**Work Distribution:**

* **Adapter Training:**
  + **PACE: Yohann**
  + **COLAB: Winston**
  + **Predibase/GCloud/RunPod: Charlie**
  + **Meet back in a few days to see what progress we made. Post updates and finding to discord as we go**
* **Benchmarking**
  + **Existing models**
  + **Our Adapter(s)**
  + **Need benchmarking infrastructure**
* **Paper Writing**
  + **Using overleaf**
  + **Some sort of joint effort**

**Reference Models:**

* <https://github.com/pipiku915/FinMem-LLM-StockTrading/tree/51defbc3134fd56098debc3e3a6222fcd29b69a7?tab=readme-ov-file>
* <https://github.com/The-FinAI/PIXIU>
* <https://github.com/AI4Finance-Foundation/FinGPT>
* https://github.com/yxli2123/LoftQ

**Data Sources:**

* [**https://huggingface.co/AdaptLLM**](https://huggingface.co/AdaptLLM)
* [**https://huggingface.co/FinGPT**](https://huggingface.co/FinGPT)
* [**https://huggingface.co/datasets/gbharti/finance-alpaca**](https://huggingface.co/datasets/gbharti/finance-alpaca)
  + **Made up of:**
  + [**https://paperswithcode.com/dataset/finqa**](https://paperswithcode.com/dataset/finqa)
  + **https://github.com/tatsu-lab/stanford\_alpaca**
* [**https://huggingface.co/datasets/winddude/reddit\_finance\_43\_250k/tree/main**](https://huggingface.co/datasets/winddude/reddit_finance_43_250k/tree/main)

**Thoughts on Data:**

The finance alpaca looks like a good option. It combines the alpaca instruction tuning set with the finance stuff from FiQA. Now just need to make sure it’s not also being used for our bench mark.

**Benchmarks**

* **Financial Phrase Bank**
* **FiQA SA**
* **Headline**
* **NER**
* **FinBen (https://arxiv.org/html/2402.12659v1)**
* **Looks like we can maybe use this to help automate:**
  + [**https://github.com/Arize-ai/phoenix?tab=readme-ov-file#llm-evals**](https://github.com/Arize-ai/phoenix?tab=readme-ov-file#llm-evals)
  + **If not could potentially give each model 10 questions and then hand grade**
* [**https://deepgram.com/learn/apibank-llm-benchmark**](https://deepgram.com/learn/apibank-llm-benchmark)
* **Could also look at hand grading on a small bank of 10 questions**
  + For benchmarking the worst case scenario is that we come up with 10 original questions and then hand grade each model. Assuming 9 models and a generous 2min per question for grading that is worst case 3 hours. Assuming we only have to do that once or twice that might be more efficient then standing up an automated infrastructure

**Update for 4/7/24**

* Use same dataset for all models
* Train different types of lora models (lora, qlora, loftq, sweep hyper params too)
* Create table in project management doc lora types and hyper params to get DOE figured out (r=4, 8), so 6-9 different experiments
* Check and make sure that the dataset does not overlap with all the benchmarks
* scrub dataset for excessively long inputs/outputs
* start figuring benchmarking
* Charlie: Check for overlap with benchmark data in dataset and then scrub data for excessive len
* [
* 5:05 PM
* ]
* Yohann/Winston: Keep working on figuring out training platform
* [
* 5:06 PM
* ]
* Reconvene Wednesday

**Datasets**

Base: [**https://huggingface.co/datasets/gbharti/finance-alpaca**](https://huggingface.co/datasets/gbharti/finance-alpaca)

Cleaned: https://drive.google.com/file/d/1-9DKFj7yGcxoJcroGV54ZJ76A6\_H8yO3/view?usp=drive\_link

| **Model Name** | **Owner** | **Notes** | **Dataset** |
| --- | --- | --- | --- |
| AdaptLLM Finance Chat |  | https://huggingface.co/AdaptLLM/finance-chat |  |
| Claude 3 Opus |  |  |  |
| Mistral 7B |  | https://huggingface.co/mistralai/Mistral-7B-v0.1 |  |
| Mistral 7b Lora Rank 16 | Charlie | Completed | Trained on Base Dataset |
| Mistral 7b Lora Rank 8 | Charlie | Completed | Trained on Base Dataset |
| Mistral 7b QLora Rank 16, bits: 8 | Yohann | Completed |  |
| Mistral 7b QLora Rank 8, bits: 8 | Yohann | Completed |  |
| Mistral 7b LoftQ Rank 16 | Winston Lai | Completed |  |
| Mistral 7b LoftQ Rank 8 | Winston Lai | Completed |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Meeting 4-24-24**

* Take stock of all models we have trained, Winston needs to work harder.
  + Rank 8 lora
  + Rank 16 lora
  + Rank 8 qlora
  + Rank 16 qlora
  + Predibase QLORA (rank=128) (backup)
  + Predibase QLORA (rank=256) (backup)
* Decide what data needs to go in the report
  + Data: Alpaca finance
  + **Table 1:**
    - **Model Configs (rank, quantizing, alpha, epochs, learnable parameters)**
  + **Table 2 Training something:**
    - **Model sizes (memory) based on hand calculations. Can fallback to estimates based on when the training loop is called if needed**
    - **Training times**
  + **Table 3 Benchmark Results**
    - **Arithmetic Score, this is just some memory math**
    - **Reasoning/Advice Score, pass/fail questions about finance questions**
    - **Summary Score**
* Information to fill tables:
  + Thursday
* Draft Report:
  + Saturday Night
* Final Report
  + Sunday Night
  + <https://www.overleaf.com/project/661ac62b34a5223ff35eb96d>

| Model | Rank | Alpha | Epochs | Lora dropout | Target Modules | Quantization | Max Len |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lora | 8 | 16 | 5 | 0.05 | empty | N/A | 550 |
| Lora | 16 | 32 | 5 | 0.05 | empty | N/A | 550 |
| Qlora | 8 | 16 | 5 | 0.05 | empty | 4 | 550 |
| Qlora | 16 | 32 | 5 | 0.05 | empty | 4 | 550 |
| LoftQ | 8 | 16 | 5 | 0.05 | empty | 4 | 550 |
| LoftQ | 16 | 32 | 5 | 0.05 | empty | 4 | 550 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Requirements Checklist

* **Introduction**
  + ~~Did we explain what we are trying to do (overall)?~~
  + ~~Did we explain the problem we are trying to solve?~~
  + ~~Did we articulate our objective with little to no jargon (outside the context of a DL student)~~
  + ~~Did we explain the current state of the art and its limitations?~~
  + ~~Did we explain who this is of interest to and what the potential impacts are?~~
  + ~~Did we explain our data, the important aspects and why we chose it?~~
* Approach
  + Did we explain our approach clearly?
  + Did we explain how we attempted to solve the problem?
  + ~~If there was anything new did we talk about it?~~
    - ~~I think a comparison of the different LoRA methods on the same data is actually something a bit new~~
  + Did we explain we thought we might be successful?
  + ~~Did we explain our expected problems?~~
  + Did we explain if the first thing we tried work?
    - Not mentioned, but should talk about what we did subsequently
* Experiment
  + ~~Did we explain how we measured success?~~
  + ~~Did we explain what experiments we used?~~
  + ~~Did we explain results quantitatively?~~
  + Did we explain results qualitatively?
  + Did we succeed or fail? Why?
  + ~~Did we provide supporting data?~~
    - ~~Tables~~
    - ~~Cited papers~~
  + ~~Did we explain the problems we encountered?~~
    - ~~Why doesnt this go in the experiment section?~~
* Additional
  + Did we use tables and figures and were they any good?
  + Did we use illustrations to explain our ideas?
  + Did we use charts or tables to illustrate any “differences”?
    - I assume this means highlight data we shout was interesting
    - Probably just bold some lines in our tables
  + Can somebody else who has taken DL mostly understand the paper?
  + Did we explain how our project relates to DL?