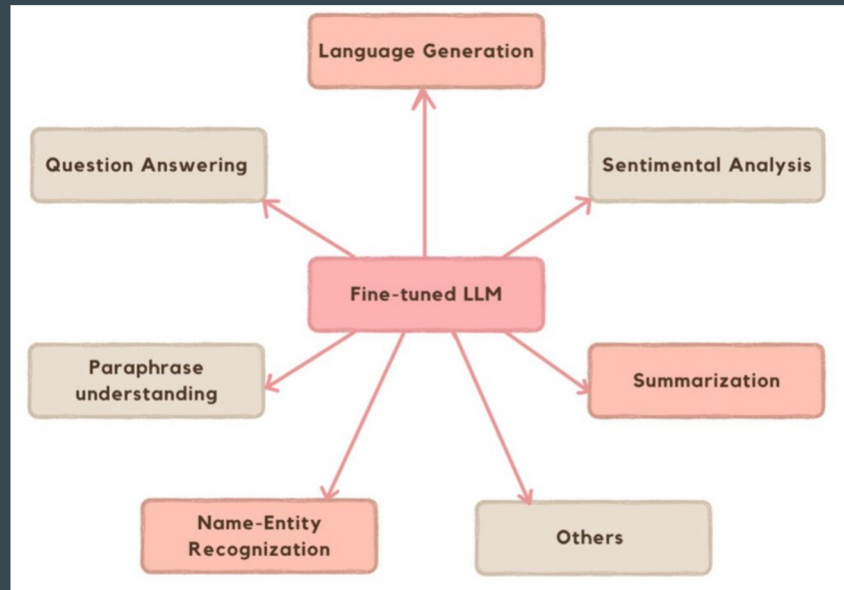
NU-GPT

- Why not traditional NLP methods?
- LLMs offer a vast knowledge base of natural language
- Fine Tuning facilitates domain adaptation
- Fine tuning (PEFT) is faster and cost efficient

LLM



Traditional NLP



Why not existing LLMs / prompt engineering?

Google Flan - T5 - False Positive

You are a grader for a undergraduate computer science course. For the given 'Question', analyze the 'Given Answer' against the 'Expected Answer' and provide a score based on the relevancy of the answer.

'Question': What is the role of a prototype program in problem solving?

'Given Answer': I dont know

'Expected Answer': To simulate the behaviour of portions of the desired software product.

Provide your score on the given scale. In order of least relevant answer to most relevant, the score can be one of the following:

'Incorrect', 'Poor', 'Fair', 'Adequate', 'Good', 'Excellent'. Do not display any other content other than the single score word."

Compute

⌘+Enter

0.6

Computation time on cpu: 0.640 s

Good

Gemma 2b - False Negative



You are a grader for a undergraduate computer science course. For the given 'Question', analyze the 'Given Answer' against the 'Expected Answer' and provide a score based on the relevancy of the answer.

'Question': What is the role of a prototype program in problem solving?

'Given Answer': High risk problems are addressed by a prototype program. A prototype program may also be used to show a company how a product will be developed.

'Expected Answer': To simulate the behaviour of portions of the desired software product.

Provide your score on the given scale. In order of least relevant answer to most relevant, the score can be one of the following: 'Incorrect', 'Poor', 'Fair', 'Adequate', 'Good', 'Excellent'. Do not display any other content other than the single score word."



Adequate

Benchmark Metrics to evaluate performance

Weighted Metric/Model	Encoder only	Encoder - Decoder			Decoder only
	BERT	Flan - T5 base			Gemma
	SVM Classifier	No Fine Tuning	Full Fine Tuning	LoRA Fine Tuning	LoRA Fine Tuning
# of Params	110 M	248 M			2 B
F1 Score	68.15%	0.03%	77.52%	71.74%	90.94%
Precision	68.06%	0.01%	78.29%	72.15%	92.34%
Recall	69.59%	1.29%	77.81%	72.57%	90.87%

- Optimal LoRA parameters tuning (target modules, rank, alpha, etc.) was performed.

Any **Key Takeaways?**



NU-GPT

- Flan T5 Base
 - Full Fine tuning (77% > base model)
 - LoRA fine tuning is close (6% < Full fine tuning)
 - LoRA trains in 1/3rd the time
- Gemma 2B
 - Best performance (13% > FlanT5, 22% > BERT)
 - More params = more knowledge?
 - Costs 30 x Flan T5 / BERT



Future steps?



NU-GPT

- Extend to multimodal applications
- Package as an integration / interactive UI for ASAG.

Question: What is the role of a prototype program in problem solving?

Answer: High risk problems are address in the prototype to make sure the program is feasible. A prototype may also be used to show a company that the software can be possibly programmed.

True Score: Good

Predicted Score: <pad> Good</s>

Thanks for
listening!