Leveraging LLM-based sentiment analysis for portfolio optimization with proximal policy optimization

Abstract

Portfolio optimization requires adaptive strategies to maximize returns while managing risk. Reinforcement learning (RL) has gained traction in financial decision-making, with Proximal Policy Optimization (PPO) demonstrating strong performance in dynamic asset allocation. However, traditional PPO relies solely on historical price data, ignoring market sentiment, which plays a crucial role in asset movements. We propose a sentiment-augmented PPO (SAPPO) model that integrates daily financial news sentiment extracted from Refinitiv using LLaMA 3.3, a large language model optimized for financial text analysis. The sentiment layer refines portfolio allocations by incorporating real-time market sentiment alongside price movements. We evaluate both PPO and SAPPO on a three-stock portfolio consisting of Google, Microsoft, and Meta, comparing performance against standard market benchmarks. Results show that SAPPO improves risk-adjusted returns, demonstrating superior Sharpe ratios and reduced drawdowns. Our findings highlight the value of integrating sentiment analysis into RL-driven portfolio management.

1 Introduction

Portfolio optimization is a central task in financial management, aiming to allocate resources across various assets to achieve maximum returns while minimizing risks. Traditionally, this problem has been addressed using techniques such as mean-variance optimization (MVO), rooted in modern portfolio theory (Markowitz, 1952). However, these static methods rely on historical data to estimate returns and covariances, often failing to adapt dynamically to changing market conditions.

The advent of reinforcement learning and deep reinforcement learning (DRL) has introduced new possibilities in solving sequential decision-making problems, including portfolio optimization. RL agents learn optimal strategies by interacting with an environment, while DRL incorporates deep neural networks to approximate complex value functions and policies. In financial contexts, DRL methods have demonstrated effectiveness in tasks such as trading, hedging, and portfolio management by leveraging the ability to learn directly from data and adapt to non-linear and non-stationary environments (Deng et al., 2016; Ye et al., 2020).

Moreover, the integration of DRL into portfolio optimization has enabled dynamic adaptability to market conditions, addressing a key limitation of traditional approaches. Techniques such as Proximal Policy Optimization (PPO) and deep Q-networks (DQN) have been widely adopted in financial research, providing robust frameworks for tackling continuous state and action spaces (Sutton and Barto, 2018; Wang et al., 2019). PPO optimizes portfolio weights through stable policy updates, allowing efficient learning of asset allocation strategies.

In this study, we first implement and evaluate PPO as a reinforcement learning framework for portfolio optimization. We analyze its performance using historical financial data and assess its ability to generate optimal portfolio allocation strategies. However, while PPO effectively captures price-based patterns, it does not explicitly account for market sentiment, which plays a crucial role in investor behavior and asset price movements (Tetlock, 2007).

We introduce a sentiment layer as an extension to PPO to address this limitation. This layer integrates real-time sentiment analysis into the reinforcement learning framework, allowing the model to incorporate external financial news sentiment into its decision-making process. Specifically, we use Refinitiv's financial news database (Refinitiv, 2024) to retrieve daily market news articles related to portfolio assets. These articles are processed using Llama 3.3, a large language model (LLM) optimized for financial text analysis, which generates a daily sentiment score for each stock (AI, 2024)

By comparing vanilla PPO with sentiment augmented PPO (SAPPO), we evaluate the impact of integrating financial news sentiment on portfolio performance. Our analysis focuses on key financial metrics, including Sharpe ratio (Sharpe, 1994), annualized returns, and drawdowns, to determine whether sentiment-based reinforcement learning enhances trading strategies. We contribute to the growing body of research on sentiment-aware reinforcement learning in finance, demonstrating how market sentiment can improve adaptability and robustness in portfolio optimization.

2 Background

Recent advancements in DRL have significantly influenced financial decision-making, particularly in portfolio optimization. DRL-based approaches allow adaptive portfolio management by learning optimal strategies through interactions with the market environment. PPO has emerged as a leading policy-gradient method in financial reinforcement learning, offering a balance between exploration and exploitation when optimizing portfolio weights over time (Sutton and Barto, 2018; Schulman et al., 2017).

However, traditional reinforcement learning models, including PPO, rely solely on structured price and volume data, overlooking market sentiment, which plays a crucial role in investor behavior, risk perception, and asset price movements (Tetlock, 2007; Baker and Wurgler, 2012; Huang et al., 2023). Financial markets are highly sensitive to external information, and news-driven events often create price fluctuations that historical price-based models fail to anticipate. This limitation has led to an increasing focus on integrating sentiment analysis into quantitative finance (Zhang et al., 2020; Chen et al., 2022).

The integration of natural language processing (NLP) and LLMs has significantly expanded the scope of reinforcement learning in financial applications. Traditional DRL models process structured numerical data but fail to leverage the vast amount of unstructured textual information, such as financial news articles, earnings reports, and analyst opinions (Lopez-Lira and Tang, 2023; Ke et al., 2019). Sentiment analysis has been widely studied in finance, demonstrating that market sentiment can improve return predictions, volatility forecasting, and risk-adjusted performance (Chen et al., 2022; Smales, 2014; Jin and Wang, 2023).

Recent research highlights the increasing role of transformer-based architectures in extracting insights from textual data. Domain-specific financial LLMs, such as FinBERT (Araci, 2019), have demonstrated superior sentiment classification performance compared to general-purpose models. The development of LLaMA 3.3 (Dubey et al., 2024) further enhances the ability to process and interpret financial news. LLaMA 3.3 is a decoder-only transformer model fine-tuned for financial text analysis, trained on a large corpus of earnings reports, market commentary, and analyst opinions. This specialization allows it to distinguish between neutral reporting, speculative opinions, and sentiment-driven market movements, making it a valuable tool for financial reinforcement learning applications.

The increasing adoption of LLM-based sentiment analysis in financial markets suggests that multi-modal approaches, incorporating both structured time-series data and unstructured textual information, may improve decision-making in algorithmic trading. Studies by Kirtac and Germano (2024) explore how financial LLMs enhance textual analysis by contextualizing financial narratives, reducing ambiguity, and refining sentiment-driven signals for investment strategies. The work of Bollen et al. (2011) further supports the idea that social media and financial news sentiment impact market returns, reinforcing the need to integrate sentiment-aware models into trading algorithms.

These developments highlight the potential for reinforcement learning frameworks to integrate financial sentiment as an additional market signal, moving beyond conventional price-based optimization techniques. The introduction of sentiment-aware reinforcement learning models, such as SAPPO, marks a shift toward more adaptive and information-rich portfolio optimization strategies. This approach aligns with existing studies demonstrating that hybrid sentiment-price models outperform purely historical price-based strategies (Ding and Duan, 2015; Liu and Zhu, 2020; Dai and Li, 2022).

3 Methodology

The financial market is represented by a state array $\mathbf{s}_n$ that consists of a vector of current portfolio weights $\mathbf{w}_n$ and a vector of current adjusted closing spot prices $\mathbf{S}_n$ for multiple assets. This setup allows decision-making based on both the agent's portfolio position and overall market conditions (Markowitz, 1952; Sutton and Barto, 2018). The discrete index $n = \lfloor t/\Delta t \rfloor$ counts trading days, where t is time and $\Delta t = 1$ day. Additionally, the agent maintains a cash account.

At the end of each trading day, adjusted closing prices are observed, daily returns are computed, and new allocation weights are selected. The portfolio is rebalanced on the next morning using a market-order strategy with trades executed at the volume-weighted average price (VWAP) over the first ten minutes of the trading session. This approach mitigates volatility effects associated with raw opening prices. The action $\mathbf{a}_n$ is the change in portfolio holdings at day n, where positive values indicate purchases and negative values denote sales

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{a}_n \tag{1}$$

To achieve self-financing, the total trade value sums to zero

$$\mathbf{a}_n \cdot \mathbf{S}_n = 0 \tag{2}$$

Transaction costs of 0.05% of the total turnover are deducted to account for realistic market frictions.

The immediate reward is the logarithmic portfolio return, which is scale-invariant

$$x_{n+1} = \log \frac{\mathbf{w}_n \cdot \mathbf{S}_{n+1}}{\mathbf{w}_n \cdot \mathbf{S}_n} \tag{3}$$

Alternatively, the relative return R_{n+1} can be used, with the relationship

$$x_{n+1} = \log(1 + R_{n+1}) \tag{4}$$

Both return formulations approximate each other for small values, ensuring stable reinforcement learning.

The state-action-value function $Q(\mathbf{s}_n, \mathbf{a}_n)$ represents the expected cumulative discounted future reward at time step n

$$Q(\mathbf{s}_n, \mathbf{a}_n) = E\left[\sum_{k=1}^{\infty} \gamma^k x_{n+k} \,\middle|\, \mathbf{s}_n, \mathbf{a}_n\right] \tag{5}$$

This function estimates how much cumulative reward can be obtained from state $\mathbf{s}_n$ onward when sampling the action $\mathbf{a}_n$ from a conditional probability density function $\pi(\mathbf{a}_n|\mathbf{s}_n)$, given the transition probability density function $p(\mathbf{s}_{n+1}|\mathbf{s}_n, \mathbf{a}_n)$ (Sutton and Barto, 2018). The discount factor $\gamma \in (0, 1]$ determines the importance of future rewards. A higher γ prioritizes long-term rewards, while a lower γ focuses on immediate returns. We set $\gamma = 0.99$.

Deep reinforcement learning employs deep neural networks to approximate the state-action-value function Q and to learn the policy π (Sood et al., 2023). We implement this with PPO that operates in continuous action spaces, enabling dynamic and adaptive portfolio rebalancing based on market trends. PPO uses a multivariate Gaussian distribution for the stochastic policy π, with the constraint in Eq. (2) ensuring that trades remain budget-neutral. The mean and covariance functions of this Gaussian distribution are learned by a deep neural network with parameters $\boldsymbol{\theta}$.

3.1 Experiment 1: PPO for Portfolio Optimization

PPO is a policy-gradient algorithm that iteratively refines the stochastic policy π (Schulman et al., 2017). The optimization objective function is defined as

$$L_{\text{PPO}} = E_{\mathbf{s}_n,\mathbf{a}_n}\left[A(\mathbf{s}_n, \mathbf{a}_n) \min\left(r(\mathbf{s}_n, \mathbf{a}_n), 1+\epsilon\right)\right] \tag{6}$$

where the advantage function $A(\mathbf{s}_n, \mathbf{a}_n)$ measures how much better an action $\mathbf{a}_n$ is compared to the expected baseline

$$A(\mathbf{s}_n, \mathbf{a}_n) = Q(\mathbf{s}_n, \mathbf{a}_n) - V(\mathbf{s}_n) \tag{7}$$

The policy ratio $r(\mathbf{s}_n, \mathbf{a}_n)$ quantifies the change between old and new policies

$$r(\mathbf{s}_n, \mathbf{a}_n) = \frac{\pi(\mathbf{a}_n|\mathbf{s}_n; \boldsymbol{\theta})}{\pi(\mathbf{a}_n|\mathbf{s}_n; \boldsymbol{\theta}_{\text{old}})} \tag{8}$$

3.2 Experiment 2: SAPPO

To enhance portfolio optimization, we introduce SAPPO, a sentiment-augmented extension of PPO that integrates real-time sentiment analysis into reinforcement learning. This framework allows the model to incorporate external financial news sentiment into decision-making.

Sentiment data from Refinitiv is processed using LLaMA 3.3, a financial LLM fine-tuned for market sentiment analysis (Face, 2024). The state representation in SAPPO is expanded to include the daily sentiment score $\mathbf{m}_n$ for each asset, yielding the augmented state

$$\mathbf{s}_n = (\mathbf{w}_n, \mathbf{S}_n, \mathbf{m}_n) \tag{9}$$

where $\mathbf{m}_n$ is a normalized sentiment score in the range $[-1, 1]$.

The SAPPO model adjusts the action policy $\pi(\mathbf{a}_n|\mathbf{s}_n)$ by incorporating a sentiment-adjusted advantage function

$$A'(\mathbf{s}_n, \mathbf{a}_n) = A(\mathbf{s}_n, \mathbf{a}_n) + \lambda \mathbf{m}_n \tag{10}$$

where λ is a hyperparameter controlling the influence of sentiment on portfolio allocations.

To prevent redundant sentiment information, a cosine similarity check ensures that only unique financial news articles contribute to the final sentiment score

$$\text{sim}(m_i, m_j) = \frac{m_i \cdot m_j}{\|m_i\| \|m_j\|} \tag{11}$$

Articles with similarity scores exceeding 0.8 within a 20-day rolling window are removed.

The SAPPO agent generates allocation decisions at market close, using both price movements and sentiment scores. Orders execute at the VWAP of the first 10 minutes of the next trading day.

3.3 Evaluation Metrics

The performance of PPO and SAPPO is evaluated using cumulative portfolio return, Sharpe ratio, maximum drawdown, and turnover. We compare model performance against benchmarks including the S&P 500, Dow Jones Industrial Average (DJI), and NASDAQ-100 (Wang et al., 2019). The Sharpe ratio quantifies risk-adjusted returns, while maximum drawdown measures peak-to-trough declines. Portfolio turnover is assessed to determine trading frequency.

A comparative analysis of PPO and SAPPO evaluates the effectiveness of integrating sentiment analysis into reinforcement learning-based portfolio management.

4 Experiments and Results

The evaluation involves backtesting the trained DRL agent on unseen market data, benchmarking its performance against traditional strategies such as buy-and-hold and equal-weighted portfolios (Jansen, 2000). Performance metrics include annualized returns, which measure the compound growth rate of the portfolio, the Sharpe ratio, which evaluates risk-adjusted returns by considering volatility (Sharpe, 1994), and maximum drawdown, which quantifies the largest peak-to-trough decline in portfolio value (Sortino and Van Der Meer, 1994). Turnover is also assessed to measure the frequency of portfolio rebalancing, as this impacts transaction costs (Treynor, 1966).

To ensure robust evaluation, we partitioned the dataset into training (90%) and testing (10%) periods. The training phase allows the DRL agent to learn optimal allocation strategies, while the testing phase evaluates generalization capabilities in unseen market conditions. The backtesting framework accounts for transaction costs, slippage, and liquidity constraints to provide a realistic assessment of the models' applicability to financial markets.

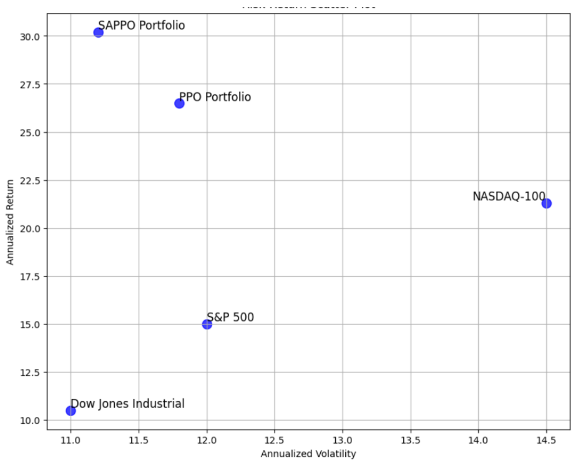

Figure 1: Risk-return scatter plot as of January 1, 2020, for the SAPPO, PPO portfolios and NASDAQ-100, DJI, S&P 500 indexes.

The reinforcement learning agent demonstrates strong performance across multiple evaluation metrics. The annualized return of the SAPPO portfolio reaches approximately 31%, while the PPO portfolio achieves around 25%. Both portfolios outperform major benchmark indices, including the NASDAQ-100 (20%), the S&P 500 (15%), and the DJI (10%). The risk-return scatter plot (Figure 1) highlights SAPPO's superior positioning in terms of volatility-adjusted returns, followed by PPO. Compared to traditional indices, SAPPO and PPO exhibit higher returns but at the cost of

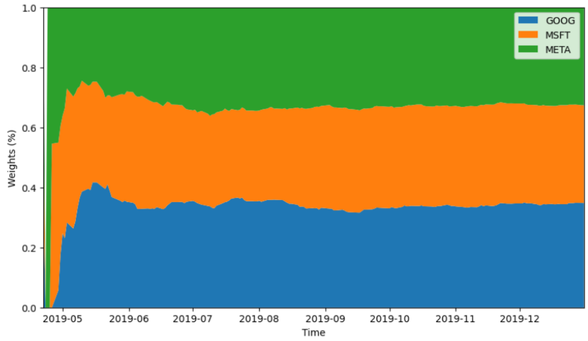

Figure 2: Portfolio weight allocation over time for the PPO portfolio, showing the distribution across Google, Microsoft, and Meta stocks.

increased volatility, indicating their ability to exploit market inefficiencies more effectively. The Sharpe ratio of SAPPO surpasses that of PPO and all benchmark indices, confirming its improved risk-adjusted performance and highlighting the effectiveness of sentiment-aware reinforcement learning in portfolio optimization (Fama and MacBeth, 1973).

Portfolio allocation dynamics (Figure 2) reveal how the PPO agent adjusts asset weights over time. The model increases exposure to Microsoft during high-volatility periods, capitalizing on its stability, while balancing Google and Meta allocations for diversification. This adaptive reallocation highlights the agent's ability to respond to market changes dynamically (Markowitz, 1952).

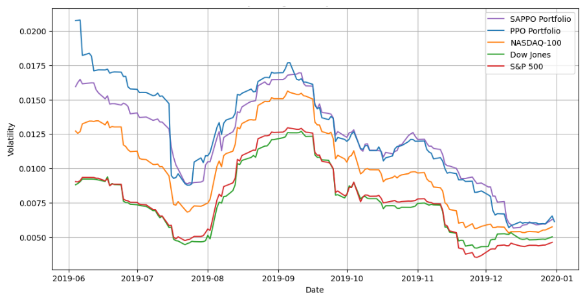

Figure 3: 30-day rolling volatility comparison of the SAPPO and PPO portfolios with the NASDAQ-100, S&P 500, and DJI indices.

Figure 3 presents the 30-day rolling volatility comparison, showing that the SAPPO and PPO portfolios exhibit higher volatility than major benchmark indices such as the NASDAQ-100, S&P 500, and DJI. The SAPPO portfolio demonstrates the highest volatility for most of the observed period, indicating a more aggressive trading strategy that reacts dynamically to market fluctuations. The PPO portfolio follows a similar trend but with slightly lower volatility, suggesting a relatively more balanced risk exposure.

Both SAPPO and PPO portfolios experience pronounced volatility spikes, particularly around mid-2019, aligning with increased market uncertainty. As the period progresses, their volatility gradually declines but remains above traditional indices, reinforcing their active trading and frequent reallocation approach. The NASDAQ-100, S&P 500, and Dow Jones exhibit more stable and lower volatility levels, consistent with their passive investment nature.

The results confirm that the sentiment-aware reinforcement learning strategies adapt quickly to market changes, capturing short-term trends efficiently. However, the higher volatility associated with SAPPO and PPO highlights the tradeoff between increased return potential and short-term risk exposure.

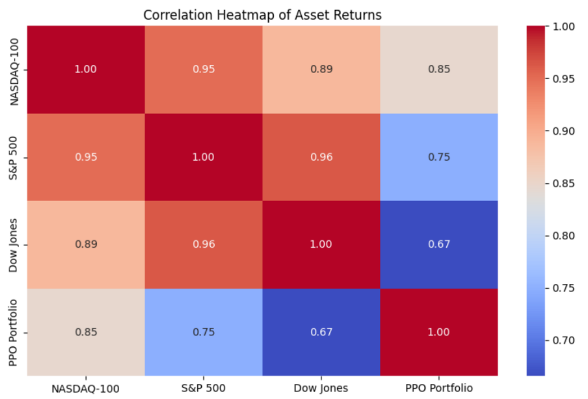

Figure 4: Correlation heatmap of asset returns comparing the PPO portfolio with major indices such as NASDAQ-100, S&P 500, and DJI.

The correlation heatmap (Figure 4) shows that the PPO portfolio maintains a moderate level of independence from major indices, with correlations of 0.67 with the DJI and 0.75 with the S&P 500. This diversification suggests that the PPO agent develops unique portfolio allocation strategies, reducing reliance on broader market movements (Campbell and Viceira, 2002).

The second experiment introduces market sentiment analysis into the PPO framework, forming the SAPPO model. By incorporating sentiment data from Refinitiv financial news sources, processed using LLaMA 3.3 via Hugging Face transformers, the agent receives an additional market signal to guide allocation decisions. This enables sentiment-driven adjustments in response to market sentiment shifts.

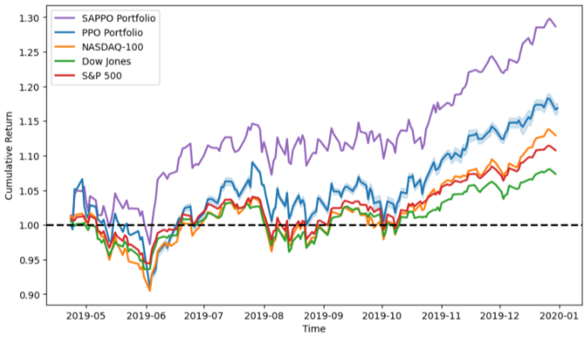

Figure 5: Cumulative return comparison of PPO and SAPPO portfolios against NASDAQ-100, S&P 500, and DJI indices.

The cumulative return comparison (Figure 5) highlights the performance improvement achieved by SAPPO over standard PPO. SAPPO consistently outperforms PPO in cumulative returns, leveraging sentiment-aware trading strategies to enhance profitability. By reacting to shifts in market sentiment, SAPPO is better equipped to capture momentum and avoid adverse market conditions.

Metric	PPO	SAPPO	NASDAQ-100
Sharpe ratio	1.55	1.90	1.25
Annualized return	26.5%	30.2%	21.3%
Max drawdown	-17.5%	-13.8%	-21.9%
Volatility	11.8%	11.2%	14.5%
Turnover rate	35%	37%	N/A

Table 1: Performance comparison of PPO versus SAPPO.

Table 1 presents a quantitative comparison between PPO and SAPPO. The Sharpe ratio of SAPPO (1.90) is higher than that of PPO (1.55), indicating improved risk-adjusted returns. Annualized returns increase from 26.5% (PPO) to 30.2% (SAPPO), demonstrating better profitability. Additionally, SAPPO exhibits a lower maximum drawdown (-13.8%) compared to PPO (-17.5%), suggesting enhanced downside protection.

These results indicate that sentiment-aware reinforcement learning enhances portfolio management by integrating external market sentiment signals. The ability to react to news-driven market sentiment fluctuations provides an additional layer of adaptability beyond price-based decision-making. The findings highlight the potential of combining reinforcement learning with financial sentiment analysis for dynamic investment strategies.

5 Impact

The integration of sentiment-aware reinforcement learning into portfolio optimization has significant implications for both academic research and real-world financial applications. By incorporating financial news sentiment into a deep reinforcement learning framework, our study demonstrates how external market signals can enhance portfolio allocation decisions beyond traditional price-based strategies. The SAPPO model consistently outperforms the vanilla PPO framework by leveraging market sentiment as an additional decision-making factor.

The findings contribute to the broader field of financial reinforcement learning by showcasing the potential of sentiment-aware trading strategies. The results suggest that sentiment-driven reinforcement learning enables agents to react more effectively to market fluctuations, capturing momentum and mitigating losses during adverse conditions. This approach is particularly relevant for institutional investors, hedge funds, and algorithmic trading firms seeking adaptive trading models that dynamically adjust to market sentiment.

Furthermore, our research underscores the growing relevance of multi-modal financial decision-making, where reinforcement learning models integrate structured market data with unstructured textual information to refine investment strategies. The application of LLaMA 3.3 for financial sentiment extraction highlights the increasing role of AI-driven financial analysis. By leveraging real-time news data, the proposed framework bridges the gap between quantitative finance and natural language processing, providing a foundation for future research in sentiment-aware algorithmic trading.

6 Limitations and Future Work

Despite its promising results, this study has several limitations that should be addressed in future research.

First, the scope of sentiment analysis is limited to financial news articles retrieved from Refinitiv and processed using LLaMA 3.3. While this dataset provides high-quality sentiment data, it does not incorporate alternative sources such as social media, earnings calls, or analyst reports, which could further improve sentiment-based decision-making. Future work could explore multi-source sentiment aggregation to enhance the robustness of sentiment signals.

Second, the experiment is conducted using a three-stock portfolio consisting of Google, Microsoft, and Meta, providing a controlled setting but lacking the diversity of a fully diversified investment portfolio. Expanding this research to larger, sector-diverse portfolios could validate the effectiveness of sentiment-aware reinforcement learning across different industries and market conditions.

Third, the backtesting period from January 2013 to January 2020 does not account for real-time market execution, liquidity constraints, or macroeconomic shocks outside this timeframe. Implementing real-time trading simulations or live trading experiments would provide a more realistic evaluation of model performance under actual market conditions, including execution delays and bid-ask spreads.

Finally, the SAPPO framework relies on historical daily news sentiment, meaning sentiment scores are updated at market close and used for allocation adjustments the following trading day. This approach does not capture intra-day sentiment shifts that could impact trading decisions. Future research could explore higher-frequency sentiment updates, integrating real-time news streams into the reinforcement learning pipeline for more responsive market adaptation.

Addressing these limitations will further refine sentiment-aware reinforcement learning models, improving their adaptability, generalization, and scalability in financial markets.

7 Conclusion

This study extends PPO by incorporating a sentiment-aware layer into portfolio optimization. By leveraging LLM-based sentiment analysis, the SAPPO framework is introduced to enhance decision-making by integrating external market sentiment signals.

Experimental results indicate that the sentiment-enhanced model achieves superior risk-adjusted returns, higher Sharpe ratios, and lower drawdowns compared to the standard PPO framework. The SAPPO portfolio consistently outperforms traditional benchmarks such as the NASDAQ-100, S&P 500, and Dow Jones Industrial Average, demonstrating that integrating sentiment analysis into reinforcement learning improves trading strategies.

The findings suggest that market sentiment serves as a valuable supplementary feature in reinforcement learning-based portfolio optimization, providing an additional layer of adaptability to evolving financial conditions. This approach offers an effective alternative to purely price-based reinforcement learning models by incorporating investor sentiment as a key factor in portfolio rebalancing.

Overall, this study highlights the importance of sentiment-aware reinforcement learning in financial decision-making and provides a foundation for future research exploring multi-modal trading models that integrate structured market data with unstructured financial news.

References

Meta AI. 2024. Llama 3.3: A multilingual large language model. Accessed: 2025-02-03.

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

M. Baker and J. Wurgler. 2012. Comovement and predictability relationships between bonds and the cross-section of stocks. *Review of Asset Pricing Studies*, 2(1):57–87.

Mao H. Bollen, J. and X. Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

John Y. Campbell and Luis M. Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press, New York, NY, USA.

Yao Chen, Bryan T. Kelly, and Dacheng Xiu. 2022. Expected returns and large language models. *SSRN Electronic Journal*.

Zhang J. Dai, Z. and C. Li. 2022. Reinforcement learning-based stock trading with sentiment analysis. *Quantitative Finance*, 22(7):1201–1220.

Yue Deng, Fang Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. 2016. Deep direct reinforcement learning for financial signal representation and trading. In *IEEE Transactions on Neural Networks and Learning Systems*, volume 28, pages 653–664, Piscataway, NJ, USA.

Zhang Y. Liu T. Ding, X. and J. Duan. 2015. Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2327–2333.

A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Yang, and A. et al. Fan. 2024. The llama 3 herd of models. *arXiv preprint*.

Hugging Face. 2024. Transformers library for natural language processing. `https://huggingface.co`. Accessed: 2024-02-03.

Eugene F. Fama and James D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636.

A. H. Huang, H. Wang, and Y. Yang. 2023. Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2):806–841.

Robert Jansen. 2000. A statistical framework for backtesting portfolio strategies. *Journal of Financial Analysis*, 56(4):67–82.

Zhang J. Jin, S. and L. Wang. 2023. Deep reinforcement learning for stock portfolio optimization with market sentiment. *Expert Systems with Applications*, 213:118971.

Zhuo T. Ke, Bryan T. Kelly, and Dacheng Xiu. 2019. Predicting returns with text data. Technical report, National Bureau of Economic Research.

K. Kirtac and G. Germano. 2024. Sentiment trading with large language models. *Finance Research Letters*, 62:105227.

Chen P. Liu, B. and N. Zhu. 2020. Hybrid deep learning model for stock price prediction. *Applied Intelligence*, 50(10):3452–3464.

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

Harry Markowitz. 1952. Portfolio selection. *Journal of Finance*, 7(1):77–91.

Refinitiv. 2024. Refinitiv financial data and analytics. Accessed: 2024-02-05.

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

William F. Sharpe. 1994. The sharpe ratio. *Journal of Portfolio Management*, 21(1):49–58.

L. A. Smales. 2014. News sentiment and bank stock returns. *European Journal of Finance*, 20(11):925–938.

Saurabh Sood, Konstantinos Papasotiriou, Matas Vaiciulis, and Tucker Balch. 2023. Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization. In *Proceedings of the 33rd International Conference on Automated Planning and Scheduling (ICAPS 2023), FinPlan Workshop*.

Frank A. Sortino and Robert Van Der Meer. 1994. Downside risk. *Journal of Portfolio Management*, 20(2):27–31.

Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA.

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

Jack L. Treynor. 1966. How to rate management of investment funds. *Harvard Business Review*, 44(1):63–75.

Yukun Wang, Xiaojun Jin, Haijun Guo, and Haomiao Xu. 2019. Deep reinforcement learning for portfolio optimization. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 2190–2193, New York, NY, USA.

Jiaqi Ye, Sheng Zhang, Jingtao Hao, and Hongtao Wang. 2020. Reinforcement-learning-based portfolio management with augmented asset movement prediction states. *Expert Systems with Applications*, 159:113594.

Junni L. Zhang, Wolfgang Karl Härdle, Cathy Y. Chen, and Elisabeth Bommes. 2020. Distillation of news flow into analysis of stock reactions. *arXiv preprint arXiv:2009.10392*.