FinRankGRPO: Optimizing LLMs for Listwise Financial Asset Ranking via Group Relative Policy Optimization

Anonymous ACL submission

Abstract

While Large Language Models (LLMs) excel at understanding unstructured financial contexts, their direct use in portfolio optimization is limited by a mismatch between next token prediction and the listwise ranking objectives required for asset allocation. They also struggle with precise numerical forecasting, leading to instability and arithmetic hallucinations. To bridge this gap, we propose FinRankGRPO, a framework that shifts LLM-based portfolio construction from direct numerical prediction to listwise ranking of financial assets. We introduce a two-stage training process, supervised finetuning on Chain-of-Thought reasoning data, followed by our Financial Asset Ranking via Group Relative Policy Optimization with a Spearman rank correlation reward that aligns generated asset rankings with ground truth market orderings. The second stage uses a novel Spearman rank correlation reward to explicitly align the model's generative preferences with ground truth market orderings. Experimental results show that FinRankGRPO outperforms traditional quantitative and state-of-the-art commercial models, achieving a Sharpe ratio of 0.636 and a Spearman correlation of 0.023. Our code and data are publicly available at: https://anonymous.4open.science/r/FinRankGRPO-7107/.

1 Introduction

Portfolio management is a cornerstone of financial investment, centered on allocating capital across different assets (equities, bonds, commodities, and alternatives) to maximize returns while minimizing risk (Markowitz, 1971; Fabozzi, 2007; Kirtac and Germano, 2025). Classical approaches such as Mean-Variance Optimization (MVO) (Markowitz, 1952) provide a standard backbone for portfolio construction given estimates of return and covariance. Later, the Black-Litterman and Entropy Pooling (EP) frameworks extend this paradigm by allowing investors to incorporate subjective views, such as relative outperformance expectations, into the portfolio construction process (Black and Litterman, 1992; Idzorek, 2007; Meucci, 2008), as illustrated as Path 1 in Figure 1. However, these frameworks assume structured numerical views (e.g., "asset A is expected to outperform asset B by 5%") and offer limited guidance on deriving such views, or relative asset rankings, from unstructured text such as market news (Mahdavi et al., 2025; Lee et al., 2025b).

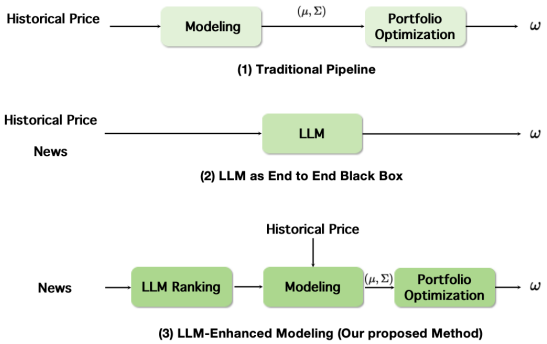

Figure 1: Comparison of portfolio management approaches. μ and Σ denote the expected returns and the covariance matrix, respectively. (1) Traditional pipelines separate return modeling and portfolio optimization. (2) LLM as End-to-end Black Box. (3) Our method generates listwise asset rankings and incorporates them into portfolio optimization.

In recent years, Large Language Models (LLMs) emerges as a promising solution, as shown in Path 2 of in Figure 1, by integrating textual and numerical data for financial analysis. FinSrag (Xiao et al., 2025) uses retrieval-augmented generation (RAG) for financial time series forecasting. LLM-BLM (Lee et al., 2025b) integrates LLM-generated prediction returns and confidence into the Black-Litterman model. Agent-based architectures, like MarketSenseAI (Fatouros et al., 2025), employ specialized AI agents to analyze multi-source financial data and produce stock selection signals.

Despite their advances, existing LLM-augmented approaches still face two critical limitations.

Firstly, there is a fundamental mismatch between standard LLM training objectives and financial investment. Most LLM-based approaches are not designed to optimize ranking relationships. Recent supervised finetuning (SFT) methods adapt LLMs to stock price (Guo and Hauptmann, 2024) or return forecasting (Chiu and Hung, 2024), but still rely on next token prediction losses that fail to capture the listwise ranking structures required for portfolio construction. Agent-based frameworks, such as TradingAgents (Xiao et al., 2024a) and FinCon (Yu et al., 2024a), also rely on generic training paradigms rather than ordinal objectives. Reinforcement learning methods share this limitation. GRPO (Shao et al., 2024) is designed for binary verification rather than ranking, while SAPPO (Kirtac and Germano, 2025) enhances portfolio optimization by integrating reinforcement learning with sentiment analysis, its reward mechanism lacks ranking specific objectives. Consequently, existing methods do not directly encourage economically meaningful asset rankings, which are essential for effective and interpretable portfolio construction.

Secondly, LLMs have limitations in reliable numerical forecasting, which weakens their direct use in quantitative portfolio management. Recent studies (Lee et al., 2025a) show that this approach is susceptible to arithmetic hallucinations and investment biases (Lee et al., 2025a). Even state-of-the-art models such as OpenAI o3-deep-research achieve only approximately 40% accuracy, suggesting that internal reasoning is still insufficient for financial numeracy forecasting (Li et al., 2026). Instead of relying on brittle point estimates, ordinal ranking offers a more robust objective by leveraging LLMs' comparative reasoning and reducing sensitivity to numerical noise.

This raises a natural question: Can we design a ranking driven reinforcement learning framework that adapts LLMs to asset ranking for portfolio construction?

To answer this question, we move from numerical prediction to robust asset ranking. Specifically, we introduce Financial Asset Ranking via Group Relative Policy Optimization (FinRankGRPO), shown as Path 3 in Figure 1, a reinforcement learning framework that optimizes LLMs with a ranking based reward. The reward is based on Spearman's correlation between the predicted asset order and ground true ordering. By converting financial reasoning into a listwise ranking problem, FinRankGRPO aligns LLMs with the ordinal structure of investment decisions. The learned rankings are then integrated into an EP framework (Meucci and Nicolosi, 2016; Meucci, 2010) as relative preference constraints over assets.

Our contributions are summarized as follows. First, we propose FinRankGRPO, a ranking-based alignment framework that reformulates financial text understanding as a listwise asset ranking task and optimizes LLMs with a Spearman based reward for structural and ordinal accuracy. Second, we integrate LLM derived rankings into an EP framework, enabling textual signals to inform mathematically grounded portfolio optimization without requiring direct numerical forecasts. Third, we conduct extensive experiments showing that FinRankGRPO outperforms traditional quantitative and LLM baselines in both ranking quality and portfolio performance.

2 Related Works

2.1 LLMs in Portfolio Management

Modern Portfolio Theory frames portfolio construction as an optimization between expected return and variance risk (Markowitz, 1952; Fabozzi, 2007), but it is sensitive to noisy return estimates (Idzorek, 2007; Black and Litterman, 1992) and can overemphasize assets with estimation errors (Michaud, 1989). The Black-Litterman model mitigates this issue by incorporating subjective views into posterior return and covariance estimates (Black and Litterman, 1992; Idzorek, 2007), while EP further generalizes view integration and supports ordinal ranking constraints (Meucci, 2010; Mercurio et al., 2020).

LLMs have recently expanded financial text understanding beyond conventional forecasting. General models such as ChatGPT (Dowling and Lucey, 2023), domain models such as BloombergGPT (Wu et al., 2023) and FinGPT (Yang et al., 2023), and reasoning systems such as Fin-R1 (Liu et al., 2025) and DianJin-R1 (Zhu et al., 2025a) show the value of domain adaptation, instruction tuning, and reinforcement learning for financial reasoning. Agent frameworks, including FinAgent (Zhang et al., 2024a) and FinCon (Yu et al., 2024b), further broaden these applications.

However, financial LLMs are still mostly used for stock picking (Xiao et al., 2024b), buy/sell/hold decisions (Zhang et al., 2024a), or numerical

views for Black-Litterman optimization (Lee et al., 2025b). FinRankGRPO instead uses LLMs to produce predictive ranking views, decoupling textual reasoning from direct numerical return or weight generation. This design reduces numerical inconsistency while preserving interpretable Chain-of-Thought (CoT) rationales for allocation decisions.

2.2 LLM-Based Ranking

LLM-based ranking has been explored in information retrieval (Zhang et al., 2024b; Khramtsova et al., 2024; Sun et al., 2023), recommendation (Gao et al., 2025; Sharma et al., 2024; Zhu et al., 2025b), and question answering (Dehghankar and Asudeh, 2025; Pradeep et al., 2023). Early methods encode items as discrete tokens and cast ranking as sequence generation (Hua et al., 2023; Zhu et al., 2024), while later prompt-based methods generate recommendation lists in natural language (Zhu et al., 2025c). Systems such as ChatRec (Gao et al., 2023) and InstructRec (Zhang et al., 2023) show that instruction tuned LLMs can follow complex ranking criteria, but their outputs may remain unstable without explicit ranking optimization. Recent work, such as RankGRPO (Zhu et al., 2025b), addresses this limitation by aligning LLM outputs with ranking specific rewards through Group Relative Policy Optimization.

Applying ranking optimized LLMs to financial asset allocation remains challenging because market signals are noisy and rankings must align with realized dynamics rather than static preferences. FinRankGRPO addresses this gap with a two-stage, reward driven framework for listwise financial asset ranking, grounding LLM reasoning in quantitative market performance through structured preference learning.

3 Method

3.1 Problem Formulation

We formulate financial portfolio management into a two-stage process (1) extracting semantic predictive rankings from unstructured text such as market news, and (2) projecting these rankings into valid portfolio weights.

Ranking Generation. Formally, given a set of N financial assets $\mathcal{A} = \{a_1, \ldots, a_N\}$, along with financial news X_{news} and a trading date X_{date}, the LLM first generates a textual output y:

$$y_{t+1} = \text{LLM}_\theta\big(\mathcal{T}(X_{\text{date}}(t), X_{\text{news}(t)}, \mathcal{A})\big) \quad (1)$$

where $\mathcal{T}(\cdot)$ represents the template we use to prompt LLM to generate ranking answers, shown in the Appendix A, θ represents the LLM parameter.

Portfolio Construction. We map the qualitative ranking y_{t+1} into continuous portfolio weights $\omega^*_{t+1} \in R^N$ through EP (Section 3.3).

3.2 Our Two-Stage Training Approach

Our method consists of two-stage training. The first stage is SFT. We finetune the model using a high quality CoT reasoning dataset. The second stage is Financial Asset Ranking via Group Relative Policy Optimization (FinRankGRPO). We propose a novel reinforcement learning objective based on Spearman correlation, specifically for the listwise financial asset ranking task. The whole framework is shown in Figure 2.

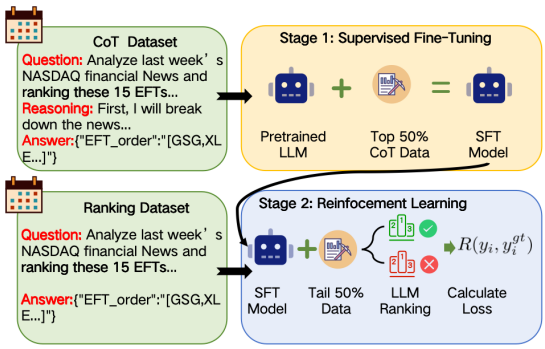

Figure 2: The two-stage construction framework of our FinRankGRPO, Stage 1 is SFT in high quality distill CoT datasets, Stage 2 is trained by our FinRankGRPO, which is a Spearman-related ranking reward.

3.2.1 Stage 1 Supervised Fine-Tuning

The SFT stage teaches the model to produce structured financial rankings from distilled CoT data generated by commercial models. For each prompt-output pair (x, y), we optimize:

$$\mathcal{L}_{\text{SFT}}(\pi_\theta) = -\log \pi_\theta\,(x, y) \quad (2)$$

where x is the input prompt and y is the target reasoning output. The fixed template $\mathcal{I}$ is shown in Appendix A.

3.2.2 Stage 2 FinRankGRPO

Although the SFT model improves reasoning capabilities, it still struggles to generate high quality rankings. To address this issue, the second stage

introduces FinRankGRPO, which redefines the process of ranking portfolios as a reinforcement learning problem with ranking optimization objectives.

The reward is based on the Spearman correlation between the model's predicted ranking and the true asset ranking. We choose Spearman over other methods for two key reasons. Firstly, unlike the Pearson correlation, which measures linear dependence on raw values and is sensitive to numerical scale, the Spearman correlation evaluates rank order independently of absolute magnitudes. Secondly, whereas metrics such as NDCG place greater importance on the top ranks, Spearman uniformly penalizes ranking errors across the entire list. This makes it better suited to obtaining a globally reliable ordering, which is critical in portfolio construction.

As a specialized adaptation of GRPO (Shao et al., 2024) for listwise financial asset ranking, FinRankGRPO directly optimizes the model via reward policy gradients, effectively aligning its generative outputs with explicit financial asset ranking objectives. For each training instance with input x, FinRankGRPO first samples a group of candidate outputs $\{o_g\}_{g=1}^{G}$ from the current policy π_θ. A rule-based extraction method $\mathcal{E}$ is then applied to each o_g to obtain a structured ranking list y_g. The sampling and extraction procedure is formalized as:

$$\begin{aligned} &\{o_g\}_{g=1}^{G} \sim \pi_\theta(\cdot \mid x), \\ &y_g = \mathcal{E}(o_g), \quad g = 1, \ldots, G. \end{aligned} \tag{3}$$

Each ranking is scored against the ground truth ordering y^{gt}:

$$R(y_g, y^{\text{gt}}) = \begin{cases} -1, & \text{if Error;} \\ \sigma\big(\rho_s(y_g, y^{\text{gt}})\big), & \text{if } \rho_s(y_g, y^{\text{gt}}) > 0; \\ \rho_s(y_g, y^{\text{gt}}), & \text{if } \rho_s(y_g, y^{\text{gt}}) \le 0, \end{cases} \tag{4}$$

where ρ_s is Spearman correlation and $\sigma(\cdot)$ amplifies positive rewards. Structural errors, including missing or hallucinated assets, receive a fixed penalty of -1.

The standardized advantage for each response is then computed based on the group's reward distribution:

$$\hat{A}_g = \frac{R(y_g, y^{\text{gt}}) - \mu_{\text{group}}}{\sigma_{\text{group}}}, \tag{5}$$

where μ_{group} and σ_{group} denote the mean and standard deviation of the rewards within the group, respectively.

Finally, the policy is optimized using the following objective:

$$\begin{aligned} \mathcal{J}_{\text{FinRankGRPO}} &= E_{x,\{o_g\}_{g=1}^{G}}\left[\frac{1}{G}\sum_{g=1}^{G}\frac{1}{|y_g|}\sum_{t=1}^{|y_g|}\ell_{g,t}(\theta)\right], \\ \ell_{g,t}(\theta) &= \min\Big(w_{g,t}(\theta)\hat{A}_{g,t}, \bar{w}_{g,t}(\theta)\hat{A}_{g,t}\Big). \end{aligned} \tag{6}$$

where $\bar{w}_{g,t}(\theta) = \text{clip}(w_{g,t}(\theta), 1-\epsilon, 1+\epsilon)$, ϵ is the clipping hyperparameter, and $w_{g,t}(\theta)$ is the importance ratio. We omit the KL term for readability.

Our method achieves direct ordinal alignment by optimizing rank correlation. This enables LLMs to produce robust and coherent asset rankings.

3.3 Integration with Entropy Pooling

To bridge the gap between discrete semantic rankings and continuous portfolio allocation, we use EP (Meucci, 2006, 2010) to convert LLM generated rankings into relative views and update the market prior into a posterior distribution used for portfolio construction.

3.3.1 Constructing View Constraints

If the LLM ranks asset i above asset j ($i \succ j$), we impose $E[R_i] \ge E[R_j]$. The ranking is encoded as:

$$\mathbf{P}E_{\mathbf{q}}[\mathbf{R}] \ge \mathbf{v}, \tag{7}$$

where $\mathbf{R}$ is the asset-return vector, $\mathbf{P}$ is the pick matrix, and $\mathbf{v} = \mathbf{0}$ for qualitative rankings. An example is provided in Appendix D.

3.3.2 Minimizing Relative Entropy

We estimate a posterior scenario distribution $\mathbf{q}^*$ that satisfies Eq. (7) while minimizing KL divergence from the empirical prior $\mathbf{p}$:

$$\begin{aligned} \mathbf{q}^* = \arg\min_{\mathbf{q}} &\sum_{k=1}^{K} q_k \ln\left(\frac{q_k}{p_k}\right) \\ \text{s.t.} \quad &\mathbf{P}E_{\mathbf{q}}[\mathbf{R}] \ge \mathbf{v}, \\ &\sum_{k=1}^{K} q_k = 1, \quad \mathbf{q} \ge \mathbf{0}. \end{aligned} \tag{8}$$

This posterior incorporates semantic views without parametric assumptions on returns.

3.3.3 Portfolio Optimization

Finally, we use posterior moments $\boldsymbol{\mu}_{\text{EP}} = E_{\mathbf{q}^*}[\mathbf{R}]$ and $\boldsymbol{\Sigma}_{\text{EP}} = \text{Cov}_{\mathbf{q}^*}[\mathbf{R}]$ to solve a long-only Maximum Sharpe Ratio problem with risk-free rate $r = 2.5\%$:

$$\omega^* = \arg\max_{\omega} \frac{\omega^\top \boldsymbol{\mu}_{\text{EP}} - r}{\sqrt{\omega^\top \boldsymbol{\Sigma}_{\text{EP}} \omega}} \quad \text{s.t.} \quad \mathbf{1}^\top \omega = 1, \quad \omega \geq \mathbf{0}. \tag{9}$$

The resulting weights ω^* dynamically concentrate capital on assets where the LLM predicts high relative outperformance, scaled by the risk estimates refined via EP.

4 Experiment

4.1 Datasets

We construct a weekly backtesting dataset for 2010–2025 over 15 highly liquid ETFs, including size indices, GICS sectors, and alternative assets. Market prices are obtained from Bloomberg[1], under our institution's academic license. Financial news headlines are collected from FNSPID (Dong et al., 2024), which follows CC BY 4.0, for 2010–2023 and The Wall Street Journal archive for 2024–2025[2]. To avoid lookahead bias, we use a chronological split, 2010–2019 for training and 2020–2025 for out-of-sample testing. The training set contains 2,516 weekly samples and is further divided into SFT and FinRankGRPO subsets; the test set contains 314 weekly samples.

We generate Chain-of-Thought (CoT) ranking supervision using GPT-4o-mini (OpenAI et al., 2024), Claude-3.5-Sonnet (Anthropic, 2024), and DeepSeek-Chat (DeepSeek-AI, 2025). For each weekly sample, models reason over historical price trends and recent news headlines, then output a ranked ETF list for the following week. Additional details are provided in Appendix B.

4.2 Basline

Our baseline includes both traditional finance models and LLM-based methods.

Traditional Finance Methods: Equal Weights (EW) is a non-parametric heuristic that allocates capital uniformly across all N assets ($w_i = 1/N$). Mean-Variance Optimization (MVO-SR) is the classical Markowitz framework. We adopt the Maximum Sharpe Ratio (SR) variant. Short-Term Momentum (Momentum)is a trend following strategy that ranks assets based on their realized returns over the preceding two week.

LLM-Based Methods: For Commercial Models, we evaluate via their APIs, including GPT-4o-mini (OpenAI et al., 2024), Claude-3.5-Sonnet (20241022) (Anthropic, 2024), and DeepSeek-Chat (V3.2) (DeepSeek-AI, 2025). For open source models, we utilize the Qwen2.5 (Team, 2024) model, covering three parameter scales, 0.5B, 1.5B, and 3B, including both the Base and Instruct variants.

4.3 Evaluation

We evaluate our proposed method using two distinct categories of metrics, rank-based metrics and financial performance metrics.

Rank-based metrics assess the quality of generated investment lists against the ground-truth ranking induced by future realized returns. We report Spearman's rank correlation (Spearman, ρ), Kendall's tau (Kendall, τ), and normalized discounted cumulative gain at k (NDCG@k), which are standard metrics for ranking evaluation.

Financial Performance Metrics assess the practical utility of the portfolios and assess the quality of output in real finance tasks. We report Annualized Return (Ann. Ret.), Annualized Volatility (Ann. Vol.) Sharpe Ratio (Sharpe), Maximum Drawdown (Max DD), Calmar Ratio (Calmar).

4.4 Main Results

Table 1 demonstrates that our proposed FinRankGRPO method, which is finetuned from the Qwen2.5-3B-Instruct model, achieves state-of-the-art performance. It outperforms traditional finance strategies and commercial large language models across a range of ranking and portfolio evaluation metrics.

FinRankGRPO works where standard GRPO fails. Although standard GRPO and FinRankGRPO are initialized from the same SFT checkpoint, standard GRPO provides little useful learning signal for financial ranking. Its exact match reward is too sparse. All sampled actions receive zero reward, and 35.14% of training steps have zero loss. FinRankGRPO addresses this issue by replacing exact-match feedback with a dense ranking reward, producing non-zero rewards for all sampled responses and non-zero loss in 99.68% of steps. This dense ordinal supervision enables

[1] https://www.bloomberg.com/
[2] https://www.wsj.com/news/archive/years

Model	Ranking Metrics				Portfolio Performance			
	Spear.	Kend.	N@3	N@5	Ann. Ret.	Sharpe	Max DD	Calmar
Ground Truth	1.000	1.000	1.000	1.000	1.049	7.655	0.138	7.618
Traditional Finance Baselines								
EW	–	–	–	–	0.085	0.302	0.356	0.239
MVO-SR	-0.007	-0.006	0.569	0.596	0.080	0.339	0.268	0.299
Momentum	-0.028	-0.021	0.544	0.581	0.041	0.107	**0.209**	0.195
LLM Prompting-Based Methods								
Deepseek-Chat	0.016	0.017	0.573	0.603	0.095	0.457	0.227	0.418
GPT-4o-mini	0.016	0.014	0.520	0.551	0.080	0.356	0.252	0.321
Claude	0.001	0.004	0.567	0.595	0.086	0.363	0.249	0.345
LLM Post-Training-Based Methods								
Qwen2.5-3B	0.002	0.001	0.032	0.033	0.070	0.289	0.266	0.276
+SFT	0.007	0.006	0.532	0.560	0.092	0.392	0.220	0.422
+GRPO	0.016	0.012	0.526	0.555	0.091	0.424	0.228	0.404
+FinRankGRPO	0.021	0.016	0.557	0.585	0.111	0.568	0.219	0.508
Qwen2.5-3B-Instruct	-0.001	-0.003	0.177	0.184	0.111	0.513	0.254	0.440
+SFT	0.019	0.015	0.522	0.552	0.092	0.444	0.249	0.375
+GRPO	0.020	0.016	0.517	0.549	0.092	0.445	0.237	0.402
+FinRankGRPO	**0.023**	**0.018**	**0.577**	**0.606**	**0.122**	**0.636**	0.211	**0.580**

Table 1: Main experimental results comparing traditional baselines, Commercial LLMs, and our fine-tuned models. Bold indicates the best performance among all methods, and underline indicates the second-best performance. All results represent the mean performance averaged over 5 independent trials. For Maximum Drawdown, a lower value is better.

effective policy optimization and leads to stronger portfolio performance.

Superior ranking performance. Our Qwen3B-Instruct FinRankGRPO model achieves the highest Spearman correlation and Kendall's tau , indicating more consistent and accurate asset ranking than all baselines. Notably, it surpasses both the best commercial LLM and traditional portfolio optimization approaches. This shows that our FinRankGRPO pipeline effectively learns the asset ranking from financial news.

Superior Portfolio Outcomes. In terms of real world investment performance, our method achieves the highest annualized return (0.122), exceeding the equal weight benchmark (0.085) and the leading commercial methods (Deepseek-Chat with 0.095). What's more, our method achieves the highest Sharpe(0.636), exceeding the leading commercial methods. These results confirm that improved ranking accuracy directly translates into superior portfolio returns.

The consistent gains in both ranking metrics and practical portfolio measures indicate that our training paradigm successfully closes the gap between pure ranking performance and actionable financial decision making. By directly optimizing for ranking-based rewards, FinRankGRPO ensures that the model's outputs are not only statistically sound but also economically meaningful in real world trading scenarios.

5 Analysis and Interpretability

5.1 The Impact of Base Model

To analyze the impact of model capacity and instruction tuning, we train Qwen2.5 models of different sizes using the same two-stage FinRankGRPO pipeline and evaluate both ranking accuracy and portfolio utility.

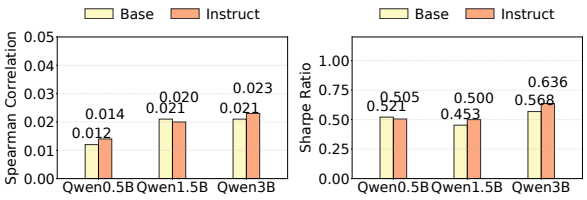

Figure 3: Performance of Different Based Models Trained with a Two Stage Framework. (Left) Spearman Correlation with Ground Truth Answers. (Right) Sharpe ratio on the Test Dataset.

Figure 3 shows that larger models generally produce better financial rankings and stronger portfolio performance, consistent with scaling law findings in general models (Kaplan et al., 2020). The

3B base model achieves a Spearman correlation of 0.021 and a Sharpe ratio of 0.568, outperforming the 0.5B model (0.012 Spearman and 0.521 Sharpe). Instruction tuning further improves alignment: Qwen2.5-3B-Instruct reaches a Spearman correlation of 0.023 and a Sharpe ratio of 0.636. These results suggest that both model capacity and instruction following ability are important for extracting reliable asset rankings from noisy financial news.

5.2 The impact of Ranking Metrics

To examine the effect of reward design, we compare FinRankGRPO with two alternative ranking rewards, NDCG@3 and Kendall's tau, using the same SFT checkpoint as initialization. This controlled setting isolates the reward metric from model capacity and supervised finetuning effects. As shown in Table 2, the Spearman-based objective performs better than both alternatives, suggesting that Spearman correlation better matches the structure of financial asset ranking.

Reward Metric	Spear.	Kendall	Sharpe	Max DD
NDCG@3	0.000	0.000	0.412	0.223
Kendall's Tau	0.000	0.002	0.441	0.220
FinRankGRPO	0.023	0.018	0.636	0.211

Table 2: Comparison of different ranking rewards using the same SFT-initialized base model.

Portfolio construction requires a coherent ordering over the full asset universe, not only accurate top ranked assets. NDCG@3 emphasizes the first few positions and therefore provides weak supervision for middle ranked and lower ranked assets, even though these assets still affect the relative-view constraints in Entropy Pooling. Kendall's tau measures pairwise consistency, but its pairwise objective is less aligned with the smooth global rank correlation needed for listwise allocation. In contrast, Spearman correlation evaluates the full list and provides a dense ordinal signal for learning stable asset preferences under noisy financial news.

5.3 The Impact of Two-Stage Training

Figure 4 illustrates the effectiveness of our two-stage training framework. For the 3B models, the experimental results show consistent improvement across training stages. Starting from the Original models, SFT raises key performance metrics, while our FinRankGRPO method further unleashes model potential, achieving the best overall performance (Spearman are 0.21 and 0.23 for Base and Instruct, respectively. Sharpe ratios are 0.568 and 0.636 for Base and Instruct, respectively). This two-stage progression demonstrates the complementary benefits of SFT for foundational improvement and FinRankGRPO for optimization refinement.

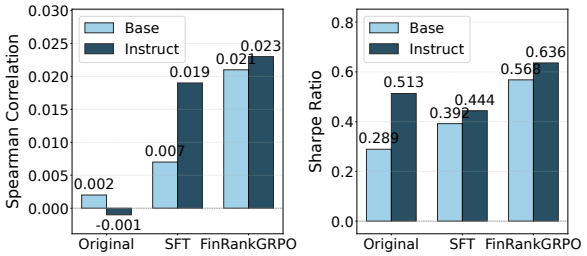

Figure 4: Performance of two-Stage training. (Left) Spearman correlation with ground-truth answers. (Right) Sharpe ratio on the test dataset.

5.4 The Ablation Study of Input Information

To validate that FinRankGRPO captures semantic signals from financial news rather than relying on superficial temporal patterns, we conduct an input ablation study by removing either trading date information or news content, as shown in Figure 5.

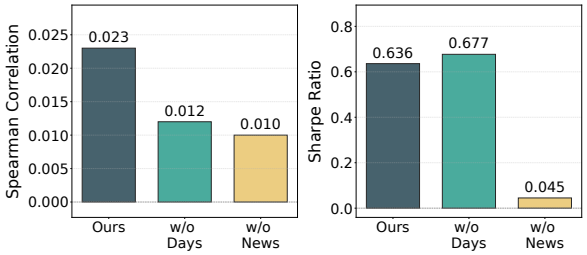

Figure 5: Performance of Input Ablation Study. (Left) Spearman correlation with ground-truth answers. (Right) Sharpe ratio on the test dataset.

Figure 5 shows that news information is crucial for portfolio performance, removing news content sharply reduces the Sharpe ratio from 0.636 to 0.045. In contrast, removing trading date information does not hurt Sharpe ratio and even slightly improves it from 0.636 to 0.677, while reducing the Spearman correlation from 0.023 to 0.012. This suggests that date information may introduce superficial temporal correlations that help ranking metrics but do not translate into better investment performance, whereas financial news provides the semantic signal needed for robust portfolio construction.

5.5 Contextual Faithfulness Analysis

To evaluate contextual faithfulness, we conduct a sentiment inversion experiment. We reverse the sentiment polarity of news summaries while preserving entities and event structures, using DeepSeek-Chat as the rewriting model. An example is shown in Appendix H. A faithful model should change its rankings under inverted news, whereas a memorization driven model should remain insensitive to the altered context.

Model	Original	Inverted	Diff
	Spear.	Spear.	Spear.
DeepSeek-Chat	0.016	-0.021	0.037
Qwen2.5-3B-Instruct	-0.001	0.006	-0.008
+SFT	0.019	-0.010	0.029
+FinRankGRPO	0.023	-0.011	0.034

Table 3: Rank correlation under original and sentiment-inverted news prompts (Spearman only). Larger differences indicate greater sensitivity to semantic direction.

Table 3 quantifies the dependency of each model on the input semantics. We observe that strong reasoning models, particularly DeepSeek and our FinRankGRPO, exhibit high contextual faithfulness, evidenced by the substantial degradation in performance under the inverted news prompts. Specifically, FinRankGRPO shows a Spearman drop of 0.034 (Δ), confirming that its decision boundary is dynamically shaped by the semantic polarity of the news.

5.6 Stress Test: the 2020 Recession Period

To evaluate whether our models genuinely understand market risk semantics rather than simply overfitting to historical patterns, we conduct a targeted analysis of the February-April 2020 recession period designated by the National Bureau of Economic Research. This period is characterized by extreme market recession, making it an ideal stress test for evaluating adaptive portfolio construction. We analyze model behavior across ranking metrics, scores, and backtest performance during the recession period.

As demonstrated in 4, FinRankGRPO effectively bridges the gap between semantic risk signals and defensive positioning. During the 2020 recession, it achieves the highest Spearman correlation and Sharpe ratio (0.4863). A case study on 2020-03-09 further validates this; while base models overlook macro warnings, FinRankGRPO prioritizes crisis-resilient assets like Gold (GLD) and Consumer Staples (XLP) and proves its capacity to ground semantic risks into actionable defensive constraints.

Model	Spear.	Kendall	Sharpe	Max DD
DeepSeek-Chat	0.051	0.039	0.193	<u>0.201</u>
GPT-4o-mini	0.042	0.050	0.386	0.211
Claude	0.050	0.046	<u>0.486</u>	**0.202**
Qwen2.5-3B-Instruct	0.049	**0.092**	-0.184	0.250
+SFT	<u>0.068</u>	0.054	-0.266	0.211
+FinRankGRPO	**0.070**	<u>0.055</u>	**0.486**	0.209

Table 4: Ranking and performance metrics during the 2020 recession period. Higher values are better for Spearman, Kendall, and Sharpe ratio; lower values are better for Max Drawdown. Bold indicates the best performance among all methods, and underline indicates the second-best performance.

6 Conclusion

Our FinRankGRPO is a novel approach for LLM-assisted asset allocation. Rather than directly using LLM to generate numerical estimates, we reformulate their role as extracting relative asset prospects from market news and producing listwise rankings. This design better matches the comparative reasoning strengths of LLMs while avoiding the numerical instability. We further connect the LLM output with Entropy Pooling. Besides, our method adopts a two-stage training procedure. The first stage uses supervised finetuning on distilled CoT reasoning data to teach structured financial ranking behavior, while the second stage applies FinRankGRPO with a Spearman-based reward to align generated rankings with realized market orderings. Empirical results demonstrate that FinRankGRPO outperforms traditional quantitative baselines and commercial models. Our work contributes a general perspective on adapting LLMs to classical financial decisions, which demonstrates that LLMs can complement established quantitative decision mechanisms.

Limitations

This work has two main limitations. First, our experiments are conducted only in English. Although English financial news covers a large portion of market information, it may not fully capture market signals expressed in other languages, especially for region specific assets and local financial events. Extending FinRankGRPO to multilingual financial corpora is an important direction for future work.

Second, our evaluation period spans 2020 to 2025, which may introduce potential lookahead

bias due to the pretraining data of modern LLMs. Although our ranking labels and portfolio evaluation are constructed chronologically, the base models may have been exposed to financial texts or market-related information from parts of the evaluation period during pretraining. Future work will further reduce this risk by using strictly time controlled model checkpoints, earlier out of sample periods.

Ethical Considerations

This work uses publicly available financial news and market data for research purposes. We do not use private personal information. The assets and market signals analyzed in this paper are based on open-source financial corpora and publicly observable market outcomes.

We used AI tools to assist with language polishing. All technical ideas, experimental design, results, and conclusions were completed and finalized by the authors.

Our study also constructs synthetic news with opposite market sentiment for analysis. These synthetic examples are used only as controlled research inputs to evaluate model behavior under sentiment perturbations. They are not intended to represent real market events, and should not be used for investment advice, public dissemination, or market manipulation.

References

Anthropic. 2024. The claude 3 model family: Opus, sonnet, haiku.

Fischer Black and Robert Litterman. 1992. Global portfolio optimization. *Financial analysts journal*, 48(5):28–43.

I-Chan Chiu and Mao-Wei Hung. 2024. Finance-specific large language models: Advancing sentiment analysis and return prediction with llama 2. *Pacific-Basin Finance Journal*.

DeepSeek-AI. 2025. Deepseek-v3.2: Pushing the frontier of open large language models.

Mohsen Dehghankar and Abolfazl Asudeh. 2025. Rank it, then ask it: Input reranking for maximizing the performance of llms on symmetric tasks. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 451–462.

Zihan Dong, Xinyu Fan, and Zhiyuan Peng. 2024. Fnspid: A comprehensive financial news dataset in time series. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Michael Mark Dowling and Brian M. Lucey. 2023. Chatgpt for (finance) research: The bananarama conjecture. *SSRN Electronic Journal*.

Frank J. Fabozzi. 2007. Robust portfolio optimization and management.

George Fatouros, Kostas C. Metaxas, John Soldatos, and Manos Karathanassis. 2025. Marketsenseai 2.0: Enhancing stock analysis through llm agents. *ArXiv*, abs/2502.00415.

Jingtong Gao, Bo Chen, Xiangyu Zhao, Weiwen Liu, Xiangyang Li, Yichao Wang, Wanyu Wang, Huifeng Guo, and Ruiming Tang. 2025. Llm4rerank: Llm-based auto-reranking framework for recommendations. In *Proceedings of the ACM on Web Conference 2025*, pages 228–239.

Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang, and Jiawei Zhang. 2023. Chat-rec: Towards interactive and explainable llms-augmented recommender system. *ArXiv*, abs/2303.14524.

Tian Guo and Emmanuel Hauptmann. 2024. Fine-tuning large language models for stock return prediction using newsflow. In *Conference on Empirical Methods in Natural Language Processing*.

Wenyue Hua, Shuyuan Xu, Yingqiang Ge, and Yongfeng Zhang. 2023. How to index item ids for recommendation foundation models. *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*.

Thomas Idzorek. 2007. A step-by-step guide to the black-litterman model: Incorporating user-specified confidence levels. In *Forecasting expected returns in the financial markets*, pages 17–38. Elsevier.

Jared Kaplan, Sam McCandlish, T. J. Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeff Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *ArXiv*, abs/2001.08361.

Ekaterina Khramtsova, Shengyao Zhuang, Mahsa Baktashmotlagh, and Guido Zuccon. 2024. Leveraging llms for unsupervised dense retriever ranking. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1307–1317.

Kemal Kirtac and Guido Germano. 2025. Leveraging llm-based sentiment analysis for portfolio optimization with proximal policy optimization. *SSRN Electronic Journal*.

Hoyoung Lee, Junhyuk Seo, Suhwan Park, Junhyeong Lee, Wonbin Ahn, Chanyeol Choi, Alejandro Lopez-Lira, and Yongjae Lee. 2025a. Your ai, not your view: The bias of llms in investment analysis. *Proceedings of the 6th ACM International Conference on AI in Finance*.

Youngbin Lee, Yejin Kim, Juhyeon Kim, Suin Kim, and Yongjae Lee. 2025b. Llm-enhanced black-litterman portfolio optimization.

Xiangyu Li, Xuan Yao, Guohao Qi, Fengbin Zhu, Kelvin J.L. Koa, Xiang Yao Ng, Ziyang Liu, Xingyu Ni, Chang Liu, Yonghui Yang, Yang Zhang, Wenjie Wang, Fuli Feng, Chao Wang, Huanbo Luan, Xiaofen Xing, Xiangmin Xu, Tat-Seng Chua, and Ke wei Huang. 2026. Findeepforecast: A live multi-agent system for benchmarking deep research agents in financial forecasting.

Zhaowei Liu, Xin Guo, Fangqi Lou, Lingfeng Zeng, Jinyi Niu, Zixuan Wang, Jiajie Xu, Weige Cai, Ziwei Yang, Xueqian Zhao, Chaojun Li, Sheng Xu, Dezhi Chen, Yun Chen, Zuo Bai, and Liwen Zhang. 2025. Fin-r1: A large language model for financial reasoning through reinforcement learning. *ArXiv*, abs/2503.16252.

Sedigheh Mahdavi, Jiating Chen, Pradeep Kumar Joshi, Lina Milena Huertas Guativa, and Upmanyu Singh. 2025. Integrating large language models in financial investments and market analysis: A survey. *ArXiv*, abs/2507.01990.

Harry Markowitz. 1952. Modern portfolio theory. *Journal of Finance*, 7(11):77–91.

Harry M. Markowitz. 1971. Portfolio selection.

Peter Joseph Mercurio, Yuehua Wu, and Hong Xie. 2020. An entropy-based approach to portfolio optimization. *Entropy*, 22(3):332.

Attilio Meucci. 2006. Beyond black-litterman in practice: A five-step recipe to input views on non-normal markets. *Available at SSRN 872577*.

Attilio Meucci. 2008. Fully flexible views: Theory and practice. *Bloomberg: Frontiers (Topic)*.

Attilio Meucci. 2010. Fully flexible views: Theory and practice. *arXiv preprint arXiv:1012.2848*.

Attilio Meucci and Marco Nicolosi. 2016. Dynamic portfolio management with views at multiple horizons. *Applied Mathematics and Computation*, 274:495–518.

Richard O Michaud. 1989. The markowitz optimization enigma: Is 'optimized'optimal? *Financial analysts journal*, 45(1):31–42.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, and Janko Altenschmidt et al. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Ronak Pradeep, Sahel Sharifymoghaddam, and Jimmy Lin. 2023. Rankvicuna: Zero-shot listwise document reranking with open-source large language models. *ArXiv*, abs/2309.15088.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junmei Song, Mingchuan Zhang, Y. K. Li, Yu Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *ArXiv*, abs/2402.03300.

Amit Sharma, Hua Li, Xue Li, and Jian Jiao. 2024. Optimizing novelty of top-k recommendations using large language models and reinforcement learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5669–5679.

Weiwei Sun, Lingyong Yan, Xinyu Ma, Pengjie Ren, Dawei Yin, and Zhaochun Ren. 2023. Is chatgpt good at search? investigating large language models as re-ranking agent. *ArXiv*, abs/2304.09542.

Qwen Team. 2024. Qwen2.5: A party of foundation models.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *ArXiv*, abs/2303.17564.

Mengxi Xiao, Zihao Jiang, Lingfei Qian, Zhengyu Chen, Yueru He, Yijing Xu, Yuechen Jiang, Dong Li, Ruey-Ling Weng, Min Peng, Jimin Huang, Sophia Ananiadou, and Qianqian Xie. 2025. Retrieval-augmented large language models for financial time series forecasting.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. 2024a. Tradingagents: Multi-agents llm financial trading framework. *ArXiv*, abs/2412.20138.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. 2024b. Tradingagents: Multi-agents llm financial trading framework. *arXiv preprint arXiv:2412.20138*.

Hongyang Yang, Xiao-Yang Liu, and Chris Wang. 2023. Fingpt: Open-source financial large language models. *ArXiv*, abs/2306.06031.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Rong Liu, Zhenyu Cui, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. 2024a. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *ArXiv*, abs/2407.06567.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan Suchow, Zhenyu Cui, Rong Liu, and 1 others. 2024b. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Advances in Neural Information Processing Systems*, 37:137010–137045.

Junjie Zhang, Ruobing Xie, Yupeng Hou, Wayne Xin Zhao, Leyu Lin, and Ji rong Wen. 2023. Recommendation as instruction following: A large language

model empowered recommendation approach. *ACM Transactions on Information Systems*, 43:1 – 37.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, and 1 others. 2024a. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th acm sigkdd conference on knowledge discovery and data mining*, pages 4314–4325.

Yongfeng Zhang, Zhiwei Liu, Qingsong Wen, Linsey Pang, Wei Liu, and Philip S Yu. 2024b. Ai agent for information retrieval: Generating and ranking. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 5605–5607.

Jie Zhu, Qian Chen, Huaixia Dou, Junhui Li, Lifan Guo, Feng Chen, and Chi Zhang. 2025a. Dianjin-r1: Evaluating and enhancing financial reasoning in large language models. *ArXiv*, abs/2504.15716.

Yaochen Zhu, Harald Steck, Dawen Liang, Yinhan He, Vito Claudio Ostuni, Jundong Li, and Nathan Kallus. 2025b. Rank-grpo: Training llm-based conversational recommender systems with reinforcement learning. *ArXiv*, abs/2510.20150.

Yaochen Zhu, Chao Wan, Harald Steck, Dawen Liang, Yesu Feng, Nathan Kallus, and Jundong Li. 2025c. Collaborative retrieval for large language model-based conversational recommender systems. *Proceedings of the ACM on Web Conference 2025*.

Yaochen Zhu, Liang Wu, Qi Guo, Liangjie Hong, and Jundong Li. 2024. Collaborative large language model for recommender systems. In *Proceedings of the ACM Web Conference*, pages 3162–3172.

A Prompt

As shown in Figure 6, we utilize a summarization prompt that instructs the model to distill a week's worth of raw Nasdaq headlines into a concise set of thematic market signals.

To generate the training data for the SFT and reinforcement learning stages, we design the prompt that enforces models to generate JSON format, shown in 7

To generate the reverse sentiment news, we design a prompt to converse the original news summary

B Dataset Details

Our investment universe contains 15 liquid ETFs that represent broad asset classes rather than individual stocks. This design makes the task closer to institutional asset allocation and requires the model to connect macroeconomic narratives with sector and asset class level implications. The full universe is shown in Table 5.

Prompt for News Summary

Instruction:Analyze today's NASDAQ financial news title and write a 200-word English analysis assessing impact on these 15 ETFs CORE SECTORS:

CORE SECTORS:
- Technology (XLK)
- Health Care (XLV)
- Financials (XLF)
- Consumer Discretionary (XLY)
- Consumer Staples (XLP)
- Industrials (XLI)
- Energy (XLE)
- Utilities (XLU)
- Materials (XLB)
- Real Estate (IYR)

MARKET CAP SEGMENTS:
- Large Cap (IVV)
- Mid Cap (IJH)
- Small Cap (IJR)

COMMODITIES:
- Gold (GLD)
- Broad Commodities (GSG)

Last Week's NASDAQ Financial News Titles (each title on a new line):{titles_text}

Figure 6: Prompt for Summary News Title.

Dataset construction consists of four steps. First, we obtain ETF market prices from Bloomberg and collect financial news headlines from two public sources: FNSPID (Dong et al., 2024) for 2010–2023 and The Wall Street Journal archive for 2024–2025. For each trading day, headlines from the preceding week are aggregated, duplicated with SimHash using a Hamming distance threshold of 25, and summarized into asset focused market digests using DeepSeek-Chat. The summarization prompt is shown in Appendix A.

Second, we use a chronological split to reduce look-ahead bias: 2010–2019 for training and 2020–2025 for out-of-sample testing. The training set contains 2,516 weekly samples, and the test set contains 314 weekly samples. Training samples are produced with a rolling weekly window, while test samples are non-overlapping calendar weeks to simulate realistic out of sample test.

Third, we generate CoT ranking data with GPT-4o-mini (**?**), Claude-3.5-Sonnet (**?**), and DeepSeek-

Prompt for Generate Finance Ranking List

Mission Objective
Analyze last week's NASDAQ financial news and rank these 15 ETFs by expected next week's return (HIGHEST to LOWEST).

ETF List to Rank
CORE SECTORS:
- XLK (Technology)
- XLV (Health Care)
- XLF (Financials)
- XLY (Consumer Discretionary)
- XLP (Consumer Staples)
- XLI (Industrials)
- XLE (Energy)
- XLU (Utilities)
- XLB (Materials)
- IYR (Real Estate)

MARKET CAP SEGMENTS:
- IVV (Large Cap)
- IJH (Mid Cap)
- IJR (Small Cap)

COMMODITIES:
- GLD (Gold)
- GSG (Broad Commodities)

Context Information
- Trading Date: {{ trading_date }}
- Last Week's NASDAQ News Analysis:
{{ news_summary }}

Output Instructions
1. Think step-by-step about news impact.
2.After your reasoning, your final answer must be **ONLY** a valid JSON object in the exact format shown in the **Example Format** below.
3. The `EFT_order` list must contain **all 15 ETF codes** exactly once (no duplicates).

Example Format:
```json
{"EFT_order": ["ETF1", "ETF2", "ETF3", ...,
"ETF15"]}
```"""

Figure 7: Prompt for Generate Finance Ranking List

Prompt for Sentiment Reversal

You are a professional linguist. Your task is to reverse the sentiment and content of news summaries while keeping the structure and topics.

IMPORTANT: Only output the reversed news summary. Do NOT add any explanations or notes in parentheses.

Task: Reverse financial direction in news sentences by converting upward trends to downward trends and vice versa.

1. Identify: Locate all phrases in the given news sentence that indicate an increase (e.g., rose, grew, climbed, surged, up, gained, higher) or a decrease (e.g., fell, dropped, declined, plunged, down, lost, lower) in asset value or financial performance.

2. Replace: Convert increase-related phrases to their opposite decrease-related phrases, and convert decrease-related phrases to their opposite increase-related phrases. Use appropriate antonyms (e.g., "rose" → "fell", "plummeted" → "soared").

3. Output: Preserve all other text, names, numbers, and structures exactly as they appear in the original sentence. Provide only the final modified sentence without any explanations or additional text.

Original news summary:
{news_summary}

Reversed news summary:"""

Figure 8: Prompt for Sentiment Inversion Protocol

Category	Tickers
Indices (Size)	IVV (S&P 500), IJH (Mid-Cap), IJR (Small-Cap)
Sectors	XLB (Materials), XLE (Energy), XLF (Financials), XLI (Industrials), XLK (Information Technology), XLP (Consumer Staples), XLU (Utilities), XLV (Health Care), XLY (Consumer Discretionary)
Alternatives	GLD (Gold), GSG (Commodities), IYR (Real Estate)

Table 5: Investment universe of 15 ETFs.

Chat (DeepSeek-AI, 2025). Each output must follow the JSON format specified in invalid outputs are discarded and replaced with an empty list. For each weekly sample, we retain the response with the highest Spearman correlation against the ground truth return ranking. Finally, to support the FinRankGRPO pipeline, the training samples are divided into SFT and FinRankGRPO subsets. For each day, we retain the response with the highest Spearman correlation against the ground-truth return ranking. Finally, to support the two stage pipeline, samples within each calendar year are sorted by Spearman correlation: the top 50% form the SFT subset, and the bottom 50% form the more challenging FinRankGRPO subset. This yearly partition preserves exposure to different market regimes in both stages.

C Experiment Setup

We present the experiment setup for the FinRankGRPO experiments across both models. All experiments were conducted using full-parameter training on 3 NVIDIA L20 GPUs with mixed precision (bfloat16) to optimize computational efficiency. Our torch version is 2.9.1, transformers is 4.57.1, trl is 0.9.6. Our model include Qwen0.5-3B, Qwen0.5-3B-Instruct, Qwen0.5-3B, Qwen1.5-3B-Instruct,Qwen2.5-3B,Qwen2.5-3B-Instruct.

SFT. In this stage, the model is trained for 3 epochs using a learning rate of 10^{-4} about 1 hours. We adopt a cosine learning rate scheduler with 10% warmup steps. Training is conducted with a per-device batch size of 1 and gradient accumulation over 4 steps. This configuration balances memory constraints with stable optimization, ensuring consistent and efficient fine-tuning across the dataset.

FinRankGRPO. Subsequently, the models were trained using the FinRankGRPO framework for 1 epoch, about 12 hours. This phase employed a learning rate of 3×10^{-6}, also with cosine decay and a 10% linear warmup. Each device uses a batch size of 1 and gradient accumulation over 4 steps. During each training iteration, prompts with a maximum length of 1024 tokens, and 4 responses were sampled per prompt for ranking based policy optimization. For training, the sampling parameters were temperature = 1.0, top-k = 50, and top-p = 1.0.

D Entropy Pooling Example

For a universe with XLF, GLD, and XLE, an LLM ranking XLF $\succ$ GLD $\succ$ XLE implies $E[R_{XLF}] \geq E[R_{GLD}]$ and $E[R_{GLD}] \geq E[R_{XLE}]$. The corresponding pick matrix is:

$$\underbrace{\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}}_{\mathbf{P}} \begin{bmatrix} E[R_{XLF}] \\ E[R_{GLD}] \\ E[R_{XLE}] \end{bmatrix} \geq \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\mathbf{v}}. \quad (10)$$

E Portfolio Optimization

In this study, we implement two distinct portfolio optimization approaches: one based on the views generated by LLMs and the Others use traditional mean-variance optimization as a baseline. The objective is to compare the performance of LLM-based portfolios against a benchmark portfolio optimized using conventional technique

F Detailed Statistics

1. **Annualized Return (Ann. Ret.)**: The geometric average amount of money earned by an investment each year over a given time period.
2. **Annualized Volatility (Ann. Vol.)**: A statistical measure of the dispersion of returns, representing risk.
3. **Sharpe Ratio**: Measures risk-adjusted performance. A higher Sharpe ratio indicates better returns for the same level of risk.
4. **Maximum Drawdown (Max DD)**: The maximum observed loss from a peak to a trough of a portfolio, before a new peak is attained.
5. **Calmar Ratio**: A measure of risk-adjusted return based on the Maximum Drawdown.

For the backtest evaluation, we perform weekly portfolio rebalancing based on the generated weights. The environment is configured to simulate realistic trading conditions. We impose a transaction fee of 5 basis points (0.05%) per turnover

to account for commissions and bid-ask spreads. Trades are assumed to be executed at the weekly closing price. Given the high liquidity of the selected top-tier ETFs, we assume zero slippage for the execution. Furthermore, management fees are excluded from the calculation to strictly isolate the alpha generation capability of the active ranking strategy.

G Recession Regime

Table 6 show the model answer for recession Regime

Market Regime: Recession *News: China's Premier Says Coronavirus Outbreak Control Remains at Crucial Stage; JD.com Trades Higher On Q4 Results, Announces CFO Succession; Fed Cuts Rates By 0.5%; Daily Markets: Investors Shrug Off Fed's Emergency Rate Cut; Apple May Soon Stop Using Intel Chips in Macs . . .* *Ground-truth Ranking:* "[GSG, **GLD**, **XLP**, XLV, XLK, IVV, XLF, XLB, XLI, XLY, XLE, IJR, IJH, XLU, IYR]"
Commercial (GPT-4o-mini) *Ranking:* "[XLK, **GLD**, XLV, **XLP**, XLU, IVV, IJH, XLI, XLB, IYR, XLY, XLF, IJR, XLE, GSG]"
Qwen2.5-3B-instruct (Base Model) *Ranking:* "[IVV, XLK, XLE, XLF, XLY, **XLP**, GSG, XLV, XLI, XLU, IYR, **GLD**, XLB, IJR, IJH]"
Qwen2.5-3B-instruct (SFT) *Ranking:* "[XLK, **GLD**, XLV, **XLP**, XLU, IVV, XLI, XLB, IYR, XLF, IJH, XLY, IJR, GSG, XLE]"
Qwen2.5-3B-instruct (FinRankGRPO Ours) *Ranking:* "[XLK, **GLD**, XLV, IVV, **XLP**, XLU, IJH, XLI, XLB, IYR, IJR, XLY, XLF, XLE, GSG]"

Table 6: Models' predicted ranking list on asset returns for the second week in Mar. 2020 of recession regime and the ground truth ranking.

H Sentiment Polarity Invertion

The table is shown in 7

I Portfolio Net Value Comparison

The table is shown in 9

Original News Summary	Sentiment-Reversed News Summary
Based on the provided news titles from December 30, 2019, the dominant themes are mixed corporate earnings, sector-specific developments, and geopolitical tensions, against a backdrop of modest market **weakness** indicated by a **declining** Dow. The impact across the 15 specified ETFs will be varied and largely indirect. **Sector ETFs:** The news is most directly impactful for **Technology (XLK)**, with **positive** updates from Netflix, NVIDIA, and Tesla's Gigafactory, though tempered by notable **outflows** in semiconductor-related funds. **Consumer Discretionary (XLY)** faces crosscurrents: **positive** Chinese EV news (Nio) contrasts with retail **distress** (Gap, Signet Jewelers, SPDR S&P Retail ETF **outflows**). **Consumer Staples (XLP)** shows clear **negative** pressure, highlighted by repeated headlines of significant fund **outflows**. **Energy (XLE)** may see volatility due to headlines on political **unrest** impacting oil and a specific stake **sale** (TC Energy), while **Health Care (XLV)** gets a mild **boost** from FDA **approval** news (Myriad). Other sectors like Industrials, Financials, and Utilities lack direct catalysts. **Market Cap ETFs:** The blend of large-cap tech strength and broader retail/industrial **sluggishness** suggests a neutral to slightly **negative** impact on **Large Cap (IVV)**. The **underperformance** of specific retailers and transports (Union Pacific) points to potential **headwinds** for **Small Cap (IJR)** and **Mid Cap (IJH)**, which are more exposed to domestic consumer and industrial cycles. **Commodity ETFs:** Geopolitical tension from the Iran headline provides a classic short-term **bid** for **Gold (GLD)** as a **safe-haven** asset. **Broad Commodities (GSG)** may see **muted** effects, balancing potential oil volatility with other commodity news like China's pork reserve release. Overall, the news flow supports a **defensive** tilt, favoring sectors with idiosyncratic **positive** news (Tech, Health Care) and **safe-haven** assets (Gold), while highlighting **weakness** in consumer-facing segments.	Based on the provided news titles from December 30, 2019, the dominant themes are mixed corporate earnings, sector-specific developments, and geopolitical tensions, against a backdrop of modest market **strength** indicated by a **rising** Dow. The impact across the 15 specified ETFs will be varied and largely indirect. **Sector ETFs:** The news is most directly impactful for **Technology (XLK)**, with **negative** updates from Netflix, NVIDIA, and Tesla's Gigafactory, though tempered by notable **inflows** in semiconductor-related funds. **Consumer Discretionary (XLY)** faces crosscurrents: **negative** Chinese EV news (Nio) contrasts with retail **resilience** (Gap, Signet Jewelers, SPDR S&P Retail ETF **inflows**). **Consumer Staples (XLP)** shows clear **positive** pressure, highlighted by repeated headlines of significant fund **inflows**. **Energy (XLE)** may see **stability** due to headlines on political **calm** impacting oil and a specific stake **purchase** (TC Energy), while **Health Care (XLV)** gets a mild **setback** from FDA **rejection** news (Myriad). Other sectors like Industrials, Financials, and Utilities lack direct catalysts. **Market Cap ETFs:** The blend of large-cap tech **weakness** and broader retail/industrial **strength** suggests a neutral to slightly **positive** impact on **Large Cap (IVV)**. The **outperformance** of specific retailers and transports (Union Pacific) points to potential **tailwinds** for **Small Cap (IJR)** and **Mid Cap (IJH)**, which are more exposed to domestic consumer and industrial cycles. **Commodity ETFs:** Geopolitical **calm** from the Iran headline provides a classic short-term **decline** for **Gold (GLD)** as a **risk-on** asset. **Broad Commodities (GSG)** may see **pronounced** effects, balancing potential oil **stability** with other commodity news like China's pork reserve **withdrawal**. Overall, the news flow supports an **aggressive** tilt, favoring sectors with idiosyncratic **negative** news (Tech, Health Care) and **risk-on** assets (Gold), while highlighting **strength** in consumer-facing segments.

Table 7: Example of sentiment reversal transformation (Trading Date: 2020-01-06). Bold text highlights inverted sentiment terms, demonstrating systematic polarity reversal while preserving factual structure.

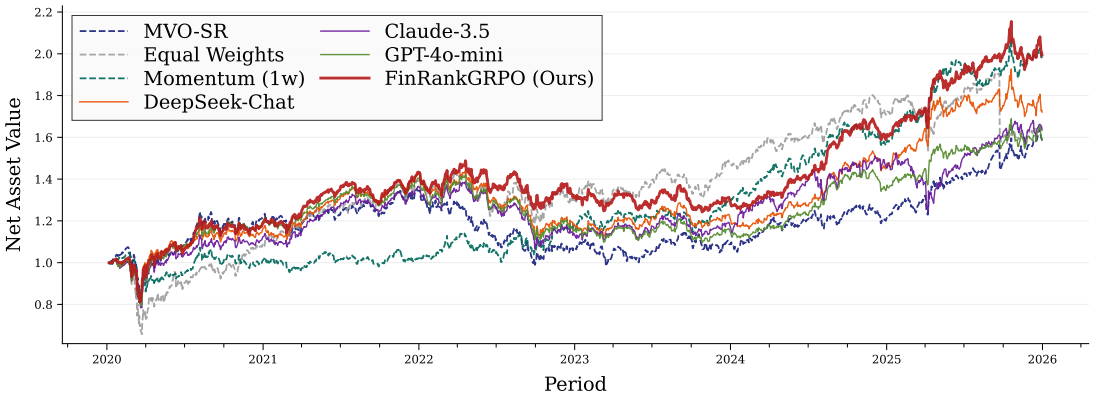

Figure 9: Portfolio net values comparison from 2020 to 2025. Our proposed method demonstrates superior performance compared to traditional baselines (MVO-SR, Momentum, Equal Weights) and LLM Ranking with EP approaches (DeepSeek-Chat, Claude, GPT-4o-mini).