TRADEXPERT: REVOLUTIONIZING TRADING WITH MIXTURE OF EXPERT LLMS

Qianggang Ding^{1,2}, Haochen Shi^{1,2}, Jiadong Guo³, Bang Liu^{1,2*}

¹ Université de Montréal, ² Mila, ³ The Hong Kong University of Science and Technology

ABSTRACT

The integration of Artificial Intelligence (AI) in the financial domain has opened new avenues for quantitative trading, particularly through the use of Large Language Models (LLMs). However, the challenge of effectively synthesizing insights from diverse data sources and integrating structured and unstructured data persists. This paper presents TradExpert, a novel framework that employs a mix of experts (MoE) approach, using four specialized LLMs, each analyzing distinct sources of financial data, including news articles, market data, alpha factors, and fundamental data. The insights of these expert LLMs are further synthesized by a General Expert LLM to make a final prediction or decision. With specific prompts, TradExpert can be switched between the prediction mode and the ranking mode for stock movement prediction and quantitative stock trading, respectively. In addition to existing benchmarks, we also release a large-scale financial dataset to comprehensively evaluate TradExpert's effectiveness. Our experimental results demonstrate TradExpert's superior performance across all trading scenarios.

1 INTRODUCTION

The fusion of artificial intelligence with financial analytics has spawned a new era of innovation, particularly with the infusion of Large Language Models (LLMs) into the realm of finance. These models, which have formerly excelled in natural language processing (NLP) tasks, are now being tailored to decode the complex and cryptic narratives of financial data. This adaptation is driven by a crucial insight: Financial markets are not just numbers-crunching engines but complicated information systems where the subtleties of news articles, reports, and economic indicators interweave to influence market dynamics.

Before the advent of LLMs, traditional financial models Zeng et al. (2023); Yang et al. (2020); Liu et al. (2020); Baek & Kim (2018), primarily relied on quantitative methods such as statistical analysis, time series forecasting, and econometric models. These models often struggled to incorporate unstructured data such as news articles or financial reports without manual intervention. As a result, the development of LLMs tailored for financial applications has progressed rapidly. Initial ventures into this domain repurposed general LLMs such as GPTs Brown (2020); Achiam et al. (2023) and LLaMAs Touvron et al. (2023a;b); Dubey et al. (2024) to interpret financial texts. However, more specialized language models such as FinBERT Araci (2019), BloombergGPT Wu et al. (2023), and FinGPT Yang et al. (2023a) have since evolved, demonstrating enhanced proficiency in understanding and predicting market movements from unstructured data. These models were specifically fine-tuned or pre-trained on vast amounts of financial corpus. This extensive training on domain-specific datasets has allowed them to better capture typical patterns in the financial corpus. Despite these advancements, the challenge remains to effectively synthesize insights from diverse data sources like historical stock prices, alpha factors, fundamental data, news articles, etc. In addition, integration of the deluge of unstructured financial texts with structured quantitative metrics still remains to be investigated with language models.

To this end, we propose the TradExpert framework, which stands at the confluence of these challenges. It leverages a Mixture of Experts (MoE) approach Eigen et al. (2013); Du et al. (2022); Shen

^{*}Corresponding author. Canada CIFAR AI Chair.

et al. (2023), involving multiple LLMs each specialized in distinct facets of financial data—news articles, market data, alpha factors, and fundamental data. This not only enhances the model's ability to process diverse data modalities but also allows for a more nuanced understanding of how different factors interact to influence market trends. Figure 1 illustrates differences among traditional, LLMbased, and MoE LLMs-based financial models. In TradExpert, each expert works with a distinct focus and produces specialized reports, which are finally summarized and analyzed by a general expert, just like the structured division of labor seen in the real world. Specifically, TradExpert employs specialized LLMs to first independently analyze different data sources, then integrates these analyses via another LLM that synthesizes insights to predict market movements and inform trading strategies. Innovatively, we utilize a reprogramming mechanism to convert time series data to embeddings aligned with LLMs. In addition, we propose two modes for the General Expert LLM, prediction mode and ranking mode, for stock movement prediction and stock trading strategies, respectively. In ranking mode, we innovatively let the LLM serve as a comparator within a relaxed sorting algorithm, enabling the selection of the Top-K ranked stocks for trading. To comprehensively evaluate our method, we have also collected a large-scale datasets, which will be publicly released.

Oue contributions can be summarized as follows.

- We present TradExpert, a novel framework that employs an MoE approach, integrating four LLMs each specialized to analyze distinct sources of financial data, imitating the structured division of labors seen in the real world.
- We utilize the LLM as a comparator within a relaxed sorting algorithm, which enables trades with the Top-K ranked stocks based on TradExpert's prediction.
- We release a comprehensive dataset encompassing a wide range of financial data, which serves as a new benchmark for financial analysis.
- Our comprehensive experiments show that TradExpert consistently outperforms state-ofthe-art baselines across all trading scenarios. Ablation studies validate the effectiveness of the modules proposed in TradExpert.

2 RELATED WORK

Financial Language Models have significantly advanced in recent years, blending NLP techniques with financial analytics to extract meaningful insights from vast amounts of unstructured financial data. To begin with, Fin-Bert Araci (2019) is a financial domain-specific variant of BERT, pretrained on a large corpus of financial communications. In 2023, BloombergGPT Wu et al. (2023) emerged as a 50-billion-parameter model trained on a vast financial dataset. FLANG Shah et al. (2022) introduced a financial language model with specialized masking and objectives. Astock Zou et al. (2022) provided a platform for studying



Figure 1: Illustration of traditional, LLM-based, and MoE LLMs-based financial models with diverse financial data sources.

NLP-aided stock auto-trading algorithms on the Chinese market. BBT-FinT5 Lu et al. (2023) advanced Chinese financial NLP with a large-scale pre-trained model. FinMA Xie et al. (2023) showcased a model fine-tuned on a multi-task instruction datasets. FinGPT Yang et al. (2023a) provided an open-source framework for financial LLMs. InvestLM Yang et al. (2023b) showed the effectiveness of instruction tuning for investment-related tasks. FinReport Li et al. (2024b) introduced a system for automatic financial report generation. Lastly, AlphaFin Li et al. (2024a) integrated retrieval-augmented generation techniques for financial analysis. Collectively, these works demonstrate the evolution of financial NLP models and benchmarks, advancing the capabilities of LLMs in financial applications.

Integration of Text and Financial Data has also been rapidly developed for stock movement prediction. StockNet Xu & Cohen (2018) developed a deep generative model that jointly exploits



Figure 2: TradExpert operates by processing distinct sources of financial data such as news texts, market data, alpha factors, and fundamental data through specialized expert LLMs. Then their reports are sumarized and sent to a General Expert which delivers the final outputs: (1) prediction of stock movement with prediction mode, (2) which of the two stocks is better or worse with ranking mode.

text and price signals for stock movement prediction. SLOT Soun et al. (2022) improved upon this by using self-supervised learning to handle sparse and noisy tweet data, capturing multi-level price trends. CH-RNN Wu et al. (2018) introduced a hybrid deep sequential modeling approach that leverages social text for stock prediction, incorporating cross-modal attention mechanisms. More recently, studies Lopez-Lira & Tang (2023); Chen et al. (2023) have explored the use of ChatGPT for stock movement prediction, comparing its performance with traditional state-of-the-art models. These works collectively demonstrate the increasing sophistication of models that integrate text and financial data, highlighting the potential for improving trading scenarios.

3 PROBLEM DEFINITION

In this study, we aim to trade stocks using a framework that incorporates Large Language Models (LLMs).

The input data comprises four primary components:

- News: Textual information from news articles pertinent to the stock and market conditions.
- Market Data: Historical OHLCV (Open, High, Low, Close, Volume) data representing the stock's trading activity.
- Alpha Factors: Quantitative indicators and signals believed to possess predictive power regarding stock price movements.
- Fundamental Data: Earnings call transcripts and fundamental metrics reflecting the company's economic health and performance.

Task 1: Stock movement prediction is a fundamental challenge in quantitative trading, which involves the prediction of future price trends based on multifaceted data sources. Formally, let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ denote our dataset, where x_i represents the input vector for the *i*-th stock on day t, and $y_i \in \{\text{Rise}, \text{Fall}\}$ is the corresponding label indicating whether the stock price will rise or fall on day t + 1. The input x_i can be expressed as:

$$x_i = \{\text{News}_i, \text{Market}_i, \text{Factors}_i, \text{Fundamental}_i\}$$
(1)

Our objective is to learn a predictive function f parameterized by θ such that $f_{\theta}(x_i) \approx y_i$, where f_{θ} is modeled using LLMs. The model outputs a binary prediction, "Rise" or "Fall" indicating the predicted stock price movement.

Task 2: Stock trading simulation involves evaluating the performance of *Buy-and-Hold* strategy based on the Top-K ranked stocks sorted by TradExpert. This task simulates real-market trading scenarios to assess the profitability and risk of TradExpert using metrics including Annualized Return (AR), Sharpe Ratio (SR), Annualized Volatility (AV), and Maximum Drawdown (MD).

4 DATASETS

In this study, we collected a comprehensive datasets encompassing various data sources including four primary components: News, Market Data, Alpha Factors, and Fundamental Data. The period covered by all data sources spans 4 years from January 1, 2020, to December 31, 2023.

4.1 STASTICS

News is collected from several reputable financial news sources, including Yahoo Finance, Reuters, InvestorPlace, GlobeNewswire, The Motley Fool, etc. This dataset comprises a total of 524,995 news articles for stocks on S&P 500 list, with an average word count of 596.4 words per article. Each news article is associated with a list of related stock tickers.

Market Data consists of historical daily OHLCV records for stocks on S&P 500 list. This dataset includes a total of 481,484 records, offering a detailed view of the stocks' trading activity over the specified period.

Alpha Factors incorporates 108 technical indicators and factors with their expressions, which are believed to possess predictive power regarding stock price movements.

Table 1: Components in each data source. † and ‡
denote generated by GPT-4 and external models,
respectively.

	Prompts	Responses
News	News articles	Movement Reasoning [†]
Market	OHLCV embeddings Statistics	Movement
Alpha	Expressions Descriptions [†] Comprehensive score [‡]	Movement
Fundamental	Earning calls scripts Fundamental metrics	Movement Reasoning [†]

Fundamental Data includes earnings call

transcripts, financial statements, and fundamental metrics. The earnings call transcripts are sourced from Seeking Alpha, with 16 transcripts (4 years, quaterly updated) available for each stock. Fundamental metrics include Earnings Per Share (EPS), Price-to-Earnings Ratio (P/E Ratio), Book Value Per Share (BVPS), etc.

4.2 DATA SPLIT

The datasets were split into training, validation, and test sets based on chronological order to ensure that future data remains unseen during the training process. The split was performed as follows: Training set: January 1, 2020, to June 30, 2022. Validation set: July 1, 2022, to December 31, 2022. Testing set: January 1, 2023, to December 31, 2023.

5 METHODOLOGY

In this study, we propose **TradExpert**, a novel framework leveraging the MoE LLMs approach, where four LLMs serve as specialized experts for distinct sources of financial data. A General Expert LLM then synthesizes the summaries of the four Expert LLMs to produce the final output. The pipeline of TradExpert is shown in Figure 2.

In TradExpert, all expert LLMs are built on the LLaMA-2-7B backbone LLM Touvron et al. (2023b) and are supervised and fine-tuned using the LoRA mechanism Hu et al. (2022). Before training and fine-tuning, we preprocess the raw datasets to construct prompts, instructions, and ground-truth responses for each LLM. An overall description of the preprocessed datasets is demonstrated in Table 1. The details will be introduced in the following.

Instruction: You are provided with a news article. Please predict how the stock will perform in the next <D> days. Your response should include your reasoning followed by a prediction of "Rise" or "Fall" in the specified format.

Format your response as follows: Reasoning: [Your reasoning here] Prediction: [Rise or Fall]

Prompt: News Article: [Insert news article text here]

Question: Given the information in the news article above, how is the stock expected to perform in the next <D> days?

Figure 3: Instruction and prompt for the News Analyst.

5.1 NEWS ANALYST

The News Analyst LLM is designed to analyze texts of news articles to predict stock movements. The prompt and instruction for fine-tuning the LLM are shown in Figure 7. The outputs from the News Analyst LLM include not only a prediction of the stock movement but also a reasoning of how the news article relates to the predicted movement in order to employ a Chain-of-Thought (CoT) Wei et al. (2022) reasoning approach. The ground-truth reasonings are pre-generated by the OpenAI GPT-4 API using instructions and prompts that incorporate the actual stock movements and the texts of news articles.

5.2 MARKET ANALYST

The Market Analyst LLM focuses on analyzing historical OHLCV (Open, High, Low, Close, Volume) data to predict stock movements. However, time series data is inherently continuous and lacks the discrete token structure that LLMs are designed to process. This misalignment poses a significant challenge in effectively utilizing LLMs on time series. To this end, we utilize a reprogramming mechanism Jin et al. (2024) to reprogram the input financial time series into text prototype representations.

Formally, let an OHLCV data instance be $\mathbf{X}^{(i)} \in \mathbb{R}^{N \times T}$ which consists of N variables across T time steps. $\mathbf{X}^{(i)}$ is first divided and embedded into a sequence of patch embeddings $\{\mathbf{X}_{P}^{(i)} \in \mathbb{R}^{N \times L_{P} \times d_{m}}\}$, where L_{P} and d_{m} are the number of patches and the patch embedding dimension respectively. The patches are then reprogrammed using a collection of text prototypes $\mathbf{E}' \in \mathbb{R}^{V \times D}$, which is achieved by linearly probing the LLM's pre-trained word embedding $\mathbf{E} \in \mathbb{R}^{V \times D}$, where V and V' are the size of the vocabulary of the LLM and the text prototypes ($V' \ll V$), and D is the embedding dimension. The reprogrammed patches are generated using a multi-head cross-attention mechanism: $\mathbf{Z}_{k}^{(i)} = \operatorname{Softmax}\left(\frac{\mathbf{Q}_{k}^{(i)}\mathbf{K}_{k}^{(i)\top}}{\sqrt{d_{k}}}\right)\mathbf{V}_{k}^{(i)}$, where query $\mathbf{Q}_{k}^{(i)} = \mathbf{X}_{P}^{(i)}\mathbf{W}_{k}^{Q}$, key $\mathbf{K}_{k}^{(i)} = \mathbf{E}'\mathbf{W}_{k}^{K}$, and value $\mathbf{V}_{k}^{(i)} = \mathbf{E}'\mathbf{W}_{k}^{V}$ for each head k. The reprogrammed embeddings $\mathbf{O}^{(i)}$ are obtained by aggregating the outputs from each attention head and projecting them to the hidden dimensions of the backbone LLM. Finally, the reprogrammed embeddings are augmented with a language description of statistics extracted from TSFresh Christ et al. (2018), serving as prompts for the Market Analyst. An example of instruction and prompt is shown in Figure 4.

5.3 ALPHA EXPERT

The Alpha Expert specializes in processing expression-based alpha factors, which are technical indicators and algorithm-generated factors believed to possess predictie power regarding stock price movements.

We leverage GPT-4's capability of understanding complex expressions to pre-generate a language description for each factor. In this way, we built our Alpha database, where an alpha record consists of:

Instruction: You are provided with historical OHLCV data of the past 20 days and a description of its statistics. Please predict how the stock will perform the next <D> day. Your response should be "Rise" or "Fall".

Prompt: <Embbedings of reprogrammed OHLCV>
Statistics: The historical prices have a minimum close of <min_val> <min_D> days ago, a maximum
close of <max_val> <max_D> days ago, and a median close of <median_val> <median_D> days
ago. The overall trend is <upward or downward>...
Question: Given the reprogrammed OHLCV data and its statistics, how is the stock expected to perform in
the next <D> days?

Figure 4: Instruction and prompt for the Market Analyst.

- Expression: The mathematical or logical formula used to compute the alpha factor based on OHLCV data. E.g. rank(ts_argmax(corr(ts_rank(close, 10), ts_rank(volume, 10), 10), 5))
- **Description**: Generated by GPT-4 with prompts that include the expression.

For each stock, we first calculate the values of all alpha factors based on OHLCV data and then derive a comprehensive score via a LightGBM-based model Ke et al. (2017). Subsequently, we select Top-K alphas that contribute most significantly to this comprehensive score. Descriptions of these Top-K alphas are retrieved from the database and, along with the calculated values, are included in the prompts and instructions for the Alpha Expert.

5.4 FUNDAMENTAL ANALYST

The Fundamental Analyst LLM specializes in analyzing fundamental data, including earnings call transcripts and financial metrics, to predict stock price movements on a quarterly basis. The procedure of the Fundamental Analyst LLM is similar to that of the News Analyst LLM, with key differences being that the fundamental data is updated quarterly and, therefore, the movement predictions are made for the next quarter. The response should include a prediction in one of the following five categories: "Strong Rise", "Moderate Rise", "No Change", "Moderate Fall", or "Strong Fall", followed by a reasoning.

5.5 GENERAL EXPERT

The General Expert LLM can operate in two distinct modes: prediction mode and ranking mode. Both modes begin by summarizing the reports (historical conversation including instructions, prompts, and responses) from the four specialized experts due to the limitations on input context length of the backbone LLM.

In **prediction mode**, used for stock movement prediction, the summarized reports are used to construct a prompt with a prediction prefix. Given the summarized reports, the General Expert LLM outputs a binary prediction indicating whether the stock will rise or fall.

In **ranking mode**, used for stock trading, the General Expert LLM functions as a comparator to establish the ranking ability. Specifically, given the summarized reports of two stocks, the General Expert LLM would determine which stock is likely to perform better in the future. To generate a Top-K ranking of stocks, we employ a relaxed comparison-based sorting similar to BubbleSort: We initially compare every pair of stocks and count the number of wins for each stock. Subsequently, we sort these counts to establish the rankings for stocks. Although algorithms like QuickSort and vanilla BubbleSort offer fewer comparisons for Top-K selection on average $\mathcal{O}(N \log N)$ and $\mathcal{O}(N \cdot K)$, we propose to use this relaxed comparison-based sorting alogrithm with $\mathcal{O}(N^2)$ due to the non-transitive nature of LLM-based comparator Liu et al. (2024). Therefore, more comparisons tend to yield more accurate rankings in practice.

The General Expert LLM is finetuned on both tasks of stock movement prediction and stock comparison simultaneously. The instructions and prompts are shown in Figure 5. Instruction: You are provided with a summarized report of the stock. Please predict whether the stock will rise or fall the next <D> day. Format your response as follows: Reasoning: [Your reasoning here] Prediction: [Rise or Fall]. Prompt: Summarized Report: [Insert summarized report here] Question: Based on the summarized report, will the stock rise or fall in the next < D > days? Instruction: You are provided with summarized reports of two stocks. Please determine which stock will perform better the next <D> day. Please output Stock AAA or Stock BBB. Prompt: Summarized Report for Stock AAA: [Report A] Summarized Report for Stock BBB: [Report B]

Figure 5: Instructions and prompts for the General Expert LLM: (Top) Prediction mode, (Bottom) Ranking mode.

Question: Based on the summarized reports, which stock will perform better in the next $\langle D \rangle$ days?

Table 2: Comparison results on stock movement prediction task. As a binary classification problem, methods are evaluated by Accuracy (Acc) and Mattheus Correlation Coefficient (MCC). Both metrics are better with higher values. The best and second best results are in **bold** and <u>underlined</u>, respectively.

	BigI	Data22	AC	CL18	CI	KM18	S&P50	0 (Ours)
	Acc ↑	$\mathbf{MCC}\uparrow$	Acc ↑	$\mathbf{MCC}\uparrow$	Acc ↑	$\mathbf{MCC}\uparrow$	Acc ↑	MCC ↑
	Hybrid	Models						
ALSTM-W	0.48	-0.01	0.53	0.08	0.54	0.03	0.55	0.10
ALSTM-D	0.49	0.01	0.53	0.07	0.50	-0.04	0.54	0.05
StockNet	0.53	-0.02	0.54	-0.02	0.52	-0.02	0.56	0.06
SLOT	0.55	0.10	<u>0.59</u>	0.21	0.56	0.09	/	/
	Large	Language I	Models					
GPT-4	0.54	0.03	0.52	0.02	0.57	0.02	0.58	0.10
Gemini	0.55	0.04	0.52	0.04	0.54	0.02	0.59	0.11
LLaMA2-7B-chat	0.54	0.05	0.51	0.01	0.55	-0.03	0.57	0.06
LLaMA2-70B	0.47	0.00	0.51	0.01	0.49	-0.07	0.57	0.02
LLaMA3-8B	0.55	0.02	0.52	0.02	0.57	0.03	0.53	0.04
FinMA-7B	0.51	0.02	0.51	0.03	0.50	0.08	0.59	0.09
FinGPT-7B-lora	0.45	0.00	0.49	0.00	0.42	0.00	0.51	0.05
InternLM	0.56	0.08	0.51	0.02	0.57	-0.03	0.60	0.06
Falcon	0.55	$\overline{0.00}$	0.51	0.00	0.47	-0.06	0.52	0.03
Mixtral-7B	0.46	0.02	0.49	0.00	0.42	-0.05	0.54	0.04
	MoE L	arage Lan	guage M	odels				
TradExpert-NM	0.59	0 .12	0.60	<u>0.15</u>	0.59	0.12	0.64	0.19

6 EXPERIMENTS

In this section, we conduct a comprehensive evaluation for TradExpert framework on two main tasks: stock movement prediction and stock trading simulation. Our experiments aims to address the following research questions: **RQ1:** How does TradExpert perform in stock movement prediction compared with state-of-the-art baselines? **RQ2:** What are the potential profits and associated risks of TradExpert in the backtesting on the real market? **RQ3:** How effective is the reasoning capability of TradExpert for unstructured data? **RQ4:** What is the significance of each expert within the TradExpert framework? **RQ5:** Why we choose the relaxed comparison-based sorting algorithm in TradExpert?

6.1 DATASETS

We include two categories of datasets in our experiments:

- **Benchmark Datasets:** We use publicly available benchmark datasets in stock movement prediction research including CIKM18 Wu et al. (2018), ACL18 Xu & Cohen (2018), and BigData22 Soun et al. (2022) datasets.
- **Proprietary Datasets:** We also utilize our proprietary datasets, which include extensive historical OHLCV data, news articles, alpha factors, and fundamental metrics for a comprehensive analysis.

6.2 EXPERIMENTAL SETUP

In our experiments, the four expert LLMs and the General Expert LLM are bulit on the LLaMA-2-7B bakcbone model Touvron et al. (2023b) and are finetuned via LoRA Hu et al. (2022) mechanism.

Stock Movement Prediction: TradExpert works in prediction mode, that is, the General Expert LLM reponses a binary prediction indicating whether a stock will rise or fall the next day. Methods are evaluated using binary classification metrics such as accuracy (Acc) and Matthews Correlation Coefficient (MCC).

Stock Trading Simulation: TradExpert works in ranking mode, that is, the General Expert LLM acts as a comparator to sort the stocks. We simulate the real profit and risk of TradExpert by executing trades based on the Top-K ranked stocks. TradExpert and baselines are evaluated using metrics including Annualized Return (AR), Sharpe Ratio (SR), Annualized Volatility (AV), and Maximum Drawdown (MD).

6.3 BASELINES

For stock movement prediction, the baselines include: (1) Hybrid Models: StockNet Xu & Cohen (2018), ALSTM-W Soun et al. (2022), ALSTM-D Soun et al. (2022), SLOT Soun et al. (2022). 2) Large Language Models: GPT-4 Achiam et al. (2023), Gemini Team et al. (2023), LLaMA2-70B Touvron et al. (2023b), LLaMA3-8B Dubey et al. (2024), FinMA-7B Xie et al. (2023), FinGPT-LlaMA2-7B Yang et al. (2023a), InternLM-7B Cai et al. (2024), Falcon-7B Almazrouei et al. (2023), Mixtral-7B Jiang et al. (2023).

For stock trading simulation, the baselines include: (1)Traditional Models: Random Forest Breiman (2001), Decision Tree Loh (2011), SVM Cortes & Vapnik (1995). (2) Deep Learning Models: A2C Mnih et al. (2016), PPO Schulman et al. (2017), SARL Ye et al. (2020), EIIE Jiang et al. (2017), and DeepTrader Wang et al. (2021). To reduce computational costs in backtesting, we evaluated all methods on datasets with stocks on the DOW 30 list, a subset of the S&P 500, with around 30 stocks.

6.4 RESULTS

6.4.1 STOCK MOVEMENT PREDICTION

We implemented all baselines ourselves or utilized existing open-source codes, except the closedsource model SLOT, for which we refer to the metrics reported in the relevant paper. To ensure a fair comparison, we only included the News Analyst and Market Analyst in TradExpert, named TradExeprt-NM. The results are shown in Table 2. As we can see, among hybrid models, SLOT achieves outstanding accuracy and MCC on the ACL18, benefitting from the proposed global market guidance. Among LLMs, InternLM shows remarkable performance, particularly on our proprietary S&P500 dataset. Our proposed TradExpert-NM, utilizing a mixture of expert LLMs approach, consistently outperformed other models across all datasets except for MCC on the ACL18, showcasing its superior performance. Noting that BigData22, ACL18, and CIKM18 are relatively small datasets with texts from tweets, while our S&P500 dataset consist of news articles with much more words. This difference in text lengths contributes to the more significant improvements obtained by TradExpert-7B-NM on the S&P500 dataset.

6.4.2 STOCK TRADING SIMULATION

	AR ↑	AV↓	SR↑	$\mathbf{MD}\downarrow$
DJI Index	13.92%	11.41%	1.22	9.70%
	Tradition	nal Models		
SVM	15.77%	26.67%	0.59	19.94%
XGBoost	21.58%	27.29%	0.79	21.90%
LightGBM	2.17%	22.74%	0.1	21.29%
	Deep Lea	arning Mo	dels	
A2C	19.16%	11.29%	1.7	9.09%
PPO	16.62%	11.51%	1.44	9.45%
EHE	23.64%	13.73%	1.72	10.07%
SARL	21.87%	14.72%	1.49	8.52%
DeepTrader	32.45%	17.86%	1.82	15.32%
	MoE La	rge Langua	nge Mod	els
TradExpert	49.79%	9.95%	5.01	6.56%

Table 3: Comparison results on stock trading simulation task with stocks on the DOW 30. Annualized Return (AR), Sharpe Ratio (SR), Annualized Volatility (AV), and Maximum Drawdown (MD) are utilized to evaluate the profits and risks of methods. The best results are in **bold**.

We perform backtesting to evaluate Trad-Expert and baselines. To reduce computational costs in backtesting, we limit the stock pool to about 30 stocks on the DOW 30, a subset of the S&P 500. For TradExpert, we implement a Buy-and-Hold trading strategy on the Top-K stocks ranked by TradExpert. The backtesting period is the same as the testing period of our datasets, which ranges from January 1, 2023, to December 31, 2023. The results summarized in Table 3 demonstrate TradExpert's superior performance across all metrics considered. Among traditional models, XGBoost achieved a relatively high return but also exhibited high volatility and drawdown, indicating greater risk. Deep learning models generally outperformed traditional models. Among them, Deep-Trader stood out with the highest return



Figure 6: Cumulative returns over time of all methods on 30 stocks on DOW 30 list. DJI Index represents the market trend.

and Sharpe ratio. TradExpert, our proposed model, significantly outperformed all other models with an exceptional AR of 49.79% and the lowest AV of 9.95%. This combination yielded an outstanding Sharpe ratio of 5.01, indicating a high return per unit of risk. Figure 6 shows the trends of cumulative returns over time for all methods.

7 ABLATION STUDY

The Impacts of Experts To evaluate the effectiveness of each expert within the TradExpert framework, we created multiple versions of TradExpert, each with a specific expert removed. By comparing the performance of these modified frameworks, we can assess the impact of each expert on the overall functionality of TradExpert. The results in Table 4 reveal the varying degrees of impact of each expert.

Га	b	e 4	4:	Ab	lation	stud	y i	for	the	impacts	s of	experts	•
----	---	-----	----	----	--------	------	-----	-----	-----	---------	------	---------	---

Configuration	AR ↑	$\mathbf{AV}\downarrow$	$\mathbf{SR}\uparrow$	$\mathbf{MD}\downarrow$
TradExpert	49.79%	9.95%	5.01	6.56%
w/o Market	30.87%	16.43%	1.88	13.29%
w/o Alpha	41.65%	11.38%	3.66	13.04 <i>%</i> 8.94%
w/o Fundamental	44.32%	10.68%	4.15	7.82%

The Market Analyst and the News Analyst

emerged as the most critical, significantly influencing profitability and risk management, as seen by the largest drop in AR and AV when they were removed, respectively. The Alpha Expert is obviously less impactful than the Market Analysts and the News Analysts. The Fundamental Analyst had the smallest effect on daily trading metrics, but provided essential long-term stability, evident from the modest changes in AR and MD upon its removal. This highlights a strategic balance in TradExpert, where each expert contributes uniquely to the final decision and prediction.

The Effectiveness of Structured Data Rea-

soning. We show the effectiveness by comparing TradExpert-MA with traditional models for structured data like OHLCV data and alpha factors. We use a genetic programmingbased symbolic regression model as our baseline, which mines alpha expressions aimed at predicting the RankIC of day T + 1's returns. TradExpert-MA is built on top of the same alphas, where News and Fundamental experts

Table 5:	Ablation	study	for	the	effecti	veness	of
structured	l data reas	oning	in p	redi	cting da	ay $T +$	1's
returns.							

Model	RankIC ↑	RankICIR ↑
TradExpert-MA	0.12	0.90
Alpha Combination	0.07	0.65

were removed to exclude affects from other sources. We compare TradExpert-MA with the combination of alphas using metrics of RankIC and RankICIR. The results are shown in Table 5. The improvements over the alpha combination demonstrate the reasoning ability of TradExpert for structured data.

The Choices of Ranking Algo-

rithm In TradExpert, we implement the Top-K ranking by sorting all stocks completely using a relaxed comparison-based algorithm, where TradExpert serves as the comparator. To justify our choice of this seemingly cumbersome approach, we conducted comparison experiments with other theoretically more efficient ranking algorithms. Specifically, our

Table 6: Ablation study for the choices of ranking algorithm. † denotes being equipped in TradExpert.

	RankIC ↑	RankICIR ↑	Time
$\textbf{RelaxedSort}^{\dagger}$	0.12	0.90	$O(N^2)$
BubbleSort QuickSort	0.06 0.03	0.65 0.38	$O(N \cdot K)$ $O(N \log N)$

alternatives include QuickSort and BubbleSort with time complexity $\mathcal{O}(N \log N)$ and $\mathcal{O}(N \cdot K)$, respectively. The comparison results in Table 6 demonstrate that our approach outperforms others, despite having a higher computational complexity. This is attributed to the non-transitive nature of LLM-based comparator. Therefore, a greater number of comparisons yield more accurate rankings in TradExpert.

8 CONCLUSION

In this study, we introduced TradExpert, a novel framework that harnesses the power of LLMs to enhance stock trading strategies. By integrating multiple specialized LLMs, each focused on distinct aspects of financial data, TradExpert provides a comprehensive and nuanced analysis that significantly outperforms traditional financial models in practice. Looking ahead, our goal is to explore how to employ TradExpert in the high-frequency trading scenario and extend its capabilities to encompass a wider range of global markets.

Limitation Although TradExpert has notable strengths, its processing time poses certain challenges. On average, it takes 4.7 seconds for a single stock with an Nvidia A5000 GPU. For daily trading, this processing time is generally manageable. However, for scenarios demanding quicker decision-making, such as high-frequency trading, TradExpert's latency becomes a notable drawback.

REFERENCES

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. arXiv preprint arXiv:2303.08774, 2023.
- Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, et al. The falcon series of open language models. *arXiv preprint arXiv:2311.16867*, 2023.
- Dogu Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.
- Yujin Baek and Ha Young Kim. Modaugnet: A new forecasting framework for stock market index value with an overfitting prevention lstm module and a prediction lstm module. *Expert Systems with Applications*, 113:457–480, 2018.

Leo Breiman. Random forests. Machine learning, 45:5-32, 2001.

Tom B Brown. Language models are few-shot learners. arXiv preprint ArXiv:2005.14165, 2020.

- Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, Xiaoyi Dong, Haodong Duan, Qi Fan, Zhaoye Fei, Yang Gao, Jiaye Ge, Chenya Gu, Yuzhe Gu, Tao Gui, Aijia Guo, Qipeng Guo, Conghui He, Yingfan Hu, Ting Huang, Tao Jiang, Penglong Jiao, Zhenjiang Jin, Zhikai Lei, Jiaxing Li, Jingwen Li, Linyang Li, Shuaibin Li, Wei Li, Yining Li, Hongwei Liu, Jiangning Liu, Jiawei Hong, Kaiwen Liu, Kuikun Liu, Xiaoran Liu, Chengqi Lv, Haijun Lv, Kai Lv, Li Ma, Runyuan Ma, Zerun Ma, Wenchang Ning, Linke Ouyang, Jiantao Qiu, Yuan Qu, Fukai Shang, Yunfan Shao, Demin Song, Zifan Song, Zhihao Sui, Peng Sun, Yu Sun, Huanze Tang, Bin Wang, Guoteng Wang, Jiaqi Wang, Jiayu Wang, Rui Wang, Yudong Wang, Ziyi Wang, Xingjian Wei, Qizhen Weng, Fan Wu, Yingtong Xiong, Chao Xu, Ruiliang Xu, Hang Yan, Yirong Yan, Xiaogui Yang, Haochen Ye, Huaiyuan Ying, Jia Yu, Jing Yu, Yuhang Zang, Chuyu Zhang, Li Zhang, Pan Zhang, Peng Zhang, Ruijie Zhang, Shuo Zhang, Songyang Zhang, Wenjian Zhang, Wenwei Zhang, Xingcheng Zhang, Xinyue Zhang, Hui Zhao, Qian Zhao, Xiaomeng Zhao, Fengzhe Zhou, Zaida Zhou, Jingming Zhuo, Yicheng Zou, Xipeng Qiu, Yu Qiao, and Dahua Lin. Internlm2 technical report, 2024.
- Zihan Chen, Lei Nico Zheng, Cheng Lu, Jialu Yuan, and Di Zhu. Chatgpt informed graph neural network for stock movement prediction. *arXiv preprint arXiv:2306.03763*, 2023.
- Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307: 72–77, 2018.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.
- Nan Du, Yanping Huang, Andrew M Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, et al. Glam: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, pp. 5547–5569. PMLR, 2022.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah

Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Olivier Duchenne, Onur Celebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaoqing Ellen Tan, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aaron Grattafiori, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alex Vaughan, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Franco, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, Danny Wyatt, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Firat Ozgenel, Francesco Caggioni, Francisco Guzmán, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Govind Thattai, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Karthik Prasad, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kun Huang, Kunal Chawla, Kushal Lakhotia, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Maria Tsimpoukelli, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikolay Pavlovich Laptev, Ning Dong,

Ning Zhang, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Rohan Maheswari, Russ Howes, Ruty Rinott, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Kohler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vítor Albiero, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaofang Wang, Xiaojian Wu, Xiaolan Wang, Xide Xia, Xilun Wu, Xinbo Gao, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yuchen Hao, Yundi Qian, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, and Zhiwei Zhao. The llama 3 herd of models, 2024. URL https://arxiv.org/abs/2407.21783.

- David Eigen, Marc'Aurelio Ranzato, and Ilya Sutskever. Learning factored representations in a deep mixture of experts. *arXiv preprint arXiv:1312.4314*, 2013.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum? id=nZeVKeeFYf9.
- Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. arXiv preprint arXiv:2310.06825, 2023.
- Zhengyao Jiang, Dixing Xu, and Jinjun Liang. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*, 2017.
- Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-Ilm: Time series forecasting by reprogramming large language models. In *The Twelfth International Conference on Learning Representations*, 2024.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.
- Xiang Li, Zhenyu Li, Chen Shi, Yong Xu, Qing Du, Mingkui Tan, and Jun Huang. Alphafin: Benchmarking financial analysis with retrieval-augmented stock-chain framework. In *Proceedings of* the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pp. 773–783, 2024a.
- Xiangyu Li, Xinjie Shen, Yawen Zeng, Xiaofen Xing, and Jin Xu. Finreport: Explainable stock earnings forecasting via news factor analyzing model. In *Companion Proceedings of the ACM on Web Conference 2024*, pp. 319–327, 2024b.
- Xiao-Yang Liu, Hongyang Yang, Qian Chen, Runjia Zhang, Liuqing Yang, Bowen Xiao, and Christina Dan Wang. Finrl: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv preprint arXiv:2011.09607*, 2020.
- Yinhong Liu, Han Zhou, Zhijiang Guo, Ehsan Shareghi, Ivan Vulic, Anna Korhonen, and Nigel Collier. Aligning with human judgement: The role of pairwise preference in large language model evaluators. arXiv preprint arXiv:2403.16950, 2024.

- Wei-Yin Loh. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1):14–23, 2011.
- Alejandro Lopez-Lira and Yuehua Tang. Can chatgpt forecast stock price movements? return predictability and large language models. arXiv preprint arXiv:2304.07619, 2023.
- Dakuan Lu, Hengkui Wu, Jiaqing Liang, Yipei Xu, Qianyu He, Yipeng Geng, Mengkun Han, Yingsi Xin, and Yanghua Xiao. Bbt-fin: Comprehensive construction of chinese financial domain pretrained language model, corpus and benchmark. arXiv preprint arXiv:2302.09432, 2023.
- Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pp. 1928–1937. PMLR, 2016.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347, 2017.
- Raj Sanjay Shah, Kunal Chawla, Dheeraj Eidnani, Agam Shah, Wendi Du, Sudheer Chava, Natraj Raman, Charese Smiley, Jiaao Chen, and Diyi Yang. When flue meets flang: Benchmarks and large pretrained language model for financial domain. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2022.
- Sheng Shen, Le Hou, Yanqi Zhou, Nan Du, Shayne Longpre, Jason Wei, Hyung Won Chung, Barret Zoph, William Fedus, Xinyun Chen, et al. Mixture-of-experts meets instruction tuning: A winning combination for large language models. arXiv preprint arXiv:2305.14705, 2023.
- Yejun Soun, Jaemin Yoo, Minyong Cho, Jihyeong Jeon, and U Kang. Accurate stock movement prediction with self-supervised learning from sparse noisy tweets. In 2022 IEEE International Conference on Big Data (Big Data), pp. 1691–1700. IEEE, 2022.
- Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023a.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288, 2023b.
- Zhicheng Wang, Biwei Huang, Shikui Tu, Kun Zhang, and Lei Xu. Deeptrader: a deep reinforcememant learning approach for risk-return balanced portfolio management with market conditions embedding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 643–650, 2021.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
- Huizhe Wu, Wei Zhang, Weiwei Shen, and Jun Wang. Hybrid deep sequential modeling for social text-driven stock prediction. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pp. 1627–1630, 2018.
- Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564, 2023.
- Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. Pixiu: a large language model, instruction data and evaluation benchmark for finance. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 33469–33484, 2023.

- Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1970–1979, 2018.
- Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the first ACM international conference on AI in finance*, pp. 1–8, 2020.
- Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. Fingpt: Open-source financial large language models. *FinLLM Symposium at IJCAI 2023*, 2023a.
- Yi Yang, Yixuan Tang, and Kar Yan Tam. Investlm: A large language model for investment using financial domain instruction tuning. *arXiv preprint arXiv:2309.13064*, 2023b.
- Yunan Ye, Hengzhi Pei, Boxin Wang, Pin-Yu Chen, Yada Zhu, Ju Xiao, and Bo Li. Reinforcementlearning based portfolio management with augmented asset movement prediction states. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 1112–1119, 2020.
- Zhen Zeng, Rachneet Kaur, Suchetha Siddagangappa, Saba Rahimi, Tucker Balch, and Manuela Veloso. Financial time series forecasting using cnn and transformer. *arXiv preprint arXiv:2304.04912*, 2023.
- Jinan Zou, Haiyao Cao, Lingqiao Liu, Yuhao Lin, Ehsan Abbasnejad, and Javen Qinfeng Shi. Astock: A new dataset and automated stock trading based on stock-specific news analyzing model. *arXiv preprint arXiv:2206.06606*, 2022.

A IMPLEMENTATION DETAILS

A.1 ENVIRONMENT

All experiments were conducted on a machine with the following specifications:

- Hardware: NVIDIA A5000 GPU x 4, AMD Ryzen Threadripper PRO 3975WX CPU, 256 GB RAM
- Software: Ubuntu 22.04, Python 3.9, PyTorch 2.2, CUDA 12.3

A.2 TRAINING

The proposed TradExpert framework utilizes a Mixture of Experts (MoE) approach, where four LLMs are specialized in processing distinct sources of financial data. All these LLMs are based on the LLaMA-2-7B Touvron et al. (2023b) model and fine-tuned using the LoRA mechanism Hu et al. (2022). Below are the specifics:

- **Expert LLMs:** Each expert LLM specializes in one of the following data sources: News, Market Data, Alpha Factors, and Fundamental Data.
- General Expert LLM: Integrates outputs from the four expert LLMs to make final predictions or rankings.
- **Fine-tuning:** We employ supervised-fine-tuning (SFT) with LoRA for all LLMs.

A.3 HYPERPARAMETERS

The key hyperparameters used for each of the LLMs in the TradExpert framework, including those fine-tuned with LoRA, are summarized in Table 7.

A.4 EVALUATION METRICS

The performance of the models was evaluated using the following metrics, with their respective equations and descriptions:

Hyperparameter	News Analyst	Market Analyst	Alpha Expert	Fundamental Analyst	General Expert
Model Architecture	LLaMA-2-7B	LLaMA-2-7B	LLaMA-2-7B	LLaMA-2-7B	LLaMA-2-7B
Learning Rate	$1e^{-4}$	$1e^{-4}$	$1e^{-4}$	$1e^{-4}$	$1e^{-4}$
Batch Size	4	4	4	4	4
Number of Epochs	30	30	30	30	30
Optimizer	AdamW	AdamW	AdamW	AdamW	AdamW
Sequence Length	2048	2048	2048	2048	2048
LoRA Rank	8	8	8	8	8
LoRA Alpha	16	16	16	16	16
LoRA Dropout	0.1	0.1	0.1	0.1	0.1
Warmup Ratio	0.1	0.1	0.1	0.1	0.1
Gradient Accumulation Steps	8	8	8	8	8

Table 7: Hyperparameters for Expert LLMs and General Expert LLM in the TradExpert framework.

A.4.1 STOCK MOVEMENT PREDICTION

Accuracy (Acc): Accuracy is the ratio of correctly predicted stock movements to the total number of predictions. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(2)

where:

- *TP* is the number of true positives (correctly predicted rises).
- TN is the number of true negatives (correctly predicted falls).
- *FP* is the number of false positives (incorrectly predicted rises).
- *FN* is the number of false negatives (incorrectly predicted falls).

Matthews Correlation Coefficient (MCC): The Matthews Correlation Coefficient is a balanced measure that considers all four quadrants of the confusion matrix (true/false positives and true/false negatives). It is especially useful for imbalanced datasets. MCC is defined as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$
(3)

MCC returns a value between -1 and 1, where 1 indicates perfect prediction, 0 indicates no better than random prediction, and -1 indicates total disagreement between prediction and observation.

A.4.2 STOCK TRADING SIMULATION

Annualized Return (AR): Annualized Return represents the geometric average amount of money earned by an investment each year over a given period. It is defined as:

$$AR = \left(\prod_{i=1}^{N} (1+r_i)\right)^{\frac{252}{N}} - 1$$
(4)

where r_i is the return on day *i* and *N* is the number of trading days. The factor of 252 is used to annualize the return (assuming there are 252 trading days in a year).

Sharpe Ratio (SR): The Sharpe Ratio measures the performance of an investment compared to a risk-free asset, after adjusting for its risk. It is defined as:

$$SR = \frac{AR - R_f}{\sigma}$$
(5)

where:

- AR is the annualized return of the portfolio.
- R_f is the risk-free rate (which is 0 in this paper for simplicity).
- σ is the annualized standard deviation of daily returns (a measure of volatility).

A higher Sharpe Ratio indicates better risk-adjusted returns.

Annualized Volatility (AV): Annualized Volatility is a measure of the risk or uncertainty of a financial instrument. It quantifies the degree of variation of trading prices over time. It is defined as:

$$AV = \sigma \sqrt{252} \tag{6}$$

where σ is the standard deviation of daily returns, and 252 is the number of trading days in a year.

Maximum Drawdown (MD): Maximum Drawdown is the maximum observed loss from a peak to a trough of a portfolio, before a new peak is attained. It is defined as:

$$MD = \max_{t \in [0,T]} \left(\frac{\text{Peak}_t - \text{Trough}_t}{\text{Peak}_t} \right)$$
(7)

where Peak_t is the maximum value at time t and Trough_t is the minimum value following the peak, within the observed time period [0, T]. Maximum Drawdown gives insight into the risk of a portfolio by identifying the largest possible loss.

A.5 REPRODUCIBILITY

To ensure the reproducibility of our results, we have included a partial of the pseudo codes and datasets used in our experiments in the supplementary material due to the limitation of file size.

B Algorithm

We illustrate the algorithm pipeline of TradExpert framework and relaxed comparison-based sorting algorithm for ranking mode of General Expert.

Algorithm 1 TradExpert Framework

- 1: **Input:** Financial datasets
 - $\mathcal{D} = \{News, Market, Alpha, Fundamental\}$
- 2: Output: Predicted stock movements or Top-K ranked stocks for trading
- 3: Step 1: Data Preprocessing
- 4: Preprocess each dataset:
 - Tokenize and truncate news articles to a fixed length.
 - Reprogram market data (OHLCV) into text prototypes.
 - Generate language descriptions for alpha factors.
 - Tokenize earnings call transcripts and generate fundamental metrics prompts.
- 5: Step 2: Expert LLMs Processing
- 6: for each dataset in \mathcal{D} do
- 7: Apply the corresponding Expert LLM:
 - **News Analyst:** Analyze news articles and predict stock movement (CoT with reasoning).
 - Market Analyst: Process market data and predict stock movement.
 - Alpha Expert: Evaluate alpha factors and predict stock movement.
 - **Fundamental Analyst:** Analyze fundamental data for quarterly predictions (CoT with reasoning).
- 8: **end for**
- 9: Step 3: Integration by General Expert LLM
- 10: Summarize the outputs of all Expert LLMs.
- 11: Use the General Expert LLM to:
 - Prediction Mode: Predict whether the stock will rise or fall.
 - Ranking Mode: Compare and rank stocks to select the Top-K for trading.
- 12: Step 4: Final Output
- 13: Return the predicted stock movements or the Top-K ranked stocks for trading.

Algorithm 2 Relaxed Comparison-Based Sorting Algorithm for Ranking **Require:** $S = \{s_1, s_2, ..., s_n\}$ {Set of stocks} **Require:** $R = \{r_1, r_2, \dots, r_n\}$ {Summarized reports for each stock} **Require:** *K* {Number of top stocks to select} 1: $C \leftarrow [0, 0, \dots, 0]$ {Initialize win counts for each stock} 2: for i = 1 to n do 3: for j = i + 1 to n do 4: $result \leftarrow General_Expert_LLM_Compare(r_i, r_j)$ 5: if *result* is True then 6: $C[i] \leftarrow C[i] + 1$ {Stock s_i wins} 7: else $C[j] \leftarrow C[j] + 1$ {Stock s_j wins} 8: end if 9: end for 10: 11: end for 12: $ranked_stocks \leftarrow Sort(S, key = C, reverse = True)$ {Sort stocks by win counts in descending order} 13: $top_k stocks \leftarrow ranked_stocks[1:K]$ {Select the Top-K stocks} 14: **return** top_k_stocks

Instruction: You are provided with a news article. Please predict how the stock will perform in the next <D> days. Your response should include your reasoning followed by a prediction of "Rise" or "Fall" in the specified format.

Format your response as follows: Reasoning: [Your reasoning here] Prediction: [Rise or Fall]

Prompt: News Article: [Insert news article text here] Question: Given the information in the news article above, how is the stock expected to perform in the next <D> days?

Figure 7: Instruction and prompt for the News Analyst.

C INSTRUCTIONS AND PROMPTS

The TradExpert framework utilizes several specialized Large Language Models (LLMs) to process distinct types of financial data, each guided by specific instructions and prompts.

- **News Analyst LLM:** The instructions and prompt for predicting stock movements based on news articles are shown in Figure 7.
- Market Analyst LLM: The instructions and prompt for analyzing historical OHLCV data are shown in Figure 8.
- Alpha Expert LLM: The instructions and prompt for evaluating alpha factors derived from OHLCV data are shown in Figure 9.
- Fundamental Analyst LLM: The instructions and prompt for predicting stock movements based on earnings call transcripts and fundamental metrics are shown in Figure 10.
- General Expert LLM (Prediction Mode): The instructions and prompt for predicting stock movements by integrating outputs from other experts are shown in Figure 11.
- General Expert LLM (Ranking Mode): The instructions and prompt for comparing and ranking stocks are shown in Figure 12.

D EXAMPLES OF DATASETS

Figures 13, 14, 15, and 16 illustrate the example datasets for the News Analyst, Market Analyst, Alpha Expert, and Fundamental Analyst LLMs, respectively, while Figure 17 presents the example dataset for the General Expert LLM, which synthesizes the inputs from all expert LLMs.

Instruction: You are provided with historical OHLCV data of the past 20 days and a description of its statistics. Please predict how the stock will perform the next <D> day. Your response should be "Rise" or "Fall".

Prompt: <Embbedings of reprogrammed OHLCV> Statistics: The historical prices have a minimum close of <min val> <min D> days ago, a maximum close of <max val> <max D> days ago, and a median close of <median val> <median D> days ago. The overall trend is <upward or downward>... Question: Given the reprogrammed OHLCV data and its statistics, how is the stock expected to perform in the next <D> days?

Figure 8: Instruction and prompt for the Market Analyst.

Instruction: You are provided with alpha factors derived from OHLCV data. Please predict the stock's movement based on the top contributing alpha factors. Your response should be "Rise" or "Fall".

Prompt: Alpha Factors: [Insert top contributing alpha factors here] Descriptions: [Insert language descriptions for alpha factors] Question: Based on the provided alpha factors, how is the stock expected to perform in the next <D> days?

Figure 9: Instruction and prompt for the Alpha Expert.

Instruction: You are provided with a summarized report of the stock's earnings call transcripts and fundamental metrics. Please predict whether the stock will rise or fall in the next quarter. Your response should include a prediction in one of the following five categories: "Strong Rise," "Moderate Rise," "No Change," "Moderate Fall," or "Strong Fall," followed by reasoning.

Prompt: Summarized Report: [Insert earnings call transcrpts and summarized fundamental metrics here] Question: Based on the fundamental information, will the stock rise or fall in the next quarter?

Figure 10: Instruction and prompt for the Fundamental Analyst.

Instruction: You are provided with a summarized report of the stock. Please predict whether the stock will rise or fall in the next $\langle D \rangle$ day.

Prompt: Summarized Report: [Insert summarized report here] Question: Based on the summarized report, will the stock rise or fall in the next <D> days?

Figure 11: Instruction and prompt for the General Expert (Prediction Mode).

Instruction: You are provided with summarized reports of two stocks. Please determine which stock will perform better in the next <D> day. Please output "Stock AAA" or "Stock BBB".

Prompt:

Summarized Report for Stock AAA: [Insert summarized report A here] Summarized Report for Stock BBB: [Insert summarized report B here] Question: Based on the summarized reports, which stock will perform better in the next <D> days?

Figure 12: Instruction and prompt for the General Expert (Ranking Mode).

Instruction: You are provided with a news article. Please predict how the stock will perform in the next <D> days. Your response should include your reasoning followed by a prediction of "Rise" or "Fall" in the specified format.

Format your response as follows: Reasoning: [Your reasoning here] Prediction: [Rise or Fall]

Prompt:

News Article:

Competition fears couldn't rattle Apple stock this year, and 2024 looks bright, say Wedbush analysts. And indeed, shares of the iPhone maker are up 49% this year.

The Wedbush analysts, led by Dan Ives, wrote in a Thursday report, "The 'iPhone China demise narrative' was a great fictional story by the bears which is far from the reality as underlying mainland China growth remains strong for the December quarter based on our Asia supply chain checks and a key asset for the core iPhone franchise."

Apple is still the firm's "top tech pick," thanks to "a strong iPhone 15 upgrade cycle," an unproblematic holiday season leading into next year, in spite of competition from Huawei in China, and growth in the Services business. Wedbush rates Apple stock at Outperform with a \$250 price target. That's 29% upside from Thursday's closing price of \$193.58.

Question: Given the information in the news article above, how is the stock expected to perform in the next 1 day?

Model Response:

Reasoning: The news article highlights strong performance and positive future prospects for Apple, particularly due to a strong iPhone 15 upgrade cycle and robust growth in the Services business. The positive outlook and the 29% upside potential as per Wedbush analysts suggest investor confidence, leading to a potential rise in stock price. Prediction: Rise

Figure 13: Example data for the News Analyst LLM.

Historical OHLCV Data (Last 20 Days):						
Date	Open	High	Low	Close	Volume	
2023-12-01	190.33	191.56	189.23	190.74	45,679,300	
2023-12-04	189.98	190.05	187.45	188.93	43,389,500	
2023-12-05	190.21	194.40	190.18	192.91	66,628,400	
2023-12-06	194.45	194.76	192.11	191.81	41,089,700	
2023-12-07	193.63	195.00	193.59	193.76	47,477,700	
2023-12-08	194.20	195.99	193.67	195.20	53,377,300	
2023-12-11	193.11	193.49	191.42	192.67	60,943,700	
2023-12-12	193.08	194.72	191.72	194.20	52,696,900	
2023-12-13	195.09	198.00	194.85	197.44	70,404,200	
2023-12-14	198.02	199.62	196.16	197.59	66,831,600	
2023-12-15	197.53	198.40	197.00	197.05	128,256,700	
2023-12-18	196.09	196.63	194.39	195.38	55,751,900	
2023-12-19	196.16	196.95	195.89	196.42	40,714,100	
2023-12-20	196.90	197.68	194.83	194.32	52,242,800	
2023-12-21	196.10	197.08	193.50	194.17	46,482,500	
2023-12-22	195.18	195.41	192.97	193.09	37,122,800	
2023-12-26	193.61	193.89	192.83	192.54	28,919,300	
2023-12-27	192.49	193.50	191.09	192.64	48,087,700	
2023-12-28	194.14	194.66	193.17	193.07	34,049,900	
2023-12-29	193.90	194.40	191.73	192.02	42,628,800	

Instruction: You are provided with historical OHLCV data of the past 20 days and a description of its statistics. Please predict how the stock will perform the next 1 day. Your response should be "Rise" or "Fall".

Prompt:

Embeddings of Reprogrammed OHLCV: [Insert embeddings here]

Statistics: The historical prices have a minimum close of \$188.93 on 19 days ago, a maximum close of \$197.59 on 10 days ago, and a median close of \$194.20 over the last 20 days. The overall trend appears to be slightly downward in the last 5 days.

Question: Given the reprogrammed OHLCV data and its statistics, how is the stock expected to perform in the next 1 day?

Model Response: Prediction: Fall

Figure 14: Example data for the Market Analyst LLM.

Top-5 Alpha Factors (Largest Values):

- ID Expression
- 26 1 * (close ts_min(close, 30)) / (ts_max(close, 30) ts_min(close, 30))
- 5 1 * (close vwap) / decay_linear(rank(ts_argmax(close, 30)), 2)
- 29 1 * ts_rank(stddev(high, 10), 5)
- 27 1 * ts_rank(correlation(rank(low), rank(adv15), 9), 15)
- 25 1 * ts_rank(correlation(rank(high), rank(volume), 5), 5)

Instruction: You are provided with alpha factors derived from OHLCV data. Please predict the stock's movement based on the top contributing alpha factors. Your response should be "Rise" or "Fall."

Prompt:

Alpha Factors:
ID 26: 1 * (close - ts_min(close, 30)) / (ts_max(close, 30) - ts_min(close, 30))
ID 5: 1 * (close - vwap) / decay_linear(rank(ts_argmax(close, 30)), 2)
ID 29: 1 * ts_rank(stddev(high, 10), 5)
ID 27: 1 * ts_rank(correlation(rank(low), rank(adv15), 9), 15)
ID 25: 1 * ts_rank(correlation(rank(high), rank(volume), 5), 5)
Descriptions:

ID 26: Measures how close the current closing price is to the minimum and maximum close in the last 30 days.
ID 5: Evaluates the deviation of the closing price from the VWAP, adjusted by a decayed rank.
ID 29: Ranks the standard deviation of the high prices over the last 10 days.
ID 27: Ranks the correlation between the low price and a 15-day average volume rank.
ID 25: Ranks the correlation between the high price and volume over the last 5 days.

The comprehensive score derived from LightGBM is: 0.88
Question: Based on the provided alpha factors, how is the stock expected to perform in the next 1 day? —

Model Response:

Prediction: Rise

Figure 15: Example data for the Alpha Expert LLM.

Summarized Earning Calls Transcript:

Apple reported strong Q4 2023 financial results with total net sales of \$89.50 billion, a gross margin of \$40.43 billion, operating income of \$26.97 billion, and net income of \$22.96 billion. The company achieved a diluted earnings per share (EPS) of \$1.46.

During the earnings call, Apple's CEO Tim Cook highlighted the company's record performance in key markets, including India, China, and Latin America. iPhone revenue set a September quarter record at \$43.8 billion, up 3% year-over-year, driven by strong demand in emerging markets. Apple's services segment also performed exceptionally well, setting an all-time revenue record of \$22.3 billion, a 16% year-over-year increase.

However, the Mac and iPad segments experienced declines in revenue due to challenging market conditions and tough comparisons to the prior year's extraordinary results, where supply disruptions led to pent-up demand being fulfilled in the corresponding quarter. Despite these challenges, Apple continued to invest in future innovations, including the launch of new M3 chips and Apple Vision Pro.

Looking forward, Apple provided guidance that it expects iPhone revenue to grow year-over-year in the December quarter and anticipated strong gross margins of between 45% and 46%. The company also highlighted its strategic focus on emerging markets, with substantial growth opportunities in India. **Fundamental Metrics:**

Total Net Sales: \$89.50 billion Gross Margin: \$40.43 billion Operating Income: \$26.97 billion Net Income: \$22.96 billion Earnings Per Share (EPS): Basic EPS: \$1.47 Diluted EPS: \$1.46

Instruction: You are provided with a summarized report of the stock's earnings call transcript and fundamental metrics. Please predict whether the stock will rise or fall in the next quarter. Your response should include a prediction in one of the following five categories: "Strong Rise," "Moderate Rise," "No Change," "Moderate Fall," or "Strong Fall," followed by reasoning.

Prompt:

Summarized Report: Apple's Q4 2023 earnings report shows strong financial performance with a total net sales of \$89.50 billion, a gross margin of \$40.43 billion, operating income of \$26.97 billion, and net income of \$22.96 billion. The earnings per share (EPS) are \$1.46 (diluted).

During the earnings call, CEO Tim Cook emphasized record iPhone revenue in several key markets and a strong performance in the services segment. The company expects continued growth in iPhone revenue and strong gross margins in the upcoming quarter, supported by ongoing investments in innovation and expansion into emerging markets.

Question: Based on the summarized report, will the stock rise or fall in the next quarter?

Model Response:

Reasoning: Apple's robust financial metrics, particularly the record iPhone and services revenue, coupled with the positive outlook for the December quarter, suggest strong market confidence. The company's strategic focus on emerging markets, along with its commitment to innovation, further supports this positive sentiment. Given these factors, it is likely that the stock will experience a moderate rise in the next quarter. Prediction: Moderate Rise

Figure 16: Example data for the Fundamental Analyst LLM.

Input from News Analyst:

- **Reasoning:** The news article highlights strong performance and positive future prospects for Apple, particularly due to a strong iPhone 15 upgrade cycle and robust growth in the Services business. The positive outlook and the 29% upside potential as per Wedbush analysts suggest investor confidence, leading to a potential rise in stock price.
- Prediction: Rise

Input from Market Analyst:

- **Historical OHLCV Data:** The stock has shown a slightly downward trend over the last 5 days, with a recent close of \$192.02.
- Prediction: Fall

Input from Alpha Expert:

• Top-5 Alpha Factors:

- ID 26: 1 * (close ts_min(close, 30)) / (ts_max(close, 30) ts_min(close, 30))
- ID 5: 1 * (close vwap) / decay_linear(rank(ts_argmax(close, 30)), 2)
- ID 29: 1 * ts_rank(stddev(high, 10), 5)
- ID 27: 1 * ts_rank(correlation(rank(low), rank(adv15), 9), 15)
- ID 25: 1 * ts_rank(correlation(rank(high), rank(volume), 5), 5)
- Prediction: Rise

Input from Fundamental Analyst:

- **Reasoning:** Apple's robust financial metrics, particularly the record iPhone and services revenue, coupled with the positive outlook for the December quarter, suggest strong market confidence. The company's strategic focus on emerging markets, along with its commitment to innovation, further supports this positive sentiment.
- Prediction: Moderate Rise

Instruction: You are provided with a summarized report of the stock. Please predict whether the stock will rise or fall the next 1 day.

Format your response as follows: Reasoning: [Your reasoning here] Prediction: [Rise or Fall].

Model Response:

Reasoning: While the Market Analyst predicts a fall based on recent downward trends in the OHLCV data, the overall sentiment from the News Analyst, Alpha Expert, and Fundamental Analyst is positive. The strong financial metrics, record iPhone and services revenue, and positive forward guidance from the earnings call provide a solid foundation for a potential rise. The Alpha Expert's positive alpha factors also support this outlook. Therefore, despite the short-term downward trend, the overall stock movement is expected to be upward. Prediction: Rise

Figure 17: Example data for the General Expert LLM.