Imperial College London

Department of Mechanical Engineering

ME4 Individual Project Final Report

Small LLMs for educational purposes

Author: Zhuangfei Gao

Supervisor: Dr Peter Johnson

00/00/0000

Number of pages: 00

Word count:0000

Abstract

Content

In content make colourful graphs like tosin

Introduction:

What is LLM?

Why LLM for these question?

Why small LLM? And how small is small LLM?

Encoder and Decoder problem?

Background:

1. Why for short answer question?

I want the story to start with charlies work, then go into the set up of the topic

1. What conditions improved since charlies work?
2. What is

Methodology:

Major difficuly:

Peftmodel module (allowing adding adaptors to modify model) is not inherently supported by torch pipeline nor huggingface pipeline, which makes deployment difficult (need to merge to a new model after testing)

LLM for classification finetuning met difficulty; for same reason the bare minimum case “long long true without further prompting” is not tested

Result

Discussion

Conclusion

AI writing declaration

5 sequencial training]/ infrastructure done

Llama3 is a decoder only model

Meaning of encoder: generates new meaningful things based on the input

Meaning of decoder: completes the prompt

Comparison between raw/finetuned casual model/finetuned classification model

Citations

Charlies report

Sophias report

Lambda Feedback related