Structuring News, Shaping Alpha: RL-Enhanced LLMs in a Hybrid Framework for Event Driven Financial Forcasting

Haohan Zhang[1,†,*], **Saizhuo Wang**[2,†], **Hao Kong**[2], **Baozhu Shang**[3]
[1]HKUST(GZ), Guangzhou, China [2]HKUST, Hong Kong, China
[3]IDEA Research, Shenzhen, China
hzhang760@connect.hkust-gz.edu.cn, swangeh@connect.ust.hk
hkongab@connect.ust.hk, shangbaozhu@idea.edu.cn

Abstract

There has been an emergent field within AI-powered financial forecasting that leverages alternative data, particularly unstructured news and event information. Existing approaches often rely on fixed sentiment lexicons or manually defined event taxonomies, while recent advances in large language models (LLMs) have inspired the use of prompt engineering to structure such events into features for predictive modeling. However, such methods, though offering flexibility across modalities, fail to adapt to the constantly shifting dynamics of financial markets. Directly using human-annotated labels to guide adaptation is impractical, as annotation in financial domains are often not explicitly defined. How, then, can we align LLM event structuring with predictive objectives in a scalable and efficient way? In this work, we propose Structuring News, Shaping Alpha, a hybrid framework that integrates reinforcement learning–enhanced LLMs with ensemble-based forecasting models. Our system employs an LLM to re-classify financial events into structured categories, which are passed as features into a downstream ensemble predictor. Crucially, the LLM's classification policy is optimized in a closed-loop setting via Proximal Policy Optimization (PPO), where the reward derives not from human supervision but from the predictive value of the resulting features, measured through information coefficient (IC) against market returns. We argue that in domain tasks such as financial forecasting, the LLM's strength lies in feature extraction, while the machine learning model excels at mapping structured features to numerical outputs. By combining these strengths, we advance a hybrid modeling paradigm in which LLMs and machine learning models each perform what they do best, yielding more adaptive and powerful event-driven prediction. Experiments on large-scale Chinese A-share stock data demonstrate that our RL-enhanced classifications yield a non-tricial information coefficient while consistently outperform carefully engineered prompt-only methods using a flagship LLM, yielding more adaptive and powerful event-driven prediction.

1 Introduction

Ever since early work demonstrated the predictive value of financial news [Tetlock, 2007], a growing body of researchSoun et al. [2022], Xu and Cohen [2018] has explored the use of textual data to extract sentiment signals that are often absent from traditional price-and-volume-based factor models.

[†]These authors contributed equally. [*] Haohan Zhang is the corresponding author.

[‡] This work is supported by Shenzhen Science and Technology Program (No. KJZD20240903103210014)

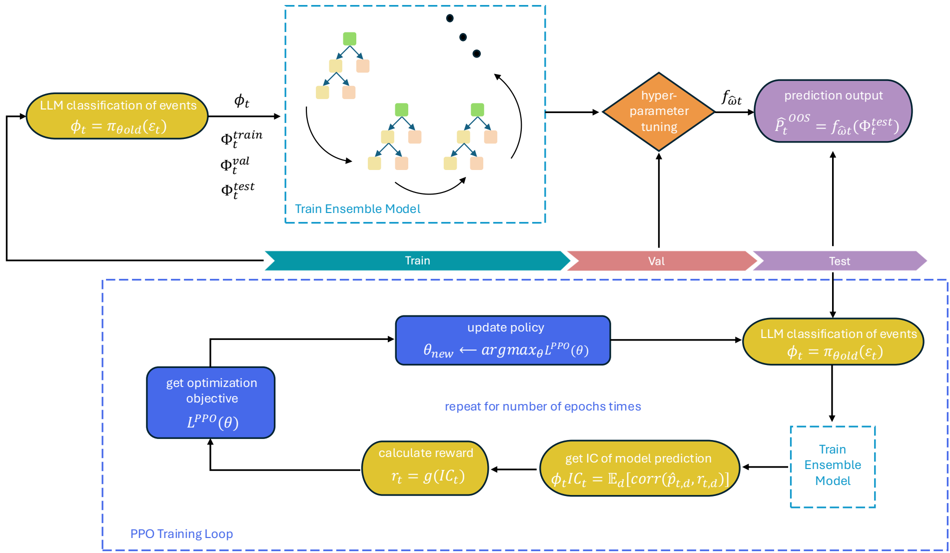

Figure 1: Hybrid model pipeline

This line of research has shown that qualitative narratives—whether in news articles, analyst reports, or social media—can capture dimensions of market behavior that numerical indicators alone fail to reveal.

The recent advent of large language models (LLMs) has further accelerated this trend. With their ability to parse unstructured text and generate human-level interpretations, LLMs appear to offer a powerful new tool for extracting insights from financial documentsLopez-Lira and Tang [2023], Xiao et al. [2024]. Through in-context learning and prompt-based querying, these models can evaluate the implications of market-relevant information in a flexible, zero-shot setting. However, the nature of the specific task at hand, not the capability of the LLMs themselves, leaves something to be desired for such an approach. Unlike general NLP tasks, financial forecasting operates in an environment with a low signal-to-noise ratio, where subtle variations in model output can have outsized implications for downstream trading decisions. Prompt-based methods do not adapt as market conditions evolve, nor do they optimize directly for predictive accuracy. Reinforcement learning (RL), or more specifically Proximal Policy Optimization (PPO)Schulman et al. [2017] provides a natural solution. Althoug RL has already proven successful in aligning LLM behavior with human preferencesOuyang et al. [2022], yet in finance, human-annotated labels are not a practical supervision source: annotations are costly, ambiguous, and inevitably lag behind market reality. What is needed instead is a dynamic reward signal—a metric that reflects how well the LLM's structured outputs support financial prediction, and one that co-evolves with the ever-changing conditions of the market itself.

In this work, we propose a hybrid forecasting pipeline that explicitly separates the tasks of semantic feature construction and numerical prediction. First, an LLM classifies raw financial news into structured event categories, distilling them into binary feature vectors at the company-day level. These features are then fed into an XGBoost ensemble predictor, which estimates the probability of a company under-performing among all listed companies. Crucially, the LLM's event-classification policy is not fixed: after each rolling evaluation, its mappings are updated via Proximal Policy Optimization (PPO), where the reward is derived from the predictive alignment of its features with realized returns (measured by information coefficient). This closed-loop design allows the system to continually adapt its feature space to shifting market regimes while leaving the supervised predictor stable and efficient.

2 Methodologies

2.1 The Hybrid Model

Model Overview and Data Composition. Our stock universe consists of all listed Chinese A-share companies from the Shanghai and Shenzhen Stock Exchanges. Figure 1 illustrates the hybrid framework, which integrates (i) an LLM-based event classifier, (ii) a supervised ensemble predictor, and (iii) a reinforcement learning loop in a rolling pipeline. For each roll, firm-day observations are split chronologically into training, validation, and test periods. The LLM we used for PPO post-training enhancement was Qwen-2.5-3B-Instruct Team [2024].

Each company-day is first encoded as a binary vector of predefined raw event types (e.g., Personnel Change, Litigation). An entry equals 1 if a news item of that raw type occurs between the previous close and the current day's open (intraday news is excluded). For example, if Personnel Change is reported at 09:23 on day T_0 and Litigation at 17:23 on T_{-1}, both raw-event entries are set to 1 for day T_0. These raw labels are produced by a RoBERTa-based classifier Liu et al. [2019] fine-tuned on manually annotated financial news (a full list is given in Appendix 2). *Importantly, these raw event vectors (around 2.4M in total) are the fixed input space; the LLM's role is to subsequently group or reclassify them into higher-level categories during policy adaptation.* The first training window spans January 2020–August 2021 (20 months), followed by a 3-month validation period (September–November 2021) and a 3-month test period (December 2021–February 2022). Subsequent rolls advance each window by three months while keeping the left training boundary fixed.

Event Classification and Supervised Prediction. At roll t, the LLM maps raw events $\mathcal{E}_t$ into 10 new abstract event classes according to its inferred impact on prices, producing transformed features $\Phi_t = \pi_{\theta_{\text{old}}}(\mathcal{E}_t)$. These re-mapped features are then split into $\Phi_t^{\text{train}}, \Phi_t^{\text{val}}, \Phi_t^{\text{test}}$. Given Φ_t^{train}, we train an XGBoost ensemble to predict whether a firm falls into the bottom $p = 40\%$ of one-day-ahead returns. Hyperparameters are optimized on Φ_t^{val} using Hyperopt/TPE, yielding tuned parameters $\widehat{\omega}_t$. The final XGBoost model trained with $\widehat{\omega}_t$ generates probability predictions, which constitute the hybrid model's output on Φ_t^{test}. "Hybrid Model" means we deliberately separate the roles of the two model components with LLM as feature constructor; it learns semantically meaningful groupings of raw event types, leveraging its interpretive power to transform input features. On the other hand, we elect XGBoost ensemble model as feature-to-numerical mapper; it specializes in converting structured features into calibrated probability estimates of financial outcomes. Crucially, these test predictions are produced *before* reinforcement learning begins, ensuring that downstream RL adaptation does not contaminate the out-of-sample evaluation.

Reward Definition and RL Adaptation. To initiate PPO, we assess the quality of LLM's event classification by computing the average daily cross-sectional information coefficient (IC) between hybrid model predictions and realized one-day-ahead open returns in the test set. The reward is defined as $r_t = g(\text{IC}_t) = C\,\text{IC}_t$ with $C = -10$, linearly scaling IC within $[-0.1, 0.1]$ and clipped to $+1$ when $\text{IC}_t < -0.1$ and to -1 when $\text{IC}_t > 0.1$. This reflects empirical evidence that pre-open event signals often exhibit short-horizon reversal. Thus, the hybrid design assigns the LLM to adaptively refine event classification (via PPO), while XGBoost remains a fixed, efficiently optimized predictor. This separation ensures interpretability, stable supervised learning, and targeted adaptation where it matters most.

2.2 PPO in a Contextual Bandit Setting

We cast PPO into a contextual bandit form. At each roll, the policy $\pi_\theta(a|x_t)$ produces an event grouping a_t given context x_t, then receives reward $r_t = g(\text{IC}_t)$. Since there are no trajectories, the advantage reduces to $\hat{A}_t = r_t$.

The PPO clipped objective is

$$L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}_t\Big[\min\big(r_t(\theta)r_t,\, \text{clip}(r_t(\theta),\, 1-\epsilon,\, 1+\epsilon)r_t\big)\Big],$$

with importance ratio $r_t(\theta) = \pi_\theta(a_t|x_t)/\pi_{\theta_{\text{old}}}(a_t|x_t)$.

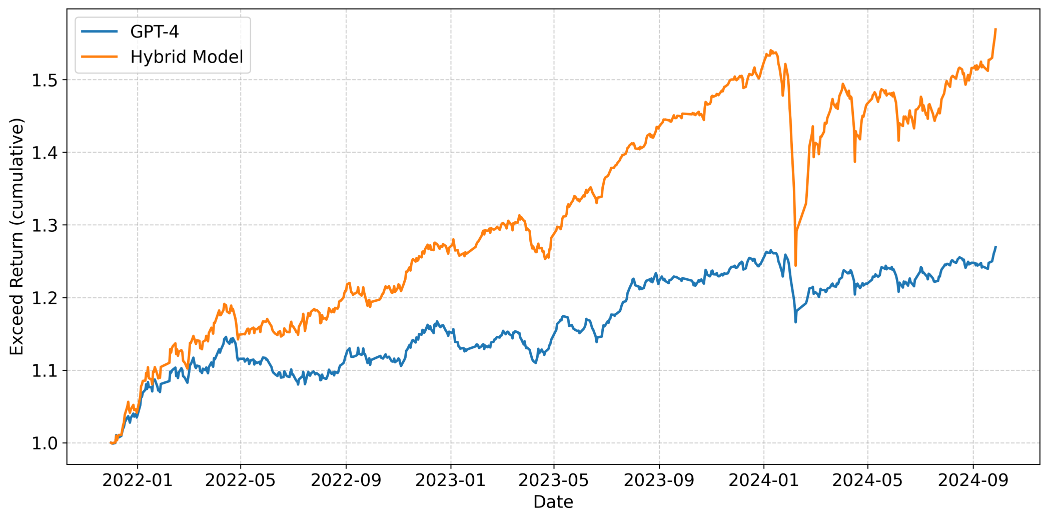

Figure 2: Cumulative exceed return of the proposed hybrid model versus GPT-4 baseline.

To prevent over-shifting, we add an adaptive KL penalty:

$$L^{\text{PPO}}(\theta) = L^{\text{CLIP}}(\theta) - \beta\,\hat{\mathbb{E}}_t\Big[\text{KL}\big(\pi_{\theta_{\text{old}}} \,\|\, \pi_\theta\big)\Big],$$

where β is dynamically adjusted.

This formulation enables stable policy updates in single-step bandit settings, adapting the LLM's event classification to maximize predictive alignment with returns.

3 Experimental Results

We evaluate our framework in a cross-sectional stock selection task on the Chinese A-share market. For each trading day in the test set, the hybrid model outputs the probability that a given stock will fall into the bottom 40% of one-day-ahead returns, based on features constructed from LLM-driven event classification. This probability is interpreted as a *negative signal*: stocks deemed less likely to be in the bottom 40% are ranked higher. Each day, we form a long portfolio by buying the set of stocks with the lowest predicted bottom-40% probability, subject to a maximum daily turnover of 5% and a transaction fee of 0.3%.

As a comparison, we evaluate GPT-4o-miniOpenAI [2024] as a direct predictor. Instead of relying on an intermediate feature-construction stage, GPT-4o-mini is provided with the raw daily event-occurrence vector and instructed to predict whether each stock will belong to the bottom 40% of returns the next day. Figure 2 reports the cumulative exceed return (relative to the CSI-1000 market benchmark) of the two strategies and Table 1 reports the metrics of backtest evaluation. Most importantly, the hybrid model yields a non-trivial ic of -1.61% from enhanced event classifications alone with no numerical feature added while the ic contribution from the GPT-4o-mini is almost neglegible. Additionally, the hybrid model consistently outperforms the GPT-4 baseline, across all metrics.

Table 1: Backtest metrics. Metrics marked with * are measured relative to the benchmark return.

Model	Annual Excess Return*	Sharpe Ratio*	Win Rate*	Average IC
GPT-4o-mini	9.48%	1.31	53.64%	-0.25%
Hybrid Model	**20.05%**	**1.60**	**59.91%**	**-1.61%**

4 Conclusion

In this work, we put forth a hybrid model paradigm that combines the interpretive strength of LLMs for semantic event structuring with the predictive efficiency of ensemble methods for numerical forecasting. Unlike end-to-end prompting baselines, our framework deliberately separates the roles of feature construction and outcome prediction, ensuring both interpretability and robustness. A key novelty lies in our design of an *IC-based reward* that directly links policy updates to predictive alignment with market returns, adapting PPO to a contextual bandit setting. Empirical results on large-scale Chinese A-share data demonstrate that this design yields non-trivial predictive information from event classification features alone, outperforming GPT-4o-mini in both statistical metrics and trading performance under realistic turnover and transaction cost constraints. These findings highlight the value of combining structured LLM-driven representations with reinforcement learning for dynamic adaptation to shifting financial environments. Future work may extend this paradigm to multi-horizon objectives, richer event hierarchies, and online market deployment.

References

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Alejandro Lopez-Lira and Yuehua Tang. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*, 2023.

OpenAI. Gpt-4o mini: Advancing cost-efficient intelligence. `https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/`, July 2024.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*, 2022.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Yejun Soun, Jaemin Yoo, Minyong Cho, Jihyeong Jeon, and U Kang. Accurate stock movement prediction with self-supervised learning from sparse noisy tweets. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1691–1700. IEEE, 2022.

Qwen Team. Qwen2.5: A party of foundation models, September 2024. URL `https://qwenlm.github.io/blog/qwen2.5/`.

Paul C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 62(3):1139–1168, 2007.

Yifan Xiao, Enze Sun, Dong Luo, and Wenhao Wang. Tradingagents: Multi-agents llm financial trading framework. *arXiv preprint arXiv:2412.20138*, 2024.

Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1970–1979, 2018.

5 Appendix

Table 2: List of raw financial event types used in this study.

Event Type	**Event Type**
Initial Public Offering (IPO)	Earnings / Performance
Individual Speech / Conduct	Personnel Change
Refinancing	Dividend / Bonus Issue
Cooperation / Partnership	Employee Stock Ownership
Insider Share Increase / Decrease	Regulatory Oversight
Legal Disputes	Production
Research and Development	Investigations and Penalties
Stock Price Increase	Stock Price Decrease
Share Buyback	Equity Freeze
Equity Incentive	Equity Pledge
Industry Policy	Industry Climate / Prosperity
Rating Upgrade	Rating Downgrade
Debt	Financial Quality
Loans	Asset Purchase / Sale
Asset Restructuring	Capital Financing
Liquidity / Capital	Sales
Risk Elimination	Risk Warning