FinZero: Launching Multimodal Financial Time-Series Reasoning

Yanlong Wang
Tsinghua University
wangyanl21@mails.tsinghua.edu.cn

Jian Xu
Tsinghua University
jianxu20@mails.tsinghua.edu.cn

Abstract

Financial time series forecasting is both highly significant and challenging. Previous approaches typically standardized time series data before feeding it into forecasting models, but this encoding process inherently leads to a loss of important information. Moreover, past time series models generally require fixed numbers of variables or lookback window lengths, which further limits the scalability of time series forecasting. Additionally, the interpretability of predictions and the uncertainty in forecasting remain areas requiring further research, as these factors directly impact the reliability and practical value of predictions. To address these issues, we first constructed a diverse financial image-text dataset (FVLDB) with over 10,000 samples. We developed the Uncertainty-adjusted Group Relative Policy Optimization (UARPO) method to enable the model not only output predictions but also assess and analyze the uncertainty of those predictions. We then proposed FinZero, a multimodal pre-trained model finetuned by UARPO to perform reasoning, prediction, and analytical understanding on the FVLDB financial time series. Extensive experiments validate that FinZero exhibits strong adaptability and scalability. After fine-tuning with UARPO, FinZero achieves an approximate 13.48% improvement in prediction accuracy over GPT-4o in the high-confidence group, demonstrating the effectiveness of reinforcement learning fine-tuning in multimodal large models, including in financial time series forecasting tasks.

1 Introduction

The field of time series forecasting has garnered increasing attention, as time-series data is widely present in various real-world industries (e.g., transportation, weather, power, finance, etc.). Extracting future trends from historical time-series information holds significant practical value. Among these, financial time series exhibit more distinctive characteristics as they are influenced by more complex factors; the asset price movements are shaped by a broad range of external macro- and micro-level influences, as well as the interplay between buyers and sellers in determining transaction prices. This implies that, in such a game-theoretic environment, any discernible patterns or identifiable features (e.g., the pronounced periodicity seen in transportation or power time series) tend to diminish once traders recognize and exploit them for profit. This "adaptive" nature of markets leads to the inability of historical patterns to fully replicate in the future. Predicting such time series is undoubtedly highly challenging. However, even marginal improvements in forecasting performance can yield substantial impacts, particularly in high-frequency trading scenarios.

To improve time-series forecasting performance, including financial time series such as exchange rate prediction, several specialized models have been designed. For example, Autoformer[26] uses a series decomposition network block to enhance the modeling of complex temporal structures; ModernTCN[18] proposes a purely convolutional architecture specifically designed for time series; FTS-Diffusion[10] constructs a dedicated generative learning framework to address the irregularity and scale invariance of financial data; SoftCLT[14] introduces a contrastive learning method tailored

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

for time-series data. Meanwhile, in recent years, the application of large pre-trained models in time-series forecasting has gained increasing attention. TEMPO[3] proposes a new decomposition method for learning time-series models by fine-tuning a language model; DAM[5] is a unified foundation forecasting model designed to efficiently and interpretably predict across multiple domains and time-series data.

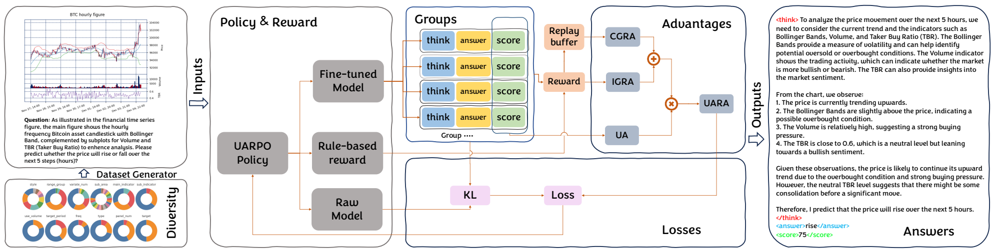

Figure 1: The Overall Pipeline of FinZero.

However, several challenges remain unresolved. First, most current time series models require standardization to transform the data into a numerical range that the model can process, such as the normalization techniques used in RevIN[13]. This inevitably leads to the loss of partial information from the original values. Second, patch-based processing is commonly adopted, but this may not fully align with the size and location of critical features. Besides, time series models usually operate with several fixed configurations, such as lookback window size, variable types and quantities, and data frequency; these significantly limit their generalizability. Although some pre-trained time-series models (e.g., DAM[5]) have partially addressed this issue, the advanced reasoning capabilities of large models have not yet been fully leveraged in time series applications. Moreover, interpretability of reasoning and uncertainty quantification in forecasting results remain critical yet understudied challenges.

To address the aforementioned challenges, we have abandoned traditional model architectures that process raw time series values and instead transformed the original time series into image compositions. Leveraging reinforcement learning fine-tuning, we enhance the visual reasoning capabilities of multimodal large model (MLM). Our focus is on financial time-series trend prediction and reasoning tasks. To support this, we constructed the FVLDB dataset, comprising over 10,000 financial time series image-text pairs. To ensure dataset diversity, we performed stratified sampling across multiple dimensions, including asset types, prediction task categories, historical sequence lengths and frequencies, time-series indicator varieties, and image styles. To tackle the inherent uncertainty and non-stationarity in financial time-series forecasting, we propose the Uncertainty-Adjusted Relative Policy Optimization (UARPO) method. UARPO evaluates both intra-group relative advantage (IGRA) (performance within a group) and cross-group relative advantage (CGRA) (performance between groups over a recent window). Additionally, it adjusts advantage levels based on prediction uncertainty (Uncertainty-Adjusted Relative Advantage, UARA).

In this work, we propose the FinZero model, as illustrated in Figure 1, which fine-tunes 3B-parameter multimodal large model via the UARPO method in the FVLDB dataset, which enables MLM to explicitly account for prediction uncertainty. Comparative experiments with GPT-4 show a 13.48% improvement in prediction accuracy in the high-confidence group, validating the effectiveness of RL-based cross-modal fine-tuning for financial time-series forecasting and reasoning. By providing confidence score and reasoning traces, FinZero helps users better understand model predictions and their rationale, ultimately supporting more informed financial decision-making, making it particularly valuable for real-world financial applications where risk assessment is paramount.

2 Methods

2.1 Uncertainty Adjusted Related Policy Optimization

The GRPO[22] is employed to fine-tune the DeepSeek-R1[6]. As an improvement over the PPO[21], GRPO eliminates the need for an additional model as a policy model (as required by methods like PPO) and leverages Group Relative Advantage sampled from multiple outputs within a group, thereby avoiding the necessity for extra value function approximation. GRPO primarily focuses on the relative advantages among multiple outputs within each sample group, while other methods like REINFORCE++[9] utilize discounted cumulative rewards to construct advantage variations that reflect the training process, which helps improve training stability. Additionally, how to reflect the uncertainty in model inference results holds significant importance, as it aids decision-making by assessing the confidence level of reasoning outcomes.

Based on the above, we propose the UARPO algorithm, which introduces two key enhancements. 1.Under the same prediction target, a multidimensional advantage function combining In-Group Relative Advantage (IGRA) within samples and Cross-Group Relative Advantage (CGRA) across groups; 2.Construction of an uncertainty function (UA) based on the model's inference confidence scores, ultimately forming Uncertainty-Adjusted Relative Advantage (UARA). The optimization objective can be expressed as Equation 1

$$\begin{aligned} J_{\text{UARPO}}(\theta) = \mathbb{E}\left[q \sim P(Q), \{o_i\}_{i=1}^{G} \sim \pi_{\theta_{\text{old}}}(O|q), \tau \in \mathcal{T}\right] \Bigg\{ \frac{1}{G} \sum_{i=1}^{G} \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \Bigg[\min \Bigg\{ \frac{\pi_\theta(o_{i,t}|q, o_{i,<t})}{\pi_{\theta_{\text{old}}}(o_{i,t}|q, o_{i,<t})} \\ \left(\hat{A}_{i,t}^{I} + \hat{A}_{t}^{S_\tau}\right) \hat{U}_{i,t}, \text{clip}\left(\frac{\pi_\theta(o_{i,t} \mid q, o_{i,<t})}{\pi_{\theta_{\text{old}}}(o_{i,t} \mid q, o_{i,<t})}, 1-\varepsilon, 1+\varepsilon\right) \left(\hat{A}_{i,t}^{I} + \hat{A}_{t}^{S_\tau}\right) \hat{U}_{i,t} \Bigg\} \\ - \beta D_{\text{KL}}\left[\pi_\theta || \pi_{\text{ref}}\right] \Bigg] \Bigg\} \end{aligned} \tag{1}$$

$$\hat{A}_{i,t}^{I} \triangleq \widetilde{r}_i = \frac{r_i - \text{mean}(\mathbf{r})}{\text{std}(\mathbf{r})} \tag{2}$$

$$\hat{A}_{t}^{S_\tau} \triangleq \widetilde{s_t^\tau} = \frac{s_t^\tau - \text{mean}(\mathbf{s}_{\mathbf{t-1,t}}^{\tau})}{\text{std}(\mathbf{s}_{\mathbf{t-1,t}}^{\tau})} \tag{3}$$

Where π_θ and $\pi_{\theta_{old}}$ are the current and old policy models, q and o are questions and outputs sampled from the question dataset and the old policy $\pi_{\theta_{old}}$, respectively. ε is a clipping-related hyper-parameter introduced in PPO for stabilizing training. $\hat{A}_{i,t}^{I}$ represents the in-group relative advantage as in GRPO, where $\mathbf{r} = [r_0, r_1, \ldots, r_i, \ldots, r_G]$, $\hat{A}_{t}^{S_\tau}$ represents the cross-group relative advantage where $\mathbf{s}_{\mathbf{t-1,t}}^{\tau} = [s_{t-l}, s_{t-l+1}, \ldots, s_t|\tau]$ and $s_t^\tau = \frac{1}{G}\sum_{i=1}^{G} r_{i,t}^{\tau}$, which indicate the advantage of the current group's overall performance relative to the average performance over multiple steps in a recent window period under the same prediction objective. $\mathbf{s}_{\mathbf{t-1,t}}^{\tau}$ is a group consisting of multiple steps with window length l. $\hat{U}_{i,t} \triangleq \alpha \cdot \frac{\text{score}-\text{const}}{100}$ is the uncertainty adjustment function, and α denotes an adjustable coefficient. The algorithmic iterative process can be described as Algorithm 1.

2.2 Rewards and Uncertainty

- **Accuracy Reward** Prediction accuracy is commonly used to evaluate the performance of reinforcement learning models and construct loss functions. Specifically, it measures the consistency between the model's predictions and the ground-truth outcomes (rise/fall) of each sample.
- **Completion Length Reward** Previous works have found that text length expansion occurs in large model RL reasoning, which is helpful for improving reasoning time and enabling complex reasoning. Therefore, we provide this type of reward. Specifically, when the text reasoning length is no more than 200 tokens, a gradually increasing reward is offered.
- **Format Reward** We add this reward to help the model learn the target output format during reinforcement learning fine-tuning.
- **Confidence Score** Prior works ([16, 27]) have explored the feasibility and methods for large models to learn task uncertainty. Given the high uncertainty inherent in financial decision-making—where uncertainty analysis is critical for model development and real-world use—we integrate model reasoning uncertainty into reinforcement learning fine-tuning. During each image-text reasoning process, the model outputs a confidence score based on the input information and its reasoning. This score quantifies the model's uncertainty about its reasoning result for the given task, enabling it to learn problem difficulty and uncertainty through training.

Table 1: Main Results of Model Accuracy Comparison.

	Volitality ACC (%)				Price ACC (%)			
Model	5	21	63	Avg	5	21	63	Avg
Naive	48.54	46.23	48.46	47.75	50.00	52.04	50.00	50.68
Qwen2.5-VL-3B	46.67	45.69	50.51	47.62	54.20	51.64	52.54	52.79
Qwen2.5-VL-7B	50.49	43.64	51.16	48.43	<u>55.55</u>	51.91	51.14	53.53
GRPO	53.68	<u>54.86</u>	52.15	<u>53.56</u>	53.24	<u>53.63</u>	<u>53.76</u>	<u>53.54</u>
GPT-4o	<u>54.28</u>	48.26	**53.38**	51.97	**56.16**	51.22	51.14	52.84
FinZero	**56.31**	**65.74**	<u>52.93</u>	**58.33**	54.52	**56.31**	**65.88**	**58.90**

3 Experiments

Setup. The FinZero utilize the Qwen2.5-VL-3B model as the backbone and fine-tuned it directly on the FVLDB dataset with the UARPO algorithm. For baselines, we selected the original Qwen2.5-VL-3B model, the Qwen2.5-VL-7B model, and the larger-scale GPT-4o. Additionally, we also fine-tuned Qwen2.5-VL-3B with GRPO, and also constructed a Naive Model, which extends the trend of the past period of time. The Adam optimizer was adopted with a learning rate of 1e-6, and the fine-tuning process ran for two epochs. All experiments were conducted on a server equipped with two 80G Nvidia A100 GPUs.

Results As shown in Figure 3, the model's rewards continuously increase during the UARPO fine-tuning process: the format reward and completion length reward rise rapidly in the early stage of training and then stabilize, while the accuracy reward also increases steadily with training; meanwhile, the loss value decreases consistently. The prediction performance of the fine-tuned model on the test set is presented in Table 1. After UARPO fine-tuning, FinZero exhibits more competitive prediction performance compared to baseline models, whether in price prediction tasks or volatility prediction tasks. While FinZero with 3B parameter size surpasses larger parameter models such as GPT-4o. Additionally, when test set samples are divided into three equal groups based on the model's uncertainty scores sorted from highest to lowest as in Table 2, it shows that for the FinZero, the prediction accuracy of samples with high confidence scores is further improved—the prediction accuracy of the highest-score group is increased by approximately 13.5% relative to that of GPT-4o. Comparing with Qwen2.5-VL-3B fine-tuned by GRPO, FinZero achieves better average prediction performance. Meanwhile, grouping based on confidence scores exhibits a more pronounced positive correlation with prediction accuracy. Besides, we illustrate the accuracy changes of the two models during the fine-tuning process, as shown in Figure 4.

4 Conclusions

In this study, we introduced FinZero, pioneering the field of multimodal financial time-series reasoning. To achieve this, we developed the FVLDB dataset specifically designed for reasoning and analysis in financial time-series tasks, and designed the UARPO algorithm, which enables the implementation of relative advantage strategies through uncertainty adjustment. Experimental results show that even when applied to a small-scale model, the UARPO method significantly enhances the model's capabilities in financial time-series reasoning and prediction. Its performance in financial time-series prediction tasks is not only competitive with larger models like Qwen-7B but also competitive with large-scale models such as GPT-4o. Furthermore, after reinforcement learning fine-tuning, the model's uncertainty scoring output provides an "uncertainty dimension" for predictions: sorting samples by this score reveals that samples with higher scores exhibit higher prediction accuracy. This indicates that such uncertainty information can serve as an effective reference for assessing prediction reliability, thereby aiding in improving overall prediction accuracy. By developing a reinforcement learning algorithm tailored for high-uncertainty financial time-series prediction scenarios and a corresponding multimodal financial dataset, this study thoroughly validates the feasibility and application potential of cross-modal reinforcement learning fine-tuning in the field of financial reasoning and prediction.

References

[1] 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.

[2] Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. arXiv preprint arXiv:2212.08073, 2022.

[3] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. TEMPO: Prompt-based generative pre-trained transformer for time series forecasting. In The Twelfth International Conference on Learning Representations, 2024.

[4] Liang Chen, Lei Li, Haozhe Zhao, and Yifan Song. Vinci. r1-v: Reinforcing super generalization ability in vision-language models with less than 3.

[5] Luke Nicholas Darlow, Qiwen Deng, Ahmed Hassan, Martin Asenov, Rajkarn Singh, Artjom Joosen, Adam Barker, and Amos Storkey. DAM: Towards a foundation model for forecasting. In The Twelfth International Conference on Learning Representations, 2024.

[6] DeepSeek-AI, D. Guo, D. Yang, H. Zhang, and et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.

[7] Lingxiao Du, Xiangyan Liu, and Fanqing Meng. R1-Multimodal-Journey: A Journey to Real Multimodal Reinforcement Learning. GitHub Repository, 2024.

[8] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948, 2025.

[9] Jian Hu, Jason Klein Liu, Haotian Xu, and Wei Shen. Reinforce++: An efficient rlhf algorithm with robustness to both prompt and reward models, 2025.

[10] Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with irregular and scale-invariant patterns. In The Twelfth International Conference on Learning Representations, 2024.

[11] Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. arXiv preprint arXiv:2412.16720, 2024.

[12] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In The Twelfth International Conference on Learning Representations, 2024.

[13] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In International Conference on Learning Representations, 2022.

[14] Seunghan Lee, Taeyoung Park, and Kibok Lee. Soft contrastive learning for time series. In The Twelfth International Conference on Learning Representations, 2024.

[15] Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, pages 18915–18923, 2025.

[16] Shudong Liu, Zhaocong Li, Xuebo Liu, Runzhe Zhan, Derek F. Wong, Lidia S. Chao, and Min Zhang. Can llms learn uncertainty on their own? expressing uncertainty effectively in a self-training manner. In EMNLP, pages 21635–21645, 2024.

[17] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625, 2023.

[18] Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In The twelfth international conference on learning representations, pages 1–43, 2024.

[19] Fanqing Meng, Lingxiao Du, Zongkai Liu, Zhixiang Zhou, Quanfeng Lu, Daocheng Fu, Tiancheng Han, Botian Shi, Wenhai Wang, Junjun He, et al. Mm-eureka: Exploring the frontiers of multimodal reasoning with rule-based reinforcement learning. arXiv preprint arXiv:2503.07365, 2025.

[20] Yingzhe Peng, Gongrui Zhang, Miaosen Zhang, Zhiyuan You, Jie Liu, Qipeng Zhu, Kai Yang, Xingzhong Xu, Xin Geng, and Xu Yang. Lmm-r1: Empowering 3b lmms with strong reasoning abilities through two-stage rule-based rl. arXiv preprint arXiv:2503.07536, 2025.

[21] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

[22] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024.

[23] Mingtian Tan, Mike Merrill, Vinayak Gupta, Tim Althoff, and Tom Hartvigsen. Are language models actually useful for time series forecasting? Advances in Neural Information Processing Systems, 37:60162–60191, 2024.

[24] Haoran Wei, Yaofeng Sun, and Yukun Li. Deepseek-ocr: Contexts optical compression, 2025.

[25] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186, 2022.

[26] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems, 34:22419–22430, 2021.

[27] Zhiqiu Xia, Jinxuan Xu, Yuqian Zhang, and Hang Liu. A survey of uncertainty estimation methods on large language models. 2025.

[28] Tian Xie, Zitian Gao, Qingnan Ren, Haoming Luo, Yuqian Hong, Bryan Dai, Joey Zhou, Kai Qiu, Zhirong Wu, and Chong Luo. Logic-rl: Unleashing llm reasoning with rule-based reinforcement learning. arXiv preprint arXiv:2502.14768, 2025.

[29] Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. In Thirty-seventh Conference on Neural Information Processing Systems, 2023.

[30] Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. Advances in neural information processing systems, 36:43322–43355, 2023.

A Related Work

A.1 Large model fine-tuning with RL

The application of large-scale reinforcement learning (RL) fine-tuning to enhance the reasoning capabilities of large language models (LLMs) has garnered increasing research attention [8, 1]. OpenAI pioneered the use of Reinforcement Learning from Human Feedback (RLHF) for LLM fine-tuning, significantly improving instruction-following competence and generation quality. The Claude model series, through its Constitutional AI framework [2], integrates predefined rule systems with self-supervised mechanisms, partially replacing human preference annotations in conventional RLHF. This approach not only ensures content compliance but also elevates model performance in complex reasoning tasks. Recent technological advancements have further validated the critical role of large-scale RL fine-tuning in boosting reasoning abilities. For instance, DeepSeek-R1 [8] and o1 [11] demonstrated substantial improvements in LLM reasoning even without supervised fine-tuning.

Although DeepSeek-R1 has open-sourced model parameters, its training code remains proprietary. Subsequent studies have attempted to replicate these methodologies: Logic-RL [28] successfully reproduced rule-based RL fine-tuning on a 7B-parameter LLM, while multimodal extensions include R1-V [4], R1-Multimodal-Journey [7], LMM-R [20], and MM-EUREKA [19]. Current multimodal reasoning research primarily focuses on two domains: 1. general multimodal reasoning (e.g., cross-modal alignment and visual question answering) and 2. agent-related reasoning (e.g., embodied decision-making and tool manipulation). However, temporal reasoning in multimodal contexts (e.g., video event prediction and longitudinal data analysis) remains underexplored.

A.2 Time Series Forecasting

General time series forecasting has increasingly garnered attention. Current general time series datasets cover multiple critical domains, including electricity, weather, traffic, and finance. Model architectures encompass various types, such as attention-based models[26, 17], CNN-based models[18, 25], GNN-based models[29], and others. Additionally, the construction of time series foundation pre-trained models has gained traction. For instance, TimeLLM [12] employs a reprogramming framework and Prompt-as-Prefix to build large-scale time series forecasting models. CALF [15] introduces a multimodal pre-trained model with a dual-branch structure integrating time series target branches and textual source branches. OneFitsAll [30] freezes most parameters of large models while fine-tuning a small subset for time series tasks. Despite the growing research on time series large models, studies indicate that current approaches still face challenges in predictive performance [23]. Furthermore, the critical reasoning capabilities and interpretability of large models remain underdeveloped for widespread application in the time series domain. To address these gaps, our work explores the performance of multimodal large models with image-text inputs on time series reasoning tasks. We design UARPO fine-tuning to enhance predictive accuracy and improve the model's grasp of uncertainty risks in reasoning. Additionally, enhancing the reasoning capabilities of large models through reinforcement learning holds significant potential for advancing time series applications.

B Discussion of Image Rasterization trade-off

Rasterizing time-series data into images inherently involves a trade-off. On one hand, it facilitates the recognition of behaviorally significant patterns and contexts in finance—such as psychological barriers at integer price levels (e.g., 3,000 points) or resistance near prior highs. On the other hand, this process may sacrifice precise numerical accuracy. Preserving key statistical values—such as the latest price or interval highs/lows—in textual or numerical form alongside the image can help mitigate this loss of precision.

Moreover, rasterization serves as an effective method for information compression and pattern recognition. Similar to how models like DeepSeek-OCR[24] use visual tokens to achieve high compression ratios for lengthy documents, converting time-series data into images may circumvent limitations in scale and computational efficiency associated with time-series-specific tokenization, as well as the challenges in modeling long-range dependencies. Nevertheless, given the unique characteristics of temporal data, maximizing numerical precision remains a valuable direction for future research.

C FVLDB Dataset

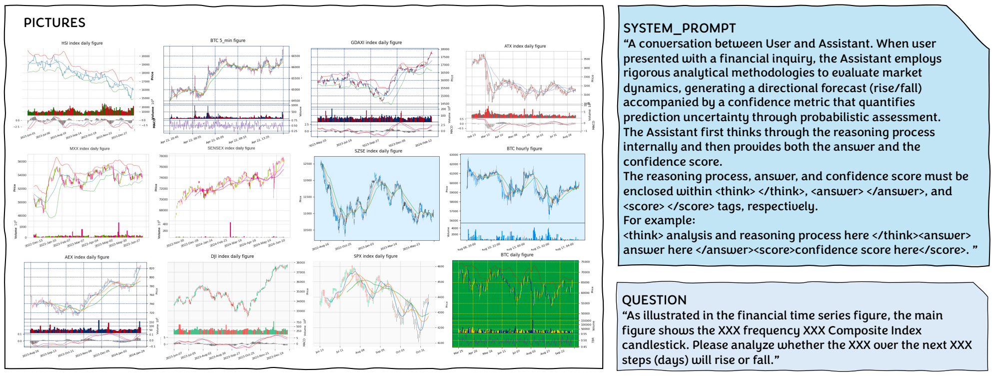

Figure 2: Overview of Image-Text Pairs for the FVLDB Dataset.

To validate our idea, we have specifically constructed a financial time-series image-text dataset (FVLDB as Figure 2) with over 10000+ samples. The images in FVLDB contain a wealth of financial assets, along with corresponding text descriptions and questions. To enhance data diversity, FVLDB includes index data from global stock markets, as well as data on cryptocurrency assets such as Bitcoin. The time-series length, sampling frequency, type, and number of features of the assets in each image are variable, and the image styles are also diverse. This flexibility enables the model to process diverse data types.

D UARPO Algorithm

Algorithm 1 Iterative UARPO

Input: Initial policy model $\pi_{\theta_{\text{init}}}$; reward model r_ϕ; task prompts $\mathcal{D}$; hyperparameters ϵ, β, μ; stack length L
Initialize: policy model $\pi_\theta \leftarrow \pi_{\theta_{\text{init}}}$; target special stack $\mathbf{s}_\mathbf{L}^\tau$
for iteration $= 1$ **to** I **do**
  Update reference model $\pi_{\text{ref}} \leftarrow \pi_\theta$
  Initialize stack $\mathcal{S}[0..L-1]$
  for step $= 1$ **to** M **do**
    Sample batch $\mathcal{D}_b \subset \mathcal{D}$
    Update the old policy $\pi_{\theta_{\text{old}}} \leftarrow \pi_\theta$
    Sample G outputs $\{o_i\}_{i=1}^G \sim \pi_{\theta_{\text{old}}}(\cdot|q)$ for each question $q \in \mathcal{D}_b$.
    Compute rewards $\{r_i\}_{i=1}^G$ and confidence scores $\{u_i\}_{i=1}^G$ for each output O_i by runnning r_ϕ.
    Compute current step average reward $\frac{1}{G}\sum_{i=1}^G r_L^\tau$ for current target τ.
    Compute $\hat{A}_{i,t}^I$ for the t-th token of o_i through group relative advantage estimation.
    if $step > L$ **then**
      Compute $\hat{A}_{i,t}^I$ for the t-th token through latest L step relative advantage estimation for target τ
    Gather two part relative advantage and multiply with coressponding confidence score
    for UARPO iteration $= 1, ..., \mu$ **do**
      Update the policy model π_θ by maximizing the UARPO objective.
    Update r_ϕ through continuous traning using a replay mechanism.
Output: π_θ

E Training Process

Figure 3: Overview of the FinZero Training.

F Comparison Grouped by Scores

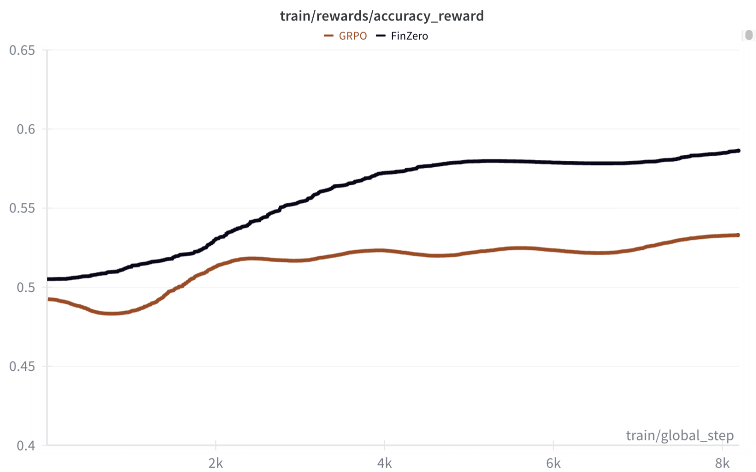

Figure 4: Accuracy Comparison of FinZero and GRPO Finetuning

Table 2: Model Prediction Accuracy Across Confidence Score Groups.

	Low (%)	Middle (%)	High (%)
Qwen2.5-VL-3B	51.2	51.7	49.3
Qwen2.5-VL-7B	47.38	47.81	54.36
GRPO	53.85	53.19	54.61
GPT-4o	49.85	49.42	54.75
FinZero	54.48	56.67	62.13

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification:

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: [NA]

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to

generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: [NA]

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification:

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [Yes]

Justification:

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.