FinRL Contest 2025 Task 1:Market-Aware In-Context Learning Framework for Proximal Policy Optimization in Stock Trading Using DeepSeek

1st Sahar Arshad
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

2nd Huma Ameer
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

3rd Nikhar Azhar
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

4th Seemab Latif
School of Electrical Engineering and Computer Science
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
seemab.latif@seecs.edu.pk

***Abstract*—This paper introduces a Market-Cap Stratified Subset (MCSS) strategy. It constructs a balanced, high-quality dataset from the Financial News and Stock Price Integration Dataset (FNSPID). Building on MCSS, we propose Market-Informed Sentiment for Trading (MIST). This is a two-phase framework that first generates sentiment scores through optimized prompt engineering with LLMs, then aligns market movements to deliver actionable trading signals. With a Proximal Policy Optimization (PPO)-based agent, MIST achieves better policy stability and learning efficiency than baseline methods.**

***Index Terms*—Reinforcement Learning, FinRL, Large Language Models, Automated Stock Trading, Deepseek**

I. Introduction

In the traditional stock market setup, the equity research team regularly produces research reports, trend forecasts, and personalized investment recommendations to offer valuable insight into potential market opportunities. This process requires considerable cognitive effort within a limited timeframe, leading to potential biases and suboptimal investment decisions. The Financial Crisis Inquiry Commission Report exposed how flawed risk management and opaque decision systems resulted in a loss of $22 trillion in the global economy and an almost 50% drop in the S&P 500 index[1]. With generative AI-based solutions, innovative equity research approaches can transform investment decisions. However, it often neglects the integration of alternative data, such as financial news, and the incorporation of market context into sentiment analysis. Recent work addresses automated stock trading by infusing Reinforcement Learning (RL) with Large Language Models (LLMs). For instance, previous work [1] demonstrates the utility of LLMs for generating sentiment and risk scores, which are then used as input to RL algorithms for decision making. In this work, we build on that foundation and propose a two-phase framework. Our key contributions are as follows:

- We introduce a novel Market-Cap Stratified Subset (MCSS) strategy. It enables efficient training on high-quality and balanced data samples in varying market capitalization bands.
- We propose Market-Informed Sentiment for Trading (MIST), a two-phase framework that combines LLM-generated sentiment scores with market-aware price signals for better interpretability.
- We also conduct an empirical evaluation to assess the impact of optimized prompts, designed for RL agent performance metrics.

II. Dataset

In this study, we use the Financial News and Stock Price Integration Dataset (FNSPID) [2], [3]. It consists of 15.7 million financial news records spanning from 1999 to 2023.

A. Filtration and Preparation

1) News Data: We begin with a subset of FNSPID dataset as used in the paper [1]. Considering the computational constraints, we apply our MCSS strategy and selected 15 companies with varying market capitalization bands. These companies are into three broad bands: Large-cap (market capitalization of approximately $10 billion or more), Mid-cap (ranging between roughly $2 billion to $10 billion), and Small-cap (up to about $2 billion, but not less than $300 million)[2]. The resulting subset consists of approximately 74K financial news.

Next, we cater for the missing articles by replacing them with corresponding headlines. Additionally, the articles exceeding 5,000 words are replaced with their respective Textrank summaries. These summaries are already available in the FNSPID dataset. The rationale behind this is to comply with the input token limitations of LLMs.

[1]The Financial Crisis Inquiry Commission Report (2012)

[2]https://finance.yahoo.com

979-8-3315-9661-3/25/$31.00 ©2025 IEEE
DOI 10.1109/IDS66066.2025.00024

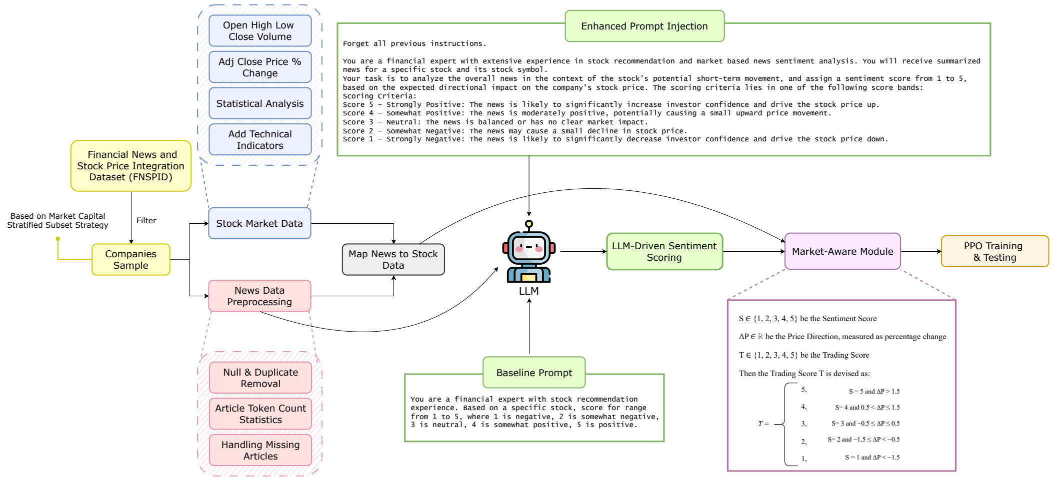

Fig. 1. Overview of the MIST Framework for Market-Informed Sentiment-Based Trading.

Company	Min %	Max %	Avg %	Med %	Std %
ADBE	-30.20	32.20	0.13	0.00	3.06
AMZN	**-94.71**	34.47	0.16	0.04	3.76
AMD	-37.93	52.29	0.11	0.00	3.74
AAPL	-75.58	**297.41**	0.12	0.00	4.13
LULU	-32.31	21.15	0.15	0.07	3.30
EBAY	-20.58	46.38	0.11	0.03	3.16
CSCO	-19.83	24.39	0.12	0.08	2.60
GOOGL	-11.63	50.22	0.11	0.07	2.08
GPRO	-26.10	31.34	-0.00	0.00	3.93
MNST	-50.00	266.63	1.51	0.00	**19.12**
PYPL	-24.59	36.71	0.07	0.10	2.55
NFLX	-40.91	42.22	0.20	0.03	3.64
NVDA	-35.23	42.41	0.19	0.12	4.20
UA	-51.20	84.26	-0.05	-0.09	4.24
TSLA	-92.43	24.40	0.19	0.13	3.94

TABLE I
DESCRIPTIVE STATISTICS OF DAILY PERCENTAGE CHANGES IN STOCK PRICES FOR SELECTED COMPANIES.

2) Stock Market Data: The stock data consists of open, close, high, and low prices from Yahoo Finance for the companies listed in Table I. We compute the percentage change for the adjusted close price of daily stocks to capture short-term market dynamics. The statistics summarized in Table I, are integrated into advanced phase of the proposed MIST.

III. METHODOLOGY

A. MIST: Market-Informed Sentiment for Trading

To enhance the quality of sentiment signals and integrate them more effectively in an automated trading policy, we introduce a novel two-phase framework called MIST. It is designed to align natural language sentiment with market behavior enabling informed trading decisions.

In the first phase, we design an enhanced prompt that incorporates few-shot examples, an explicit output format, and logically structured instructions as illustrated in Figure1. Furthermore, the prompt defines clear, tiered scoring criteria with their corresponding interpretations. It explicitly establish the llm's role, domain expertise, and the specific objective of sentiment tagging task. The prompt also requires a natural language reason in the response. It is essential for financial applications that demand auditability and interpretability, particularly in automated trading pipelines.

In the second phase of MIST, we devise an advanced Market-Aware module as shown in Figure 1. This module explicitly considers how short-term stock price direction can either reinforce or contradict the sentiment extracted from financial news. The sentiment scores computed in phase 1 are integrated with market trends to get the final trading score.

Investors often rely on patterns in daily returns, so in this module, the thresholds to determine stock price directions are based on the analysis of daily adjusted close percentages. These are empirically derived from a descriptive statistical analysis of daily percentage changes as shown in Table I. Finally, we synthesize the trading score based on the Mapping Rules. The rationale behind this fusion is complex nature of financial markets which we encapsulate in our framework.

B. Reinforcement Learning with Infused Signals

With the final trading score computed, we retrain the PPO agent using the same architecture as in the baseline. However, the input state now includes both traditional technical indicators and the infused trading score. The integration of these infused signals enhances the agent's ability to generalize across regimes and to adapt its policy in both risk-neutral and risk-sensitive market scenarios.

IV. EXPERIMENTS AND RESULTS

To evaluate the impact of different prompt designs on trading performance, we conduct a series of experiments on 3K data subset. In each experiment, a reinforcement learning agent is trained using PPO, where the scores generated by different prompts serve as input signals.

A. Experimental Design

We design three experiments to progressively evaluate the effects of prompt engineering on the agent's performance. **Experiment 1**: Sentiment scores are generated using a base prompt with minimal structure. The PPO agent is trained

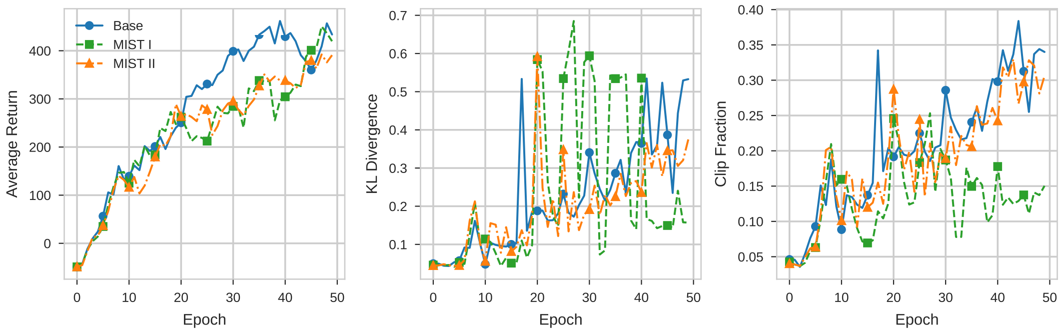

Fig. 2. PPO Training Results per Epoch

using price data and the corresponding LLM-generated scores. **Experiment 2**: Sentiment scores are derived from an enhanced prompt. The PPO agent is trained using both price data and sentiment scores. **Experiment 3**: A market-aware advance module is used that incorporates sentiments and daily trends. The PPO agent is trained with the same input configuration as in the previous experiments. In all three settings, the PPO agent is trained in the same environment and data subset to ensure comparability. We compared PPO-specific metrics throughout the training and testing, to assess learning progress and policy stability.

B. Experimental Setup

To compute the scores, we utilize the DeepSeek-R1 7B model [4] locally via the Ollama framework[3], which enables low-latency inference and fine-grained control over prompt design. PPO training is conducted on Google Colab using a CPU-enabled environment. All the other training parameters are kept consistent with the baseline setup described in [1].

C. Results and Discussion

The training results in Figure 2 depicts that the base model follows a conservative learning strategy. It achieves the highest raw episode return but with less stable updates. In contrast, the two-stage MIST framework offers a more stable and efficient policy learning. It provides the best trade-off between performance and robustness for real-world deployment.

The test results as shown in Table II, shows that the base model slightly outperforms in Information and Rachev ratios. However, it exhibits higher downside risk with its CVaR value. MIST benefits from structured prompt engineering, which enhances the quality of sentiment signals and shows a more balanced risk-reward profile. All experimented policies exceed the NASDAQ benchmark, with MIST Phase II demonstrating the strongest risk-adjusted returns by combining sentiment with market-aware features.

[3]https://ollama.com

Metric	Base	MIST I	MIST II
IR ↑	**0.293**	0.173	0.194
CVaR ≈ 0	-0.788	**-0.594**	-0.595
Rachev Ratio ↑	**6.68**	3.27	4.06
Sharpe Ratio ↑	**4.66**	2.77	3.10

TABLE II
TEST METRICS FOR BASE PPO AND MIST VARIANTS

V. CONCLUSION AND FUTURE DIRECTION

Based on the results obtained from current experiments, we conclude that a sentiment-only model may encounter limitations when external factors such as macro economic shocks or technical corrections are excluded. Conversely, price movements alone may lack interpretability or causal factors. Our proposed framework MIST, provides a holistic causal-effect relationship while offering deep understanding of market dynamics. The framework is designed for next-day prediction using end-of-day news and market data. It processes an average of 3–4 filtered news articles per company during post-market hours and retrieves market data via APIs, enabling low-latency deployment aligned with standard trading practices. Future work will focus on enhancing PPO through hyperparameter tuning and integrating explicit risk modeling. To address PPO's limitations, we will explore advanced RL methods such as CVaR-PPO and GRPO. Recent work [5] shows that structuring news into event timelines supports robust trading strategies and enables clustering of key market factors, aiding more informed decision-making. We aim to incorporate structured reasoning from sentiment-tagged news and event-based representations to enhance market context. Integrating such mechanisms can strengthen decision-making by aligning actions with evolving market narratives and latent causal drivers.

REFERENCES

[1] M. Benhenda, "Finrl-deepseek: Llm-infused risk-sensitive reinforcement learning for trading agents," *arXiv preprint arXiv:2502.07393*, 2025.
[2] Z. Dong, X. Fan, and Z. Peng, "Fnspid: A comprehensive financial news dataset in time series," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 4918–4927.
[3] K. Wang, K. Xiao, and X.-Y. L. Yanglet, "Parallel market environments for finrl contests," *arXiv preprint arXiv:2504.02281*, 2025.
[4] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi *et al.*, "Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning," *arXiv preprint arXiv:2501.12948*, 2025.
[5] S. Arshad, N. Azhar, S. Sajid, S. Latif, and R. Latif, "Cross-lingual news event correlation for stock market trend prediction," 2024. [Online]. Available: https://arxiv.org/abs/2410.00024