Integrating Large Language Models and Reinforcement Learning for Sentiment-Driven Quantitative Trading

Wo Long, Victor Xiao
Columbia Business School, Columbia University

May 16, 2025

Outline

Introduction and Objectives

Background:

- Sentiment signals offer a forward-looking view, complementary to backward-looking technical indicators.
- Combining these distinct signals effectively remains a challenge.
- This paper proposes a reinforcement learning–driven framework to dynamically integrate sentiment and technical signals, and evaluates its effectiveness through portfolio performance.

Objectives:

1. Examine whether sentiment signals from financial news and generated by LLM have predictive power over stock returns.
2. Evaluate whether sentiment signals enhance the traditional trading strategies based on technical indicators.
3. Assess whether reinforcement learning effectively integrates sentiment and technical signals into a dynamic trading strategy.

Literature Review

Three key fields of literature: Design of the Framework, LLM, RL and Look-ahead bias.

- Zhou and Mehra (2025) — **End-to-end trading system** that leverages LLMs for sentiment analysis using data from financial news and social media.
- Liu et al. (2023) — **FinGPT**, an open-source, domain-specific LLM that offers simple and effective strategies for fine-tuning on alternative data sources.
- Liu et al. (2020) — **FinRL**, RL framework that 1) automates stock trading using real market environments; 2) incorporates constraints like transaction costs and risk aversion.
- Sarkar and Vafa (2024), Kim et al. (2024) — Methods to mitigate **look-ahead bias** include masking company- and time-specific entities, Out-of-Sample back-testing, *etc.*

Contribution

- Introduce FinGPT-generated sentiment signals into the trading framework.
- Evaluates whether reinforcement learning effectively translates LLM-generated sentiment and technical signals into actionable trading strategies.
- Address look-ahead bias through entity masking and out-of-sample (OOS) back-testing.
- Evaluate performance using real market data from 2018–2025. Compare rule-based vs. RL-based integration of signals.

Data

News Data:

- Source: Thomson Reuters from 2018 to 2025.
- Universe: 44 Stocks in SP 500 with the most active news coverage.
- Granular Level: Daily (trading day timestamp) + Company + News Piece (One company may have multiple news coverage per day).
- Drop news articles released after 4 p.m. on each trading day.
- Employ LLaMA 8B to condense raw news articles into concise, company-level daily summaries.

Stock Data:

- Source: CRSP, Bloomberg from 2018 to 2025.
- Include: Price (close, open, *etc.*), trading volume (shares)
- Granular Level: Daily (trading day timestamp) + Company
- Price and volume are adjusted for dividend, stock split, and other corporate actions.

LLM Pipeline

$$(\text{News}_{i,t}^{\text{raw}},\ \text{Metadata}) \xrightarrow{\text{LLaMA 8B + vLLM}} \text{Summary}_{i,t} \xrightarrow{\text{FinGPT}} (\text{Sentiment}_{i,t},\ \text{Confidence}_{i,t})$$

LLM Trading System Flowchart

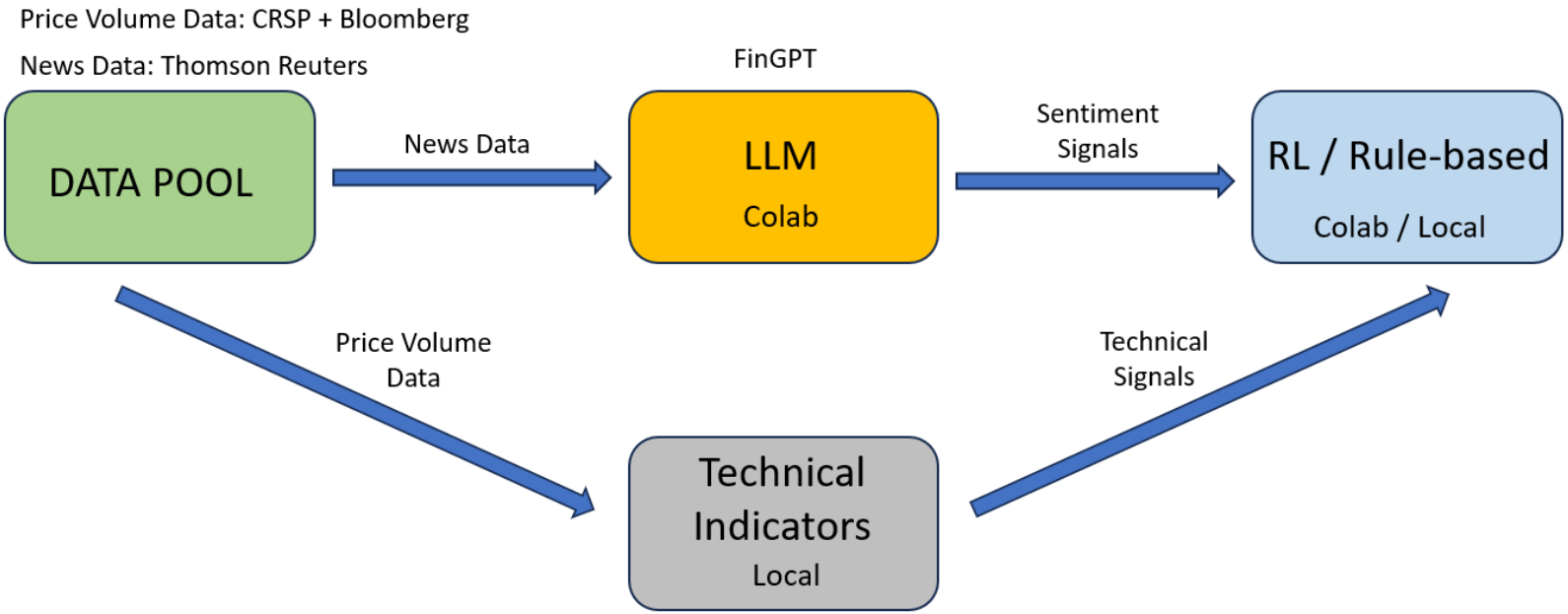


Figure: Trading System Illustration

Large Language Model: FinGPT

FinGPT:

- Pre-trained on data available up to November 2023. The company-level daily news summarized by Llama 8b then processed by FinGPT for sentiment classification.
- Each summary piece is categorized into one of three sentiment classes: [Positive, Neutral, Negative], along with a confidence score (logit). The confidence score is a key input to both rule-based and RL-driven trading strategies.

Key Points:

- Sentiment signal $sentiment_{i,t}$ for company i on trading day t is generated using the summarized news of company i on trading day t.
- Note that the summarization on trading day t only involves news released before 4 p.m. on the same day.

Efficient Inference with vLLM

Motivation: Summarizing 80k+ news articles demands high throughput and low latency.

vLLM Framework :

- Utilizes token-level continuous batching and *paged attention* to minimize GPU memory fragmentation.
- Delivers up to 4× higher throughput than HuggingFace Transformers for comparable model sizes.
- Integrates seamlessly with our LLaMA 8B summarizer, enabling thousands of article summaries per GPU-hour.

Implementation in This Study:

- Batched 128 article chunks per forward pass; max tokens set to 1024.
- End-to-end latency < 200 ms per article, summaries cached and streamed to FinGPT.
- Sampling temperature $= 0.3$, top-p $= 0.9$ to ensure factual, focused outputs.

Addressing Look-ahead Bias

To address potential look-ahead bias:

- Set aside the 2024–2025 period for out-of-sample testing. Since FinGPT is pre-trained only on data through 2023, this helps mitigate look-ahead bias.
- Anonymize news articles by masking company names, product names, and time-specific information, making it virtually impossible for the model to infer firm identities from text structure.
- Conduct robustness checks to ensure that sentiment signals are not driven by leaked future information.

Technical Indicators

We generate technical signals from stock price and volume data (2018–2025), then integrate them with sentiment scores:

- Technical Indicators:
  - **RSI:** Overbought/oversold signals.
  - **VWAP:** Intraday value benchmark.
  - **MACD:** Trend strength and momentum.
  - **Garman-Klass Volatility:** Efficient volatility estimator.
  - All features z-score normalized. Signals built using $t-1$ data.
- Trading volume and realized volatility metrics are also computed
- These features are then combined with sentiment analysis outputs for downstream trading decisions via our RL model.

Rule-based Strategy:

We construct rule-based signals by combining technical indicators and sentiment data:

- **Raw Score Construction:**

$$\textit{Alpha Signal}_{i,t} = \frac{\textit{Volume}_{i,t-1}}{\textit{MA20}_{i,t-1}} + \textit{MACD}_{i,t-1}$$

$$\textit{Sentiment Signal}_{i,t} = \textit{Confidence}_{i,t} \cdot I_{\{1:\textit{Positive};\ -1:\textit{Negative};\ 0:\textit{Neutral}\}}$$

✓ Both raw $\textit{Alpha Signal}_{i,t}$ and $\textit{Sentiment Signal}_{i,t}$ are z-scored again after construction.

- **Sentiment Integration:**

$$\textit{Combined Signal}_{i,t} = w_t \cdot \textit{Alpha Signal}_{i,t} + (1 - w_t) \cdot \textit{Sentiment Signal}_{i,t}$$

Results

We use a universe of 44 stocks selected in S&P 500 from 2018 to 2025, setting aside data from 2024 to 2025 for back-testing and evaluating the out-of-sample performance of the proposed strategies.

Execution Assumption:

- Orders are executed at the same day's closing price.
- Long-only Strategy with initial investment of \$1 million, a level considered small enough to avoid exerting significant market impact.
- Transaction cost is set conservatively at 5bps.

Caveats:

- Trades are executed on trading day t using the close price $close_{i,t}$.
- A transaction cost of 5 basis points is applied to each trade - high and conservative for the large-cap stocks in the portfolio - mitigate potential overstatement of performance results.

OOS Performance - Rule-based Strategy

Trading Rules: The *Combined Signal*$_{i,t}$ serves as the ranking basis for long-short portfolio construction. Stocks are ranked by the signal each day and assigned into 5 quintiles. The following portfolio longs the top 20% stocks.

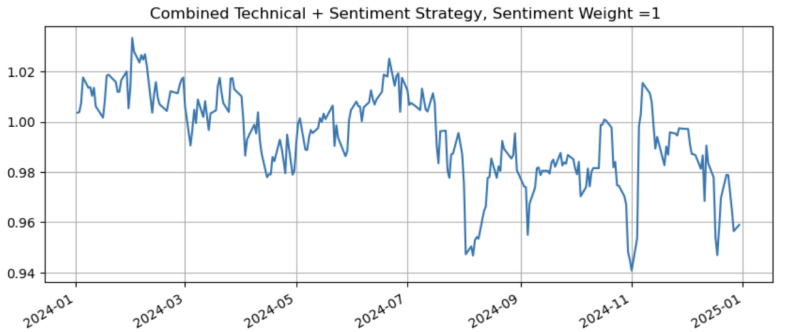
Strategy	Ret (%)	Vol (%)	Sharpe	Sortino	MDD
Sentiment weight = 0	12.85	16.09	0.83	0.07	-0.099
Sentiment weight = 0.5	6.95	15.27	0.52	0.05	-0.114
Sentiment weight = 1	-4.13	13.03	-0.26	-0.02	-0.090

✓ All performance metrics are annualized.

OOS Performance - Combined Strategy

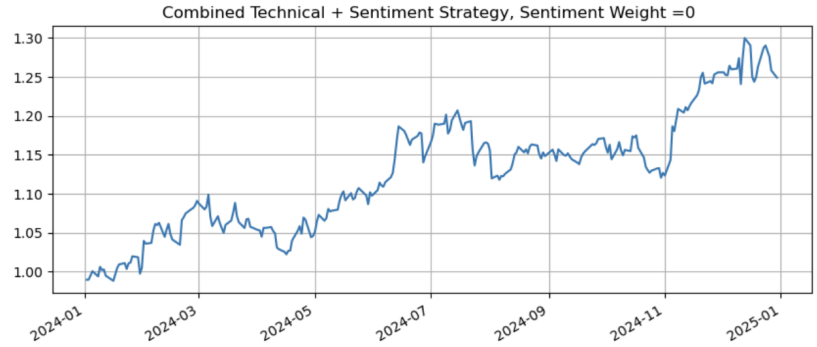
	Annualized Return	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown
0	0.128521	0.160856	0.830893	0.074498	−0.098813

	Annualized Return	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown
0	0.069511	0.152696	0.51541	0.047206	−0.113978

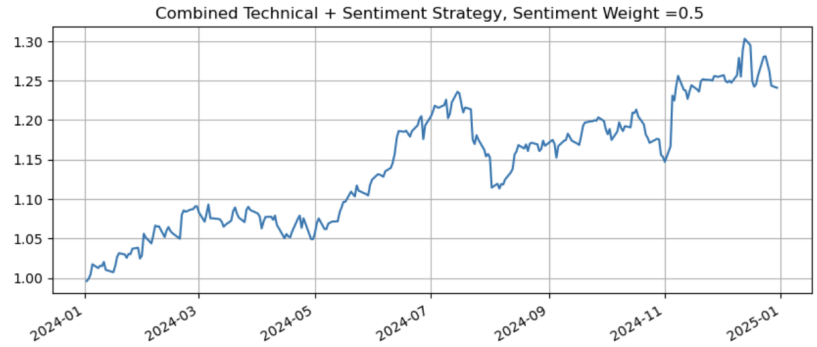


	Annualized Return	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown
0	−0.041259	0.130251	−0.258214	−0.024003	−0.089669

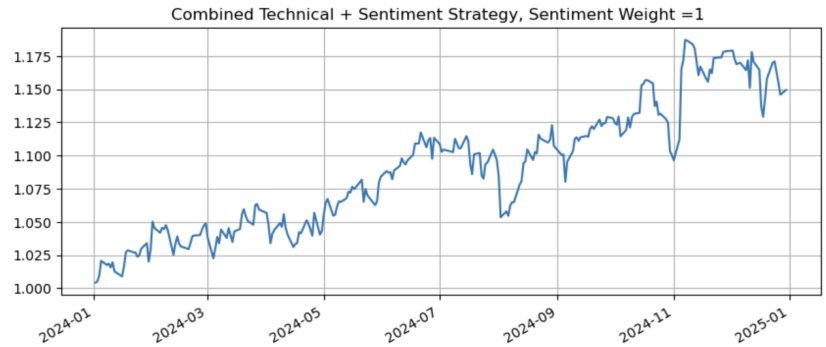
With Transaction Cost



	Annualized Return	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown
0	0.25069	0.160787	1.469892	0.132237	−0.073853



	Annualized Return	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown
0	0.242362	0.152606	1.496335	0.136718	−0.099225



	Annualized Return	Volatility	Sharpe Ratio	Sortino Ratio	Max Drawdown
0	0.150229	0.130565	1.135495	0.106772	−0.057201

Without Transaction Cost

OOS Performance - Rule-base Strategy

FF-5 Decomposition for OOS period:

Strategy / Coef	Market	SMB	HML	RMW	CMA	UMD
Sentiment weight = 0	0.03	0.07	-0.07	0.13	0.15	0.08
Sentiment weight = 0.5	0.04	0.11	-0.03	0.16	0.15	0.17
Sentiment weight = 1	-0.01	0.07	0.02	-0.05	0.11	0.17

- Strategy without transaction costs is used for the 2024–2025 regression period.
- No coefficients are statistically significant.
- R^2 values range from 0.01 to 0.02.

Reinforcement Learning: TD3 Algorithm

Twin Delayed Deep Deterministic Policy Gradient (TD3) is an actor-critic method that addresses overestimation bias in deterministic RL algorithms.

- **Two Critic Networks:** Maintains two Q-value estimators to reduce overestimation.
- **Policy Smoothing:** Adds noise to target policy actions for more stable learning.
- **Why TD3 for Trading?**
  - Continuous-action framework suits portfolio allocation, where asset weights can vary smoothly.
  - Robust to noisy reward signals and market volatility.
  - Model-free approach allows flexible adaptation to different market conditions.

Reinforcement Learning: State

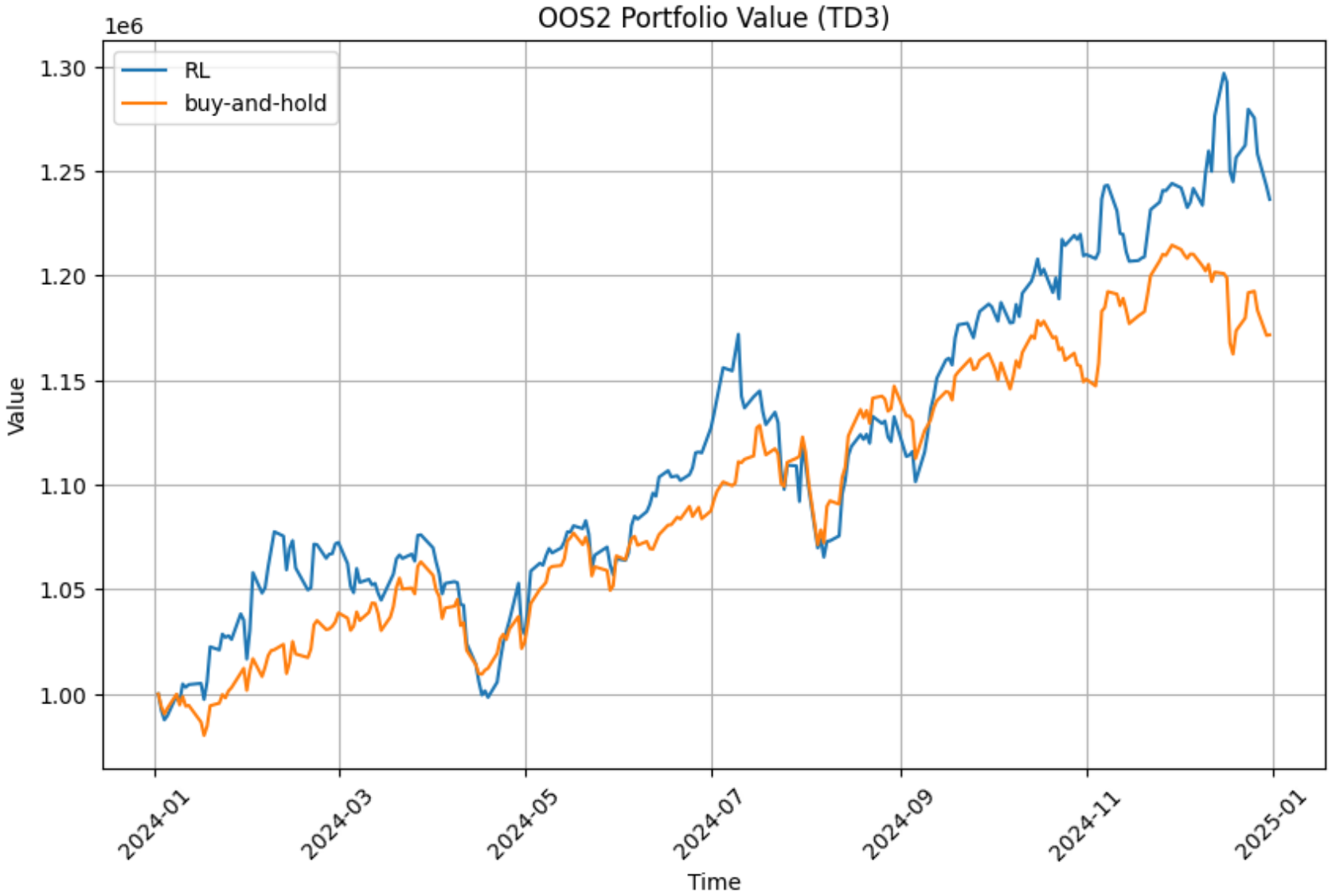
- **Reward:** Portfolio return - transaction cost (10 bps)
- **State (observation):**
  - r_{t-1}
  - RSI_{14}
  - MACD
  - Sentiment score
  - $\frac{\mathrm{Volume}_t}{\overline{\mathrm{Volume}_{20}}}$
  - $\frac{\mathrm{RV_t}}{\overline{\mathrm{RV}_{20}}}$
  - $\mathrm{GK\ Volatility}$
  - $\mathit{PortfolioWeight}'_{t-1}$

Reinforcement Learning Transition

- At t, observe portfolio weight W_t, O_t
- Take action A_t to adjust portfolio weight to W_t'
- Market moves from t to $t+1$, generate reward R_{A_t} and new portfolio weight W_{t+1}

$$(W_t,\ O_t) \xrightarrow{\text{Action}} (A_t,\ W_t') \xrightarrow{\text{Market move}} (W_{t+1},\ R_{A_t})$$

RL OOS Performance



RL

Out-of-Sample Performance

Performance Comparison

Metric	**RL**	**Buy & Hold**
Annualized Return (%)	23.65	17.17
Annualized Volatility (%)	13.46	10.06
Sharpe Ratio	1.38	1.20
Sortino Ratio	1.96	1.59
Turnover (%)	52.27	0.00
Max Drawdown (%)	–9.09	–5.06

FF5 Regression Summary

- $R^2 = 0.657$, $\text{Adj } R^2 = 0.648$; no intercept alpha ($\hat{\alpha} \approx 0$, $p = 0.934$)
- Significant loadings:

$$\beta_{\text{mktrf}} = 0.91^{***}, \quad \beta_{\text{smb}} = -0.17^{**}, \quad \beta_{\text{umd}} = -0.14^{**}$$

- Other factors: $\beta_{\text{hml}}, \beta_{\text{rmw}}$ insignificant; β_{cma} marginal ($p = 0.09$)
- **Conclusion:** Returns largely driven by market exposure due to long-only constraint, with defensive (large-cap) and negative momentum tilts.

Conclusion and Discussion

- The sentiment-based strategy exhibits a high turnover rate and its performance declines significantly when transaction costs are applied.
- The project can be accessed through https://github.com/mchlong/GR5293

Limitation and Future Direction

- Due to the lack of GPU resources, the current analysis is limited to a smaller universe of stocks. We focus on mega-cap stocks, as they receive more consistent news coverage—an important consideration given our limited access to diverse news sources. Expanding the universe to include smaller-cap stocks would require additional data sources to ensure adequate sentiment signal quality.
- Augment the pipeline with chain-of-thought reasoning to further enhances the contextual understanding of a given news feed. However, computation costs will also needs to be carefully considered.

References

- Bernard, E., Blankespoor, E., & deHaan, E. (2023). *Using GPT models to measure the complexity of business transactions.* SSRN.
- Lopez-Lira, A., & Tang, Y. (2023). *Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models.* SSRN.
- Glasserman, P., Mamaysky, H., & Qin, J. (2023). *New News is Bad News.* SSRN.
- Zhou, Z., & Mehra, R. (2025). *An End-To-End LLM Enhanced Trading System.* arXiv:2502.01574.
- Kim, S., Li, S., & Zhang, Y. (2024). *Financial Statement Analysis with Large Language Models.* Becker Friedman Institute.
- Sarkar, S., & Vafa, K. (2024). *Lookahead Bias in Pretrained Language Models.* SSRN.
- Liu, X.-Y., et al. (2020). *FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance.* GitHub.
- Liu, X.-Y., et al. (2022). *FinRL-Meta: Market Environments and Benchmarks for Data-Driven Financial Reinforcement Learning.* NeurIPS.

Github Repo Link

https://github.com/mchlong/GR5293