

# Fine-Tuning LLMs on Finance-Alpaca Dataset: *A Comparative Study of LLMs for Financial Text Generation Tasks*



Capstone Project  
By Jasmine Coloma  
Professor Ying Mao  
Fall 2023

# Abstract and Introduction

## MOTIVATION:

LLMs can encounter challenges when understanding complex financial language

## RESEARCH QUESTION:

1. *What is the comparative impact of fine tuning LLMs on a financial instruction dataset?*
2. *How do these models differ in their ability to comprehend and generate finance specific text?*

## OBJECTIVE:

- To conduct a comparative analysis of LLMs on financial text dataset
- See how well each LLM can adapt to financial language

# Dataset

- **Alpaca Finance dataset:**
  - 4 features ("text," "instruction," "input," "output").



**Dataset Viewer**

Auto-converted to Parquet </> API Go to dataset viewer

Split

train (68.9k rows)

Search this dataset

text string · classes	instruction string · lengths	input string · lengths	output string · lengths
1 value	9 489	0 2,63k	0 17k
	For a car, what scams can be...		The car deal makes money 3 ways. If you pay in one lump payment. If the payment is greater than what...
	Why does it matter if a Central Bank...		That is kind of the point, one of the hopes is that it incentivizes banks to stop storing money and...
	Where should I be investing my...		Pay off your debt. As you witnessed, no "investment" % is guaranteed. But your debt...
	Specifically when do options expire?		Equity options, at least those traded in the American exchanges, actually expire the Saturday...
	Negative Balance from Automatic...		Automatic excersisions can be extremely risky, and the closer to the money the options are, the...
	Approximation of equity value for...		Generally "default" means that the company cannot pay off their debts, and since debt holders get...

< Previous 1 2 3 ... 690 Next >

# Methodology

## Model Quantization

Load models using 4-bit precision

## Base Model Evaluation

Perplexity Metric

Model Inference

## Fine Tune Model

Using QLoRA

Task: text generation

Trained for 100 steps and 1 epoch

## Model Inference and Evaluation

Training loss comparison

Perplexity comparison

Model inference comparison

# Step 1 - Model Selection: LLM Comparisons

Selection criteria: autoregressive models

Model	Llama 2	Falcon	Bloom	GPT2
Model size	7B	7.5B	3B	1.5B

# Step 2 - Model Quantization

- Motivation for 4 bit-precision
  - Challenge of running/training LLMs on consumer hardware
  - Bitsandbytes library

# Step 3 - Base Model Evaluation

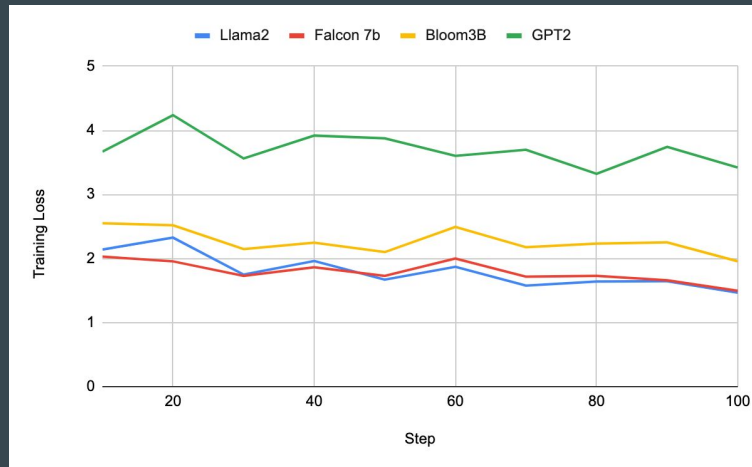
- **Perplexity Metric:** evaluates how well a model predicts a sequence of words
  - Lower perplexity = better performance
  - Evaluated perplexity over 100 samples of the dataset
- **Model Inference**
  - Asked the model - “Why should portfolios be diversified?”

# Step 4 - Fine Tune Strategy

1. SFTtrainer
  - a. Training parameters:
    - i. 100 steps
    - ii. 1 epoch,
    - iii. learning rate =  $2e-4$
    - iv. batch size 4
2. QLoRA
  - a. Allows for efficient fine-tuning on limited resources
  - b. Helps reduce memory without sacrificing performance



# Step 5 - Experiment Results: Training Loss Comparison



Steps	Llama 2	Falcon	Bloom	GPT2
10	2.14	2.0287	2.5505	3.6686
100	1.4675	1.4953	1.9554	3.4193
Total loss	0.6725	0.5334	0.5951	0.2

## Step 5 - Experiment Results: Perplexity Comparison

Model Name	Base Model	Fine Tuned Model	Performance Gain
Llama 2	2.58	3.62	+1.04
Falcon	1.73	1.69	-0.04
BLOOM	3.14	2.96	-0.18
GPT2	1.3125	1.016	-0.13

# Step 5 - Experiment Result: Model Inference

Analysis of responses to the query "Why should portfolios be diversified?"

Falcon Base Model Response:	Falcon Fine Tuned Response:
<p>Portfolios should be diversified because it <b>reduces the risk</b> of the portfolio. Diversification reduces the risk of the portfolio by spreading the risk across different assets. Diversification also reduces the risk of the portfolio by spreading the risk across different industries.</p>	<p>Diversification is a key principle of investing. It helps to <b>reduce the risk</b> of a portfolio by spreading investments across different asset classes and sectors. This can help to <b>reduce volatility</b> and <b>increase returns</b> over time. Additionally, diversification can help to <b>smooth out fluctuations</b> in the market, making it easier to manage risk. Diversification can also help to <b>reduce the impact</b> of individual investment losses. By spreading investments across different asset classes and sectors, a portfolio can be more ...</p>

*Overall fine tuned responses: more refined, coherent, and detailed*

# Result Conclusions

- Llama 2 had **highest** training loss but worse perplexity
- Bloom exhibited **most improvement** in perplexity
- GPT2 & Falcon showed consistent improvement
- All models exhibited more detailed yet clear explanations in model inference.

# Limitations and Future Work

- Evaluation limited to a select set of models; other models might be better
- Training for only 1 epoch
  - a. Potential data inconsistencies, biases, etc.
- Exclusively concentrates on financial text generation

# Conclusions

- Research contributes to understanding LLM adaptability to finance domain
- Can serve as a base analysis for further research on this topic