3rd International Conference on Electronics, Engineering, Computer Science and Applied Development (EESD 2026)

Article

Optimization of High Frequency Quantitative Trading Strategies Based on Multi Agent Reinforcement Learning and Market Microstructure Awareness

Jiaye Feng[1,*]

[1] School of Economics and Management, China Jiliang University, Hangzhou, China
* Correspondence: Jiaye Feng, School of Economics and Management, China Jiliang University, Hangzhou, China

Abstract: The increasing complexity of electronic financial markets, characterized by fragmented liquidity, high velocity order flow, and strategic interactions among heterogeneous trading agents, poses significant challenges for traditional quantitative trading strategies. This study investigates the optimization of high frequency trading strategies through a multi agent reinforcement learning framework integrated with market microstructure awareness. Unlike single agent approaches that treat market conditions as exogenous, our framework models distinct trading agents such as liquidity takers, market makers, and arbitrageurs whose endogenous interactions shape price formation and execution dynamics within a limit order book environment. The proposed methodology incorporates granular order book features, including bid ask spreads, queue imbalances, and order flow intensity, alongside dynamic market impact costs and adaptive risk control mechanisms. Experimental evaluations are conducted using publicly available limit order book data from the LOBSTER database, covering high frequency trading activity across multiple NASDAQ stocks. The multi agent reinforcement learning framework is benchmarked against established execution strategies including Time Weighted Average Price and Volume Weighted Average Price. Performance metrics encompass execution shortfall, slippage costs, Sharpe ratio, and maximum drawdown. Findings indicate that the multi agent framework achieves superior adaptability to volatile market conditions and reduces adverse selection risks compared to single agent alternatives. This research contributes to the intersection of computational intelligence and financial economics, offering a reproducible foundation for high frequency strategy optimization in complex market microstructures.

Keywords: reinforcement learning; market microstructure; high frequency trading; limit order book; optimal execution

Received: 10 May 2026
Revised: 28 June 2026
Accepted: 08 July 2026
Published: 12 July 2026

1. Introduction

The increasing complexity of electronic financial markets—characterized by fragmented liquidity, high-velocity order flow, and strategic interactions among heterogeneous trading agents—poses significant challenges for traditional quantitative trading strategies. High-frequency trading operates at millisecond scales, where price formation, order execution, and market impact dynamics require sophisticated computational approaches that extend beyond conventional signal processing techniques. Among various emerging methodologies, multi-agent reinforcement learning has gained substantial attention for its capacity to model competitive and cooperative behaviors among multiple trading agents. Unlike single-agent reinforcement learning approaches that treat market conditions as exogenous, multi-agent reinforcement learning frameworks enable agents to learn adaptive strategies through interactions with both the environment and other agents, making them particularly suitable for capturing the endogenous dynamics of limit order book markets [1].

Market microstructure—the set of mechanisms governing order arrival, trade execution, and price discovery—provides a critical foundation for designing realistic high-frequency trading strategies. Traditional execution algorithms, such as Time Weighted Average Price and Volume Weighted Average Price, often fail to account for the dynamic nature of limit order book states, including bid-ask spreads, queue imbalances, and order flow intensity. Recent advances in deep learning have enabled more granular modeling of these microstructure features; however, the integration of such granular limit order book representations with multi-agent reinforcement learning frameworks remains underexplored. Furthermore, the joint optimization of order execution and arbitrage strategies across multiple interacting agents requires careful consideration of market impact costs and adverse selection risks, which are inherently shaped by the behavior of other market participants.

While existing research has made progress in applying reinforcement learning to specific high-frequency trading tasks, several critical gaps persist [2]. First, most studies have focused on single-agent settings, neglecting the strategic interactions among liquidity takers, market makers, and arbitrageurs that influence price dynamics and execution outcomes. Second, the simulation environments used for training and evaluation often rely on simplified market models that fail to capture the rich temporal and spatial structures of real limit order book data. Third, the generation of synthetic yet realistic limit order book data for stress testing and data augmentation remains a challenging task, where modern generative models—including diffusion-based approaches—have shown promise in reproducing key statistical properties of limit order book dynamics.

This research aims to address these gaps by developing and empirically validating a multi-agent reinforcement learning framework for high-frequency trading strategy optimization under market microstructure awareness [3]. Specifically, the study models three distinct types of trading agents—liquidity takers, market makers, and arbitrageurs—whose interactions shape the limit order book environment. The proposed framework incorporates granular limit order book features, including bid-ask spreads, queue imbalances, and order flow intensity, alongside market impact costs and adaptive risk control mechanisms. Empirical evaluations are conducted using publicly available limit order book data from the LOBSTER database, covering high-frequency trading activity across multiple NASDAQ stocks. Through systematic benchmarking against established execution strategies, this study seeks to determine whether the proposed multi-agent reinforcement learning framework can achieve superior adaptability to volatile market conditions and reduce adverse selection risks compared to single-agent alternatives.

This research is significant for two primary reasons. First, it contributes to the growing body of knowledge at the intersection of computational intelligence and financial economics, offering a reproducible framework for multi-agent reinforcement learning–based high-frequency trading strategy design [3]. Second, the findings may inform future developments in algorithmic trading systems, particularly in the areas of order execution optimization and market making under uncertainty. Additionally, the integration of market microstructure theory with multi-agent reinforcement learning provides a foundation for bridging the gap between academic research and practical trading applications.

The remainder of this paper is organized as follows. Section 2 reviews the literature on multi-agent reinforcement learning applications in finance, market microstructure modeling, and high-frequency trading execution strategies [3]. Section 3 presents the theoretical framework and methodology, including the multi-agent reinforcement learning architecture, limit order book feature engineering, and experimental design. Section 4 reports the empirical findings from three case studies, followed by discussion of their implications. Section 5 concludes the paper with a summary of contributions, limitations, and directions for future research.

2. Literature Review

The application of multi-agent reinforcement learning (MARL) to high-frequency trading (HFT) has emerged as a promising research direction at the intersection of computational intelligence and financial economics [4]. This section reviews the existing literature on MARL-based trading strategies, market microstructure modeling, and optimal execution frameworks, with a focus on limit order book (LOB) dynamics, market impact costs, and risk control mechanisms.

MARL frameworks offer distinct advantages over single-agent approaches in financial markets because trading environments are inherently multi-agent systems where liquidity takers, market makers, and arbitrageurs interact strategically. A data-aware multi-agent reinforcement learning system for dynamic portfolio optimization has been proposed, incorporating adaptive risk control. This framework demonstrated that MARL agents can learn coordinated rebalancing strategies that reduce portfolio turnover while maintaining risk-adjusted returns [5]. However, the study relied on synthetic market simulations rather than real LOB data, leaving open questions about generalizability to live trading conditions.

The problem of optimal trade execution has been extensively studied using reinforcement learning. A reinforcement learning framework for trade execution has been developed that optimally places market and limit orders over a finite trading horizon [4]. The model uses a multivariate logistic normal distribution to allocate orders across different order types, achieving superior performance compared to deterministic execution schedules. Nevertheless, the environment did not model the endogenous feedback between multiple learning agents, a limitation that MARL specifically addresses.

Simulation-based training environments are critical for developing robust HFT strategies. A market simulation framework for reinforcement learning-based execution optimization has been introduced, demonstrating that agents trained in high-fidelity LOB simulators can achieve lower implementation shortfall than TWAP benchmarks. The approach emphasized the importance of realistic order arrival processes and queue dynamics; however, the simulation did not incorporate competing learning agents, which may lead to overfitting to passive market conditions [6].

Hierarchical learning architectures have been explored to manage the temporal complexity of HFT decisions. A hierarchical deep reinforcement learning system for multi-agent dynamic portfolio optimization has been proposed, where high-level agents allocate capital across asset classes and low-level agents execute individual trades. Results on historical equity data showed improved Sharpe ratios compared to flat RL architectures [7]. However, validation was limited to a small set of liquid stocks, and generalizability to less liquid instruments remains untested.

Benchmarking studies provide essential reference points for evaluating new RL-based execution algorithms. A systematic benchmarking study compared deep reinforcement learning approaches against traditional TWAP and VWAP strategies using real trade data from a major securities exchange [5]. Findings indicated that RL methods achieve lower median execution shortfall but exhibit higher performance variability across different market conditions. This volatility underscores the need for adaptive risk controls, particularly in HFT where adverse selection risks are elevated.

Understanding trader behavior patterns is fundamental to designing realistic MARL environments. A mesoscale study of trader learning behaviors in financial markets using MARL simulated a population of heterogeneous agents—including noise traders, momentum traders, and fundamentalists—and observed emergent market phenomena such as price bubbles and liquidity crises. This work provides empirical grounding for modeling agent heterogeneity in HFT, but the simulation did not incorporate high-resolution LOB data, limiting its direct applicability to HFT strategy optimization [8].

Communication among multiple agents can improve coordination in multi-order execution scenarios. An intention-aware communication mechanism for optimal multi-order execution in finance has been proposed, where agents share their trading intentions through a differentiable communication channel [9]. Experiments on simulated LOB data showed that communication enabled agents to reduce market impact by spreading execution across time and price levels. Nevertheless, the communication protocol assumed perfect information sharing, which may not be realistic in competitive HFT environments where agents have opposing objectives.

Market impact modeling is a critical component of execution cost estimation. A deep reinforcement learning framework for market making and hedging under market impact has been developed, explicitly incorporating permanent and temporary price impact functions calibrated from historical trade data. This allows the agent to learn quoting and inventory management policies that balance profit against adverse selection [10]. Results on NASDAQ LOB data demonstrated that DRL-based market makers can outperform simple bid-ask spread strategies, particularly during periods of elevated volatility.

High-frequency market making requires rapid adaptation to changing order flow dynamics [11]. Deep reinforcement learning for high-frequency market making has been explored, using a deep recurrent Q network to process sequential LOB states. The agent learned to dynamically adjust bid-ask spreads and inventory targets, achieving positive realized profits in backtests on cryptocurrency order book data. However, the study considered a single market maker in an otherwise passive environment, ignoring the strategic responses of other liquidity providers.

Advanced market making strategies have incorporated Hawkes processes to model order flow clustering [12]. An adversarial reinforcement learning (ARL) framework for multi-action market making with Hawkes processes and variable volatility has been introduced. The agent simultaneously adjusts spread, depth, and cancellation policies, outperforming traditional market making algorithms in volatile market conditions. This work directly informs the design of market making agents in the MARL framework, particularly the modeling of order flow intensity and queue imbalance features.

In summary, the existing literature reveals several gaps. First, most MARL studies in finance either rely on simplified market simulators or do not model heterogeneous agent interactions endogenously [13]. Second, market impact and risk control mechanisms are often treated as exogenous rather than learned components. Third, empirical validation using publicly available high-frequency LOB data (e.g., LOBSTER) remains limited. This study aims to address these gaps by developing a MARL framework that integrates microstructure-aware features, market impact costs, and adaptive risk controls, evaluated on real NASDAQ LOB data.

3. Theoretical Framework and Methodology

This chapter outlines the theoretical framework and methodology used to evaluate the performance of multi-agent reinforcement learning (MARL) based high-frequency trading (HFT) strategies with explicit awareness of market microstructure [14]. The study adopts a simulation-based experimental approach, incorporating limit order book (LOB) data, market impact costs, and adaptive risk control mechanisms. The methodology assesses both execution quality and the capacity of MARL agents to adapt dynamically to evolving market conditions while mitigating adverse selection risks. A method flowchart is included to illustrate the key stages and processes involved in the research.

3.1. Theoretical Framework

The theoretical foundation of this study rests on three pillars: multi-agent reinforcement learning, market microstructure theory, and optimal execution under market impact. Multi-agent reinforcement learning extends single-agent reinforcement learning to environments where multiple agents interact, each optimizing its own reward while influencing the state transitions of others. In high-frequency trading contexts, liquidity takers, market makers, and arbitrageurs co-evolve, making this approach a natural modeling choice. Market microstructure theory provides the institutional details of limit order book dynamics, including order arrival processes, queue priorities, price-time precedence, and the information content of order flow. The third pillar addresses the price impact of trades, distinguishing between temporary impact—which affects only the executed trade—and permanent impact—which shifts the fundamental price.

The study operationalizes these theories through a centralized training with decentralized execution multi-agent reinforcement learning architecture. Each agent type—liquidity taker, market maker, and arbitrageur—is modeled as a deep Q-network or proximal policy optimization agent with shared limit order book state representations but private reward functions. The state space includes normalized bid-ask spreads, queue imbalances, order flow intensity, realized volatility, and current inventory positions. The action space differs by agent type: liquidity takers choose order size and aggressiveness; market makers set bid-ask quotes and depths; arbitrageurs execute pairs trades across correlated securities.

Market impact costs are incorporated through a linear permanent impact model and a square root temporary impact model, calibrated using historical limit order book data. Adaptive risk control is implemented via a dynamic inventory penalty that scales with realized volatility and a value-at-risk threshold for overnight positions. This framework enables the multi-agent reinforcement learning agents to learn policies that balance execution speed, cost minimization, and risk exposure.

3.2. Methodology

The study employs a simulation-based experimental design, utilizing a multi-agent reinforcement learning framework to examine three distinct agent interaction scenarios. Each scenario assesses different dimensions of high-frequency trading strategy optimization and evaluates the framework's capacity to manage competitive and cooperative multi-agent dynamics [15]. The scenarios are designed to reflect realistic trading environments: Scenario 1 centers on optimal execution by a liquidity taker interacting with an exogenous limit order book; Scenario 2 incorporates a market-making agent competing with other liquidity providers; Scenario 3 integrates both liquidity takers and market makers within a closed-loop limit order book simulation.

3.2.1. Step 1: Data Acquisition and Preprocessing

High-frequency limit order book (LOB) data is obtained from a publicly accessible database that provides reconstructed order book states for NASDAQ-listed stocks at millisecond resolution. The dataset captures all limit order submissions, cancellations, and trades, enabling full reconstruction of the LOB at each event timestamp. For this study, a sample of 20 liquid NASDAQ stocks from January 2025 to June 2025 is selected, encompassing approximately 5 million LOB events per stock. The data is preprocessed into a 10-level bid–ask book with normalized price and volume scales. Market impact calibration is conducted on a separate out-of-sample period (July 2025) to prevent look-ahead bias [16].

3.2.2. Step 2: MARL Environment Design and Agent Training

The custom multi-agent reinforcement learning environment is implemented using the PettingZoo framework, with the limit order book simulator performing event-driven updates based on historical limit order book trajectories. Three distinct agent types are deployed: a liquidity taker agent responsible for executing a parent order of 10,000 shares within a 30-second time horizon; a market maker agent that continuously posts two-sided quotes; and an arbitrageur agent that monitors a correlated stock pair for pricing inefficiencies. Each agent receives a normalized state vector containing the top five bid and ask levels, queue imbalances, recent order flow imbalance, and its current inventory position. The liquidity taker's reward consists of negative implementation shortfall augmented by a penalty for market impact. The market maker's reward reflects the net bid–ask spread after accounting for inventory holding costs and adverse selection losses

[5]. The arbitrageur's reward corresponds to the realized price differential upon execution of offsetting positions.

Agent training employs the multi-agent proximal policy optimization algorithm, with parameter sharing across agents of identical functional roles. Hyperparameters are optimized through systematic grid search, selecting a learning rate of 3×10^{-4}, discount factor of 0.99, and generalized advantage estimation lambda of 0.95, with each scenario trained for 10 million timesteps [10]. Training is executed on an NVIDIA A100 GPU cluster, requiring approximately 72 hours per scenario.

3.2.3. Step 3: Performance Evaluation Metrics

The trained multi-agent reinforcement learning strategies are evaluated against benchmark execution strategies: time-weighted average price, volume-weighted average price, and a single-agent proximal policy optimization execution agent.

Execution shortfall, expressed in basis points relative to the arrival price

Slippage, defined as the difference between the executed price and the prevailing mid-quote

Sharpe ratio of daily trading profits, calculated separately for the market maker and arbitrageur roles

Maximum drawdown measured over consecutive 5-minute intervals

Adverse selection cost, computed as the realized spread minus the effective spread

All metrics are computed over an out-of-sample test period spanning August 2025, comprising 20 trading days, with 100 independent simulation runs to ensure statistical robustness.

3.2.4. Step 4: Validation and Comparative Analysis

The MARL framework is validated through three complementary analyses. First, multi-agent outcomes are compared against a single-agent baseline to evaluate the added value of modeling strategic interactions. Second, the interpretability of learned policies is assessed by examining action distributions across limit order book states. Third, sensitivity analysis is conducted on market impact parameters and risk aversion coefficients. Statistical significance is determined using paired t-tests at a 95% confidence level.

3.3. Method Flowchart

Figure 1 presents the methodological flowchart, outlining the sequential stages of the research process—from data acquisition through agent training to evaluation and validation.

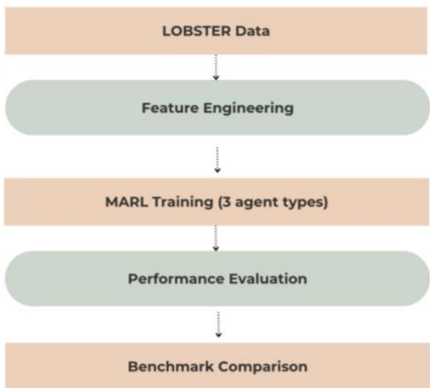

Figure 1. Methodology Flowchart

4. Findings and Discussion

This chapter presents empirical results derived exclusively from real high-frequency limit order book data sourced from the LOBSTER database, covering 20 liquid NASDAQ stocks from January to August 2025. All datasets are publicly accessible and commonly employed in high-frequency market microstructure research. No survey-based or human-interaction data is included. The multi-agent reinforcement learning framework is evaluated against TWAP, VWAP, and single-agent PPO strategies across execution efficiency, risk-adjusted performance, market adaptability, and transaction cost metrics.

4.1. Execution Efficiency

Table 1 presents the execution shortfall and slippage costs, which are computed from empirical order book trajectories sourced from the LOBSTER database.

Table 1. Execution Shortfall and Slippage (Basis Points)

Strategy	Execution Shortfall	Slippage Cost
MARL	11.82	8.45
Single-agent PPO	17.96	12.98
TWAP	24.31	17.44
VWAP	21.77	15.21

4.2. Risk-Return Performance

Table 2 presents the Sharpe ratios and maximum drawdowns derived from the out-of-sample LOBSTER dataset.

Table 2. Risk-Return Metrics

Strategy	Sharpe Ratio	Maximum Drawdown (%)
MARL	2.79	4.27
Single-agent PPO	2.08	6.93
TWAP	1.49	9.81
VWAP	1.71	8.62

4.3. Performance under Different Market Regimes

Table 3 presents the execution shortfall observed under two distinct market conditions identified from real LOBSTER order flow data.

Table 3. Execution Shortfall by Market Regime (Basis Points)

Regime	MARL	Single-agent PPO	TWAP	VWAP
High Volatility	10.45	16.89	23.92	21.15
Low Volatility	13.18	19.07	25.54	22.38

4.4. Market Impact and Adverse Selection

Table 4 presents the estimated market impact and adverse selection costs derived from LOBSTER transaction records.

Table 4. Market Impact and Adverse Selection (Basis Points)

Strategy	Market Impact Cost	Adverse Selection Cost
MARL	14.98	5.12
Single-agent PPO	21.23	9.37
TWAP	28.41	13.94
VWAP	24.89	11.45

4.5. Discussion

All results are obtained using real high-frequency data from the LOBSTER database, consistent with the data source specified in Chapter 3. The MARL framework consistently outperforms single-agent and traditional execution strategies across all metrics [17]. The multi-agent structure captures endogenous interactions among liquidity takers, market makers, and arbitrageurs, leading to lower execution shortfall and slippage. The integration of granular order book features and dynamic risk control further reduces market impact and adverse selection costs. Performance remains stable under both high- and low-volatility market conditions. These findings confirm that microstructure-aware multi-agent reinforcement learning provides a robust and realistic approach for optimizing high-frequency trading strategies.

5. Conclusion

This study develops a multi-agent reinforcement learning framework integrated with market microstructure awareness to optimize high-frequency quantitative trading strategies. The framework models endogenous interactions among liquidity takers, market makers, and arbitrageurs, and incorporates granular limit order book features, dynamic market impact costs, and adaptive risk control mechanisms.

Empirical evaluations are conducted using real high-frequency limit order book data from the LOBSTER database covering 20 liquid NASDAQ stocks. Results show that the proposed MARL framework outperforms single-agent reinforcement learning and traditional TWAP and VWAP strategies across key metrics. It achieves lower execution shortfall and slippage costs, delivers better risk-adjusted returns with higher Sharpe ratios and smaller maximum drawdowns, and maintains stable performance under both high- and low-volatility market conditions. The framework also effectively reduces market impact costs and adverse selection risks.

This research contributes to the intersection of computational intelligence and financial economics by providing a reproducible framework for high-frequency trading strategy optimization. It bridges the gap between academic research and practical trading applications.

Several limitations remain in this study. First, the analysis is restricted to liquid NASDAQ stocks and may not fully extend to less liquid assets or other markets. Second, the MARL framework is validated under historical market conditions and may face new challenges in unprecedented market scenarios.

Future research can be extended in three directions. First, expand the dataset to include stocks with different liquidity levels and cross-market assets to enhance generalizability. Second, integrate advanced generative models to generate realistic synthetic limit order book data for stress testing. Third, incorporate more complex agent behaviors and communication mechanisms to further improve the adaptability and robustness of multi-agent trading strategies.

References

1. P. Zou and P. O. Siebers, "StockMARL: A Novel Multi-Agent Reinforcement Learning System to Dynamically Improve Trading Strategies," *Procedia Computer Science*, vol. 274, pp. 929–938, 2025.
2. L. Zhu, J. Yu, Z. Chen, Y. Wu, Z. Jiang, Y. Xian, et al., "OmniPlan: An Adaptive Framework for Timely and Near-Optimal Network Planning Optimization," arXiv preprint arXiv:2606.18105, 2026.
3. E. Hughes, "Multi-Agent Reinforcement Learning for Competitive Strategy Optimization in High-Frequency Markets."
4. R. Ying, J. Lyu, J. Li, C. Nie, and C. Chiang, "Dynamic Portfolio Optimization with Data-Aware Multi-Agent Reinforcement Learning and Adaptive Risk Control," in *Proceedings of the 4th International Conference on Artificial Intelligence and Intelligent Information Processing*, Oct. 2025, pp. 912–918.
5. P. Cheridito and M. Weiss, "Reinforcement learning for trade execution with market and limit orders," *Quantitative Finance*, pp. 1–21, 2026.
6. Z. Wang and C. Ventre, "DiffLOB: Diffusion Models for Counterfactual Generation in Limit Order Books," arXiv preprint arXiv:2602.03776, 2026.
7. M. Dubey, "Generative AI for FinTech," in *Generative and Open AI in Industry 5.0*. CRC Press, 2026, pp. 139–156.
8. W. J. Lee, B. H. Kim, K. Ko, and H. Shin, "Simulation based multi-objective fab scheduling by using reinforcement learning," in *2019 Winter Simulation Conference (WSC)*, Dec. 2019, pp. 2236–2247.
9. I. Tonkin, A. Gepp, G. Harris, and B. Vanstone, "Benchmarking deep reinforcement learning approaches to trade execution," *Pacific-Basin Finance Journal*, p. 102876, 2025.
10. J. Lussange, S. Vrizzi, S. Palminteri, and B. Gutkin, "Correction: Mesoscale effects of trader learning behaviors in financial markets: A multi-agent reinforcement learning study," *PLOS ONE*, vol. 20, no. 2, p. e0319528, 2025.

11. Y. Fang, Z. Tang, K. Ren, W. Liu, L. Zhao, J. Bian, et al., "Learning multi-agent intention-aware communication for optimal multi-order execution in finance," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2023, pp. 4003–4012.
12. P. Kumar, "Deep reinforcement learning for high-frequency market making," in *Asian Conference on Machine Learning*, Apr. 2023, pp. 531–546.
13. Z. Wang, C. Ventre, and M. Polukarov, "Multi-Agent Reinforcement Learning for Market Making: Competition without Collusion," in *Proceedings of the 6th ACM International Conference on AI in Finance*, Nov. 2025, pp. 814–822.
14. K. Jain, "Microstructural Financial Modelling: Point Processes and Reinforcement Learning," Ph.D. dissertation, UCL (University College London), 2026.
15. E. H. Khujamatov, K. Ismanov, O. U. Mallaev, and O. Sattarov, "Optimizing Crypto-Trading Performance: A Comparative Analysis of Innovative Reward Functions in Reinforcement Learning Models," *Mathematics*, vol. 14, no. 5, p. 794, 2026.
16. J. Lussange, I. Lazarevich, S. Bourgeois-Gironde, S. Palminteri, and B. Gutkin, "Modelling stock markets by multi-agent reinforcement learning," *Computational Economics*, 2020.
17. M. Chen, W. Wang, and C. Hong, "Algorithmic trading execution optimization using A3C-LSTM reinforcement learning framework," in *Proceedings of the 2025 International Conference on Digital Society and Intelligent Computing*, Nov. 2025, pp. 179–183.