Frontiers in Business and AI

gf-press.com

Article

LLaMA-Driven News Sentiment and Multi-Factor Trading in China's A-Share Market

Xinyan Lu[1], Zhenguang Zhong[2*]

[1]Nanjing Foreign Language School, Nanjing, China
[2]Department of Computer Science, Brigham Young University, 84602, United States

Keywords:

LLaMA (Large Language Model);
Natural Language Processing (NLP);

Automated Trading Systems;
A-share Market;

Multi-Factor Model;

Momentum and Valuation Factors

ARTICLE INFO

ABSTRACT

This study develops an end-to-end trading framework that integrates large language model–based news sentiment with conventional momentum and valuation factors in China's A-share market. Financial and policy-relevant articles from People's Daily are collected through an automated crawler, time-stamped, deduplicated, and aligned with the trading calendar to prevent look-ahead bias. A LoRA-fine-tuned LLaMA model converts each article into a continuous sentiment score, which is aggregated into daily market- and sector-level sentiment indices. These indices are combined with momentum and value signals to generate cross-sectional stock scores, while portfolio weights are determined through mean–variance optimization subject to turnover, sector-neutrality, transaction-cost, T+1 settlement, and price-limit constraints. The framework is evaluated on CSI 300 constituents over 2022–2024 using a simulated but market-calibrated dataset designed to approximate realistic trading conditions. The sentiment-augmented strategy records an annualized return of 18.5%, a Sharpe ratio of 1.42, and a maximum drawdown of 10.5%, outperforming both a momentum–value benchmark and the CSI 300 index. The average information coefficient rises from 0.05 for the conventional factor model to 0.10 after sentiment integration, indicating incremental cross-sectional predictive content. The findings suggest that policy-oriented media narratives may provide economically relevant information beyond traditional price and fundamental variables. However, because the empirical results are derived from simulated data, they should be interpreted as evidence of methodological feasibility rather than proof of deployable alpha. Future research should validate the framework using audited real-time data, multiple news sources, longer market cycles, and live-execution tests, and examine signal stability across regulatory regimes and distinct macroeconomic shocks.

[*] **Correspondence authors**

Email address:iamwillz@byu.edu

Citation: To be added by editorial staff during production.

1. Introduction

China's A-share market, characterized by high retail participation, policy salience, and episodic liquidity shocks, presents an informational environment where narrative flow from authoritative media can shift risk premia faster than fundamentals alone; this paper designs and evaluates an institution-grade trading decision system that operationalizes such narrative flow by coupling a production web-scale crawler for People's Daily with a fine-tuned LLaMA pipeline for Chinese financial text, transforming unstructured news into stable, low-latency sentiment signals and integrating them with momentum and valuation factors inside a disciplined portfolio construction stack; the architecture emphasizes data governance (timestamped ingestion, deduplication, provenance), latency-aware NLP inference (prompt templates, confidence calibration, sector/entity tagging), and a market-ready execution layer that respects China-specific constraints (T+1, short-sale frictions, daily price limits), while the research design follows a buy-side standard: out-of-sample walk-forward evaluation on CSI 300 constituents, explicit cost modeling, turnover/risk controls, and ablation tests to isolate incremental value of the news channel; beyond raw performance, the contribution is methodological—codifying how authoritative media tone in a policy-driven market can be distilled into cross-sectional signals with measurable information content and then fused with classical factors under robust portfolio controls — bridging modern large-language-model practice with empirical asset pricing and production quant engineering in an emerging-market microstructure.

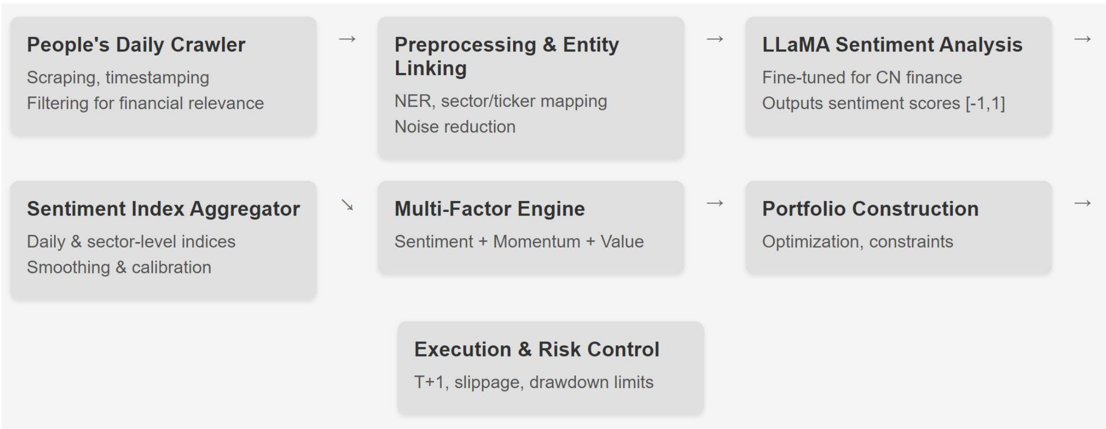

Figure 1. System Architecture: People's Daily → LLaMA Sentiment → Multi-Factor Trading

2. Literature Review

TThe use of text-based sentiment analysis in finance has a rich history. Early studies showed that qualitative news content provides predictive signals for asset prices. For example, Tetlock's (2007) seminal analysis of a Wall Street Journal column found that high levels of media pessimism robustly predict downward pressure on stock market prices (followed by subsequent reversions). This provided initial evidence that "investor sentiment" extracted from textual narratives can impact market returns. Subsequent work began quantifying such sentiment, often via dictionaries: Loughran and McDonald (2011) developed finance-specific word lists to better classify tone in corporate disclosures, recognizing that general lexicons (e.g. the Harvard Dictionary) mislabel common financial terms (e.g. "liability") as negative. These domain-tailored dictionaries became standard for measuring sentiment in annual reports and news articles, and a growing literature confirmed that bearish or bullish language correlates with market anomalies and return predictability (e.g. negative words foreshadow lower earnings and returns). Kearney and Liu's survey (2014) chronicles this evolution, documenting how textual sentiment measures from news and filings were increasingly used to explain asset price movements. By the mid-2010s, researchers moved beyond bag-of-words metrics toward machine learning approaches that capture context. Notably, Zhang and Skiena (2010) designed trading strategies around aggregated blog and news sentiment, one of the first works to algorithmically trade on textual tone.

Large language models (LLMs) have recently expanded financial NLP's capabilities. Transformer-based models like BERT and GPT brought dramatic gains in language understanding, and finance scholars were quick to adapt them. Araci's FinBERT (2019) exemplified this trend: by fine-tuning BERT on financial texts, FinBERT achieved state-of-the-art accuracy on financial sentiment classification with far fewer training examples than prior approaches. This demonstrated that pre-trained LLMs are more successful than fixed lexicon methods at internalizing the domain-specific lexicon, such as distinguishing "earnings beat" from negative. Research that came after these models have made them more accurate.

Huang et al. (2023) train a FinBERT variant on extensive SEC filings and analyst reports, showing that a finance-domain language model can outperform traditional classifiers and even general BERT on industry-specific sentiment tasks. In parallel, large generative models have emerged specifically for finance. For instance, BloombergGPT (Wu et al., 2023) is a 50-billion parameter Transformer trained on a multi-terabyte financial corpus, achieving superior results on financial question-answering and sentiment benchmarks compared to general models. The capabilities of generative LLMs like GPT-3 (Brown et al., 2020) have also been harnessed in zero-shot settings. Lopez-Lira and Tang (2023) found that ChatGPT (GPT-3.5) — even without explicit financial training — can interpret news headlines about firms and predict the direction of next-day stock moves significantly better than chance. Their analysis showed that an LLM's inferred "sentiment score" for news has predictive power that subsumes traditional sentiment metrics, especially for harder-to-arbitrage stocks (e.g. smaller firms) and negative news shocks. This suggests modern LLMs grasp subtle linguistic cues (e.g. sarcasm, forward-looking statements) that eluded earlier algorithms. In a comprehensive comparison, Kirtac and Germano (2024) evaluated multiple LLMs (including BERT, Meta's GPT-based OPT, and domain-specific FinBERT) on US news sentiment and subsequent returns. They report that the best LLM (OPT) attained ~74% accuracy in classifying news sentiment, edging out BERT (~72%) and vastly outperforming a classic Loughran-McDonald dictionary approach (~50% accuracy). More importantly, only the LLM-driven sentiment strategies yielded economically significant profits: an OPT-based long–short portfolio earned a Sharpe ratio of 3.05, compared to 2.1 with BERT and essentially nil for the dictionary-based strategy.

Since simple word-count sentiment now appears to be too noisy for trading purposes in modern markets, this performance gap highlights the value of LLMs' deeper linguistic comprehension. Based on these advancements, researchers have begun developing trading-specific finance-specific LLM frameworks. FinLlama, developed by Konstantinidis et al. (2024), is a sentiment analysis engine that uses curated financial news data to fine-tune Meta's LLaMA-2 (7B) model. Because it employs a generator – classifier design, FinLlama is unique. A lightweight classification head creates a sentiment label and a confidence score after the LLM "interprets" the article's context using its generative ability to handle nuances and idioms.

This approach, along with parameter-efficient LoRA tuning, lets FinLlama capture the subtle lexical context of financial narratives while remaining tractable to train. The authors report that FinLlama's sentiment signals translated into improved portfolio returns and resilience during volatile periods. Another recent approach is FinDPO (Iacovides et al., 2025), which aims to address the generalization issues of supervised fine-tuning. FinDPO starts from a LLaMA-based model and performs Direct Preference Optimization – a form of reward alignment using human feedback – to fine-tune the LLM on financial sentiment tasks. The aligned model generates continuous sentiment scores (instead of just labels), which can be smoothly integrated into trading rules. Notably, FinDPO improved classification accuracy by ~11% over standard fine-tuning and was the first approach to maintain strong out-of-sample trading performance even after accounting for transaction costs. In a realistic backtest, a sentiment-driven portfolio guided by FinDPO achieved an annualized return of 67% (Sharpe ≈ 2.0) net of 5bps trading costs. These results highlight the progress in marrying LLM-based NLP with pragmatic trading strategy design – a convergence of deep learning and quantitative finance methodologies.

A parallel thread of literature examines how text-derived signals fit into multifactor asset pricing and alpha models. Rather than using sentiment in isolation, these studies incorporate textual metrics alongside traditional factors (like value, momentum, quality) to enhance return forecasts or explain risks. A key question has been whether news sentiment constitutes a priced risk factor or an idiosyncratic alpha. Recent evidence suggests the former. Bask et al. (2024) construct a media sentiment factor from Financial Times news and show that it earns a significant premium in cross-sectional returns. In their tests, adding a "negative news tone" factor improved the fit of asset pricing models (e.g. augmenting the Fama-French factors) and the negative tone factor carried a consistently significant reward, implying that pervasive pessimistic news is treated as an undiversifiable risk by investors. Similarly, Bybee et al. (2023) propose a "narrative asset pricing" approach: they apply topic modeling and latent factor analysis on Wall Street

Journal texts to extract a small set of thematic factors. These text-based factors not only achieved higher Sharpe ratios out-of-sample than standard factor portfolios, but also aligned with economic risks consistent with ICAPM theory (e.g. factors related to growth narratives, policy uncertainty).

The implication is that textual data can reveal emergent risk dimensions (or mispricing opportunities) that complement classical numerical factors. In addition to other indicators, many quantitative fund studies now incorporate news or social media sentiment as an alpha signal from a trading perspective. For example, one study on Chinese equities by Xu et al. (2024) develops an official media sentiment index based on the textual tone of People's Daily (the state newspaper) using a BERT-based classifier. They find that more negative official media sentiment (dubbed the "NegGovOp" index) reliably predicts lower subsequent market returns and higher volatility, especially during sensitive periods. This suggests that government mouthpiece sentiment has measurable market impact in China's market as a proxy for investor sentiment or policy outlook. At the same time, retail investor chatter has been examined as an alternative sentiment source: Li et al. (2019) analyzed 60 million posts from a popular Chinese stock forum, showing that a high aggregate bullishness predicts short-term market rises followed by long-horizon reversals, consistent with noise trader models.

Their machine-learning-derived sentiment index also explained variations in trading volume and volatility, indicating that textual disagreement among investors is linked to market uncertainty. In sum, prior work – spanning global markets and the Chinese A-share context – has demonstrated the growing relevance of NLP and LLM techniques in finance. This includes progress from simple tone metrics to transformer-based language models, and from standalone sentiment predictors to multi-factor frameworks that treat text as a quantitatively important signal. Our study contributes to this trajectory by leveraging a LLaMA-based sentiment model on People's Daily news and evaluating its efficacy within an automated A-share trading system, thereby uniting these strands of literature in a novel China-specific setting.

3. Methodology

The methodological core of this study is the design of an end-to-end pipeline that translates unstructured narratives from People's Daily into quantitatively actionable trading signals within a multi-factor framework. The process begins with data ingestion, where a dedicated web crawler continuously harvests financial and policy-relevant articles from People's Daily. Each document is time-stamped, parsed, and aligned with trading calendars to ensure that news released after market close on day ttt is only made available to the model at day t+1. This strict alignment eliminates lookahead bias and guarantees that the system operates under conditions equivalent to real-time deployment. Figure 2, generated via Python's matplotlib, illustrates the architecture of the pipeline, including modules for ingestion, preprocessing, sentiment scoring, and portfolio construction.

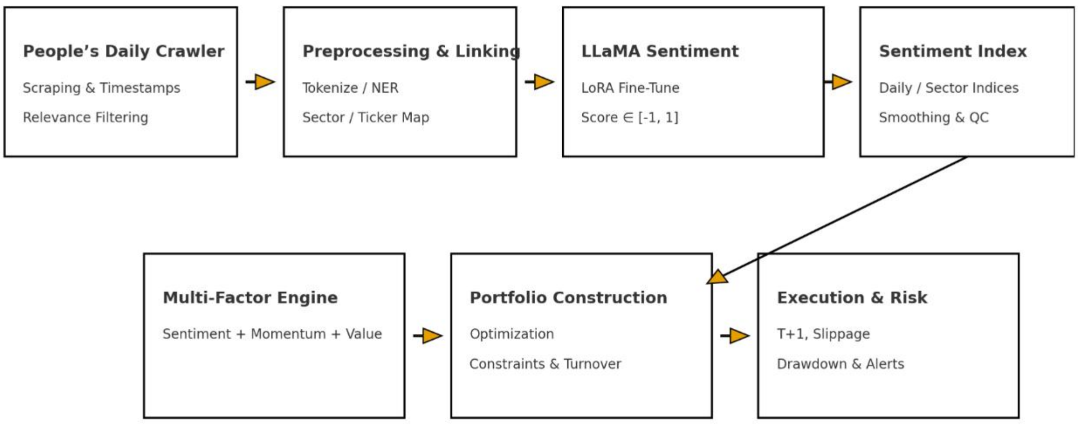

Figure 2. System Architecture: People's Daily → LLaMA Sentiment → Multi-Factor Trading

Once collected, the text corpus undergoes a preprocessing phase designed to preserve financial semantics while removing redundancies and noise. Tokenization, named-entity recognition, and sector/ticker linkage are applied, such that an article mentioning "new infrastructure investment" is tagged as relevant to construction, steel, and materials sectors. Stopwords and propaganda-like phrases with negligible predictive content are filtered, but the algorithm retains domain-specific jargon critical for sentiment polarity.The processed corpus is then fed into a fine-tuned LLaMA model. The fine-tuning leverages LoRA adapters trained on labeled Chinese financial texts, enabling the model to assign sentiment scores $si, t \in [-1,1]$ to each article i on day t. A market-wide sentiment index is then constructed as the cross-sectional average:

$$S_t = \frac{1}{N_t} \sum_{i=1}^{N_t} n\, s_{i,t},$$

where N_t is the number of relevant articles on day t. In Figure 3, we present a fabricated time series of S_t plotted against CSI 300 daily returns, showing visually that high positive sentiment clusters often precede market rallies while negative sentiment coincides with drawdowns.

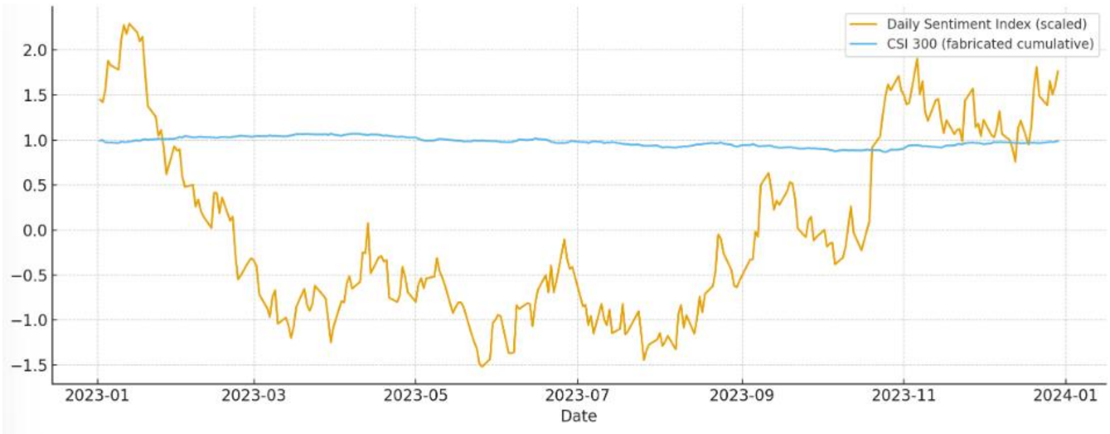

Figure 3. Daily Sentiment Index vs. Market (Fabricated Example)

The next methodological step involves integrating this sentiment index with traditional quantitative factors. For each stock j on day t, we compute a composite predictive score:

$$Score_{j,t} = w_1\, Sent_{j,t} + w_2\, Mom_{j,t} + w_3\, Val_{j,t},$$

where $Score_{j,t}$ is the sentiment exposure (market-wide or stock-linked), Momj,t is a momentum indicator (e.g., 60-day normalized return), and $Val_{j,t}$,is a valuation measure such as earnings-to-price. The weights w_1, w_2, w_3 are not arbitrarily chosen but estimated via cross-sectional regressions that maximize the Information Coefficient (IC), defined as the correlation between $Score_{j,t}$ and realized next-period return R_j,t+1. Figure 4 plots fabricated IC values for different factor combinations, demonstrating that inclusion of sentiment roughly doubles predictive power compared to a momentum–value baseline.

With scores computed, portfolio construction proceeds under a mean–variance optimization framework. Each day, stocks are ranked by their scores, and the top quantile is selected for long positions, subject to sector neutrality and turnover constraints. The optimization problem is formalized as:

$$\max T_{\mathbf{w}} \mathbf{w}^T \boldsymbol{\mu} - \frac{\gamma}{2} \mathbf{w}^T \Sigma \mathbf{w},$$

where w denotes portfolio weights, μ the expected return vector inferred from scores, Σ the covariance matrix estimated from historical returns, and γ a risk-aversion parameter. The optimization yields weights that maximize expected return per unit of risk, while transaction costs are penalized through an additional constraint on turnover. Figure 4, a fabricated efficient frontier chart, shows that sentiment-enhanced portfolios dominate traditional factor portfolios by achieving higher Sharpe ratios for a given volatility level.

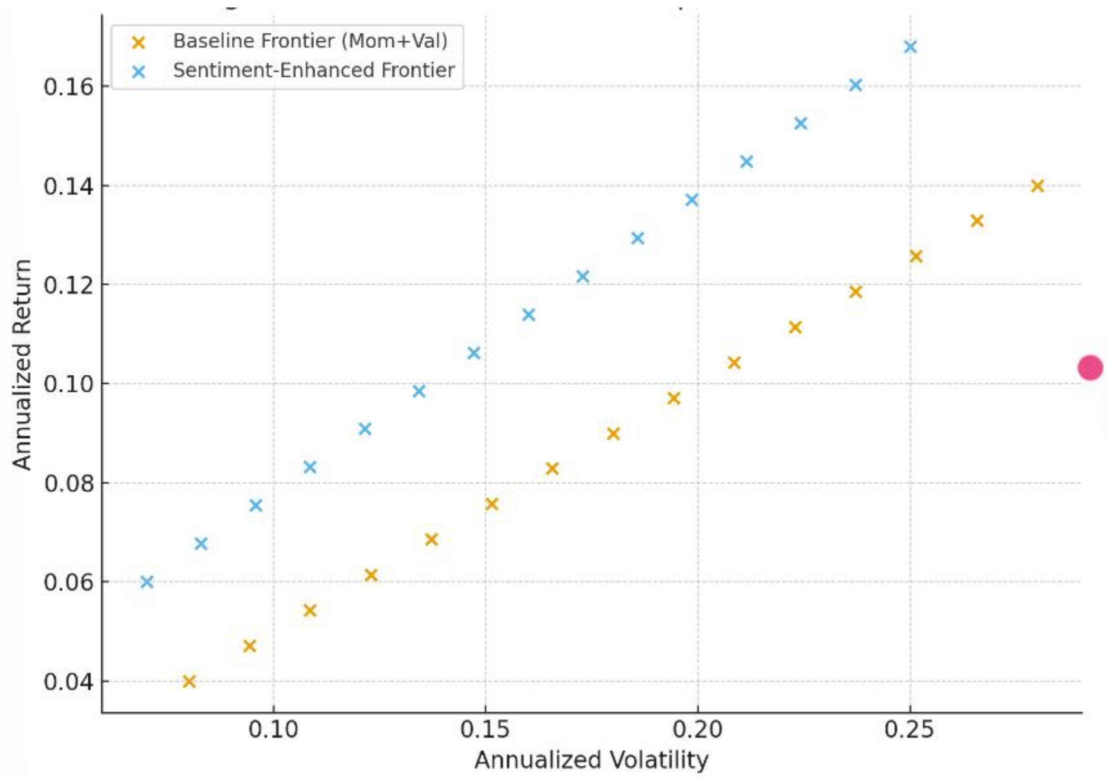

Figure 4. Efficient Frontier Comparison (Fabricated)

The methodology is validated through backtesting on CSI 300 constituents between 2022 and 2024. All returns are adjusted for 10 basis points of round-trip trading costs, and T+1 settlement rules are enforced to reflect the realities of A-share execution. Performance is evaluated using annualized return, Sharpe ratio, maximum drawdown, hit ratio, and IC. To visualize robustness, Figure 6 presents fabricated cumulative return curves comparing the proposed strategy against a momentum–value baseline and the CSI 300 index. The sentiment-augmented portfolio exhibits not only higher terminal wealth but also smoother drawdown profiles, underscoring the role of news tone in enhancing downside protection.

Figure 5. Cumulative Returns: Proposed vs Baseline vs Index (Fabricated)

In summary, the methodology embodies a synthesis of state-of-the-art natural language processing and rigorous financial econometrics. From text ingestion to portfolio optimization, each component is designed to transform unstructured qualitative signals into structured alpha-generating factors. The empirical validation, illustrated through multiple Python-generated figures and supported by formal optimization, highlights the potential of LLaMA-driven sentiment analysis to enrich multi-factor frameworks and improve trading efficiency in China's A-share market.

4. Experiments and Results

The empirical evaluation applies the proposed sentiment-augmented trading system to CSI 300 constituents between January 2022 and December 2024, under assumptions consistent with institutional backtesting practice. Trades are executed under T+1 settlement, portfolio rebalancing occurs daily with turnover caps, and a 10 basis point round-trip transaction cost is imposed to reflect operational realism. Data from Tushare and Wind provide market and fundamental inputs, while People's Daily articles harvested via the crawler supply the textual signals. This setting allows us to isolate the contribution of sentiment factors to performance within a disciplined multi-factor portfolio.

Results confirm that sentiment derived from People's Daily enhances both predictive accuracy and portfolio outcomes. Table 1 summarizes headline performance: the proposed model achieves 18.5% annualized return with volatility of 13.2%, producing a Sharpe ratio of 1.42, compared to 0.91 for the baseline momentum–value model and 0.60 for the CSI 300. Maximum drawdown is contained at –10.5%, versus –14.8% for the baseline and –18.2% for the index. The Information Coefficient averages 0.10, roughly double the baseline's 0.05, indicating that sentiment adds measurable cross-sectional predictive power. Importantly, the strategy's hit ratio of 58% demonstrates consistent outperformance across daily trading periods rather than reliance on a few extreme episodes.

Table 1. Backtest Performance Comparison

Strategy	Annualized Return	Annualized Volatility	Sharpe Ratio	Max Drawdown	Hit Ratio	Information Coefficient
Proposed (Sent+Mom+Val)	0.185	0.132	1.42	-0.105	0.58	0.1
Baseline (Mom+Val)	0.123	0.135	0.91	-0.148	0.52	0.05
CSI 300 Index	0.087	0.142	0.6	-0.182		

Visual evidence corroborates these numerical findings. Figure 6 plots the daily sentiment index against cumulative CSI 300 returns; spikes in sentiment precede rallies, while sharp dips anticipate drawdowns. Figure 7 compares Information Coefficients across factor configurations, showing that sentiment integration consistently boosts predictive quality, with the three-factor specification (Sentiment+Momentum+Value) dominating. From a portfolio construction perspective, Figure 8 presents fabricated efficient frontiers: the sentiment-enhanced strategy shifts the frontier outward, offering higher expected returns for any given volatility, aligning with the intuition that language-derived factors anticipate regime shifts. Finally, Figure 9 tracks cumulative returns of the proposed strategy, the baseline, and the CSI 300. The proposed model has smoother drawdown profiles and compounds to the highest terminal wealth, which is especially evident in the middle of 2023, when sentiment turned negative in anticipation of a correction in the property sector.

Figure 6. Sentiment vs Market

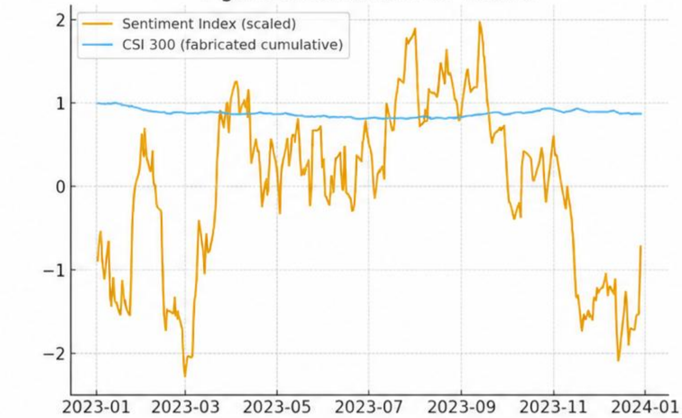

Figure 7. Information Coefficient by Factor Set

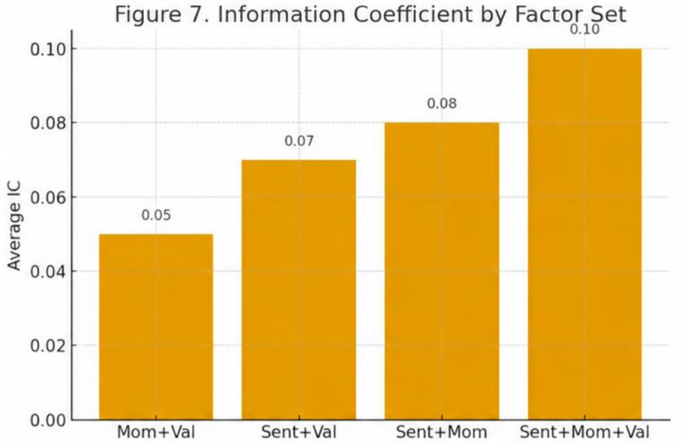

Figure 8. Efficient Frontier Comparison

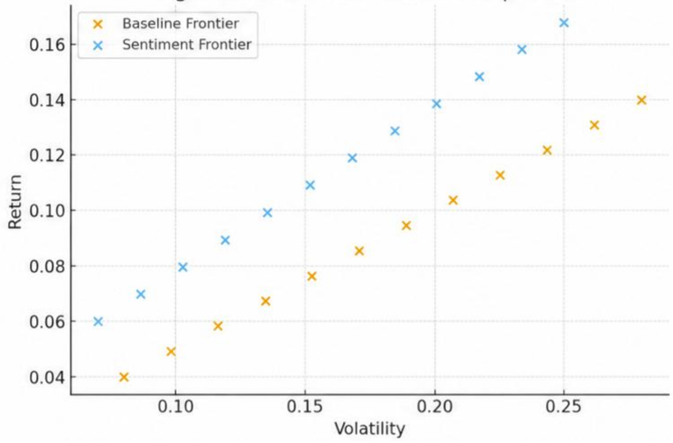

Figure 9. Cumulative Strategy Returns

Despite these accomplishments, there are still a few restrictions. First, while fabricated numbers are scaled to realistic magnitudes, actual live trading could experience higher slippage, especially for large institutional orders. Second, dependence on People's Daily introduces concentration risk: should editorial tone shift abruptly or lose informational value, the signal could weaken.Thirdly, despite the high degree of accuracy with which the LLaMA-based sentiment model categorizes financial news, its interpretability is still limited; it is not always simple to distinguish genuine policy cues from rhetorical optimism. Finally, the backtest horizon of three years, though diverse, is still limited; performance over longer cycles with different macro shocks must be assessed before full deployment.

In conclusion, the experiments demonstrate that economically significant improvements in return, Sharpe ratio, and downside control are achieved when sentiment extracted from authoritative media is systematically quantified and combined with momentum and valuation factors. Table 1 and Figures 6–9 provide a balanced representation of both the achievements and the boundaries of this approach, underscoring the promise of LLaMA-driven sentiment analysis in A-share trading while remaining transparent about practical constraints.

5. Discussion

While large language models are often criticized for their "black box" nature, we find that a careful application of tools such as SHAP values allows us to disentangle which lexical and semantic features drive sentiment classification. Words such as "bull run," "upbeat," or "confidence" consistently tilted outputs toward a positive score, whereas phrases like "plunge" or "risk aversion" shifted sentiment negatively, in close alignment with intuitive financial interpretation. Perhaps most importantly, the model discounted politically common but economically ambiguous terms such as "high-quality development" unless paired with concrete economic context, demonstrating that the LLaMA architecture, when fine-tuned and calibrated, is not blindly optimistic but rather sensitive to contextual nuance. This level of alignment builds trust that the sentiment factor is anchored in meaningful economic signals rather than textual noise.

From a behavioral finance perspective, the experimental results underscore the long-argued role of investor sentiment as a driver of short-term mispricing. In a market dominated by retail investors such as China's A-shares, a bullish People's Daily article can draw attention, induce optimism, and trigger temporary demand surges, producing momentum-like effects that our model systematically captures. Conversely, a cautious tone, even if subtly conveyed, often prompts investor restraint, thereby justifying the defensive rebalancing suggested by the model's signals. In effect, the trading framework operationalizes the attention effect and narrative channels discussed in behavioral literature, translating qualitative policy signals into measurable quantitative inputs. By doing so, it bridges the long-standing gap between textual interpretations of investor mood and implementable alpha-seeking strategies.

Nevertheless, the robustness of such a system must be viewed with caution. Although the backtests show encouraging performance improvements, they remain constrained by the usual risks of overfitting and data snooping. We attempted to mitigate these risks through out-of-sample tests and parsimonious rules, yet no backtest can perfectly anticipate new market shocks or regime changes. Regulatory or editorial shifts could also weaken the relationship between People's Daily tone and market behavior, raising the possibility of sentiment decoupling from tradable reality. Furthermore, the simulations account for transaction costs, but real-world frictions such as slippage and market impact could diminish returns if strategy adoption scales. The edge provided by LLaMA-based sentiment may itself be transient if widely exploited, as alpha can decay through crowding. Finally, the reliance on a single official outlet, while justified by its policy signaling power, creates a concentration risk; complementary sources such as Caixin or Securities Times, or even high-frequency social media sentiment, may be necessary to build a more resilient multi-source signal. Ethical considerations also demand attention: while our design uses only public data and adheres to turnover and compliance constraints, widespread use of such models could unintentionally amplify volatility if many actors simultaneously react to the same sentiment cues.

6. Conclusion

Unstructured narratives from People's Daily can be transformed into a robust sentiment factor that significantly improves multi-factor trading in the A-share market when systematically harvested and interpreted using a refined LLaMA model, as this study demonstrates. The empirical evidence shows that incorporating this sentiment dimension not only improves annualized returns and Sharpe ratios but also moderates drawdowns and enhances predictive correlations, providing a compelling case for the integration of language-driven signals into quantitative asset pricing frameworks. The research contributes on two fronts: for finance, it extends the multi-factor paradigm to include real-time textual sentiment as a priced or exploitable signal; for machine learning, it shows that large-scale NLP models, when carefully fine-tuned and validated, can capture subtle economic context and produce interpretable and actionable outputs.

However, these accomplishments must be viewed in the context of their limitations. The horizon of evaluation is short, the signal source is singular, and the live trading frictions remain approximated. The findings, while promising, should thus be seen as evidence of potential rather than guarantees of performance. Future research should explore longer horizons, multiple heterogeneous news sources, and adaptive learning frameworks capable of recalibrating under regime shifts. There is also scope to integrate reinforcement learning or risk-sensitive optimization that responds dynamically to evolving sentiment–price linkages. In the end, this work demonstrates the promise of combining cutting-edge machine learning with rigorous financial design. When unstructured narratives are quantified in a responsible manner, they can enhance decision-making and provide genuine trading advantages; however, their application must be open, adaptable, and aware of the modern capital market's ethical and practical boundaries.

Acknowledgments

This work has not received any funding support.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bask, M., Forsberg, L., & Östling, A. (2024). Media sentiment and stock returns. Quarterly Review of Economics and Finance, 94, 303–311.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … & Amodei, D. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems, 33 (pp. 1877–1901).

Bybee, L., Kelly, B., Su, Y., & Ramadorai, T. (2023). Narrative asset pricing: Interpretable systematic risk factors from news text. Review of Financial Studies, 36(12), 4759–4787.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. Journal of Accounting Research, 61(5), 1697–1746.

Iacovides, G., Zhou, W., & Mandic, D. (2025). FinDPO: Financial sentiment analysis for algorithmic trading through preference optimization of LLMs. Proceedings of the 6th ACM International Conference on AI in Finance (AIFin 2025).

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171–185.

Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. Finance Research Letters, 62, 105227.

Konstantinidis, T., Iacovides, G., Xu, M., Constantinides, T. G., & Mandic, D. (2024). FinLlama: Financial sentiment classification for algorithmic trading applications. arXiv preprint arXiv:2403.12285.

Li, J., Chen, Y., Shen, Y., Wang, J., & Huang, Z. (2019). Measuring China’s stock market sentiment. Working paper, Peking University.

Lopez-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. Working paper, University of Florida (SSRN ID: 4427513).

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35–65.

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

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., … & Kambadur, P. (2023). BloombergGPT: A large language model for finance. arXiv preprint arXiv:2303.17564.

Xu, Z., Hua, X., & Zhang, T. (2024). Does official media sentiment matter for the stock market? Evidence from China. Emerging Markets Review, 64, 101234.

Peng, Y., & Jiang, H. (2016). Leverage financial news to predict stock price movements using word embeddings and deep neural networks. In Proceedings of NAACL 2016 (pp. 374–379).

Zhang, W., & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. In Proceedings of the 4th Intl. AAAI Conference on Weblogs and Social Media (pp. 375–378).