Table 3: Ablation analysis on three conference. ✓denote the inclusion of components.

MAM	SDM	RAM	ARR(%)	TR(%)	SR	CR	SoR	MDD(%)	VoL(%)	ENT	ENB
✓			40.89	179.66	1.81	5.27	51.23	28.61	1.66	1.73	1.22
	✓		43.25	193.97	1.93	6.27	53.41	24.99	1.63	1.89	1.33
		✓	35.53	148.93	1.88	5.94	52.01	21.73	1.34	1.65	1.19
	✓	✓	48.59	228.07	2.79	8.54	58.85	19.21	1.38	2.37	1.39
✓		✓	46.42	213.94	2.51	7.82	55.21	20.52	1.28	2.18	1.35
✓	✓		52.71	256.12	2.86	9.14	61.46	21.85	1.33	2.55	1.43
✓	✓	✓	**58.68**	**299.55**	**3.11**	**11.38**	**66.94**	**16.86**	**1.23**	**2.97**	**1.49**

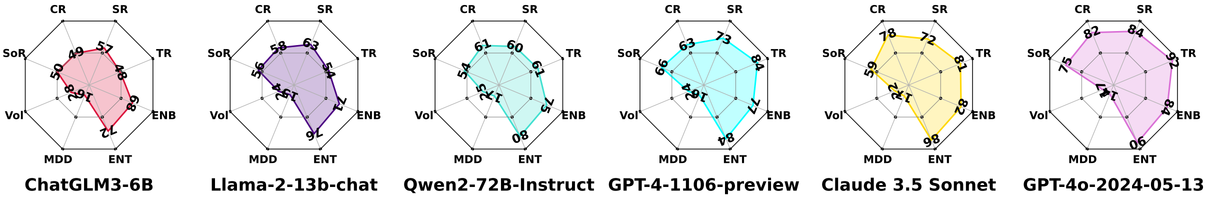


Figure 5: Ablation analysis on several LLM backbones, from open-source to closed-source models. The numbers presented in the figure have been normalized and converted into percentage values.

large-scale language models in financial decision-making. FinAgent achieves an impressive ARR of 45.31% and SR of 2.25, significantly outperforming the top RL-based model, AlphaMix+ (ARR: 32.51%, SR: 1.49). The cumulative returns graph Figure 4 clearly illustrates the superior performance trajectory of LLM-based methods, particularly from mid-2022 onwards.

3) QuantAgents, our proposed multi-agent system, demonstrates superior performance across all evaluation metrics, achieving the highest ARR (58.68%) and SR (3.11), surpassing the best baseline (HedgeAgents) by 19.15% and 30.02%. QuantAgents also excels in risk management with the lowest MDD (16.86%) and VoL (1.43%), while achieving the highest portfolio diversity with ENT (2.97) and ENB (1.49), indicating a more balanced and robust investment strategy. Figure 4 vividly showcases QuantAgents' outstanding performance, with its cumulative returns curve consistently above all other methods, reaching approximately 300% by the end of the test period. This remarkable performance can be attributed to the synergistic collaboration of specialized agents within QuantAgents, each contributing unique expertise in investment management, strategy development, risk control, and market analysis.

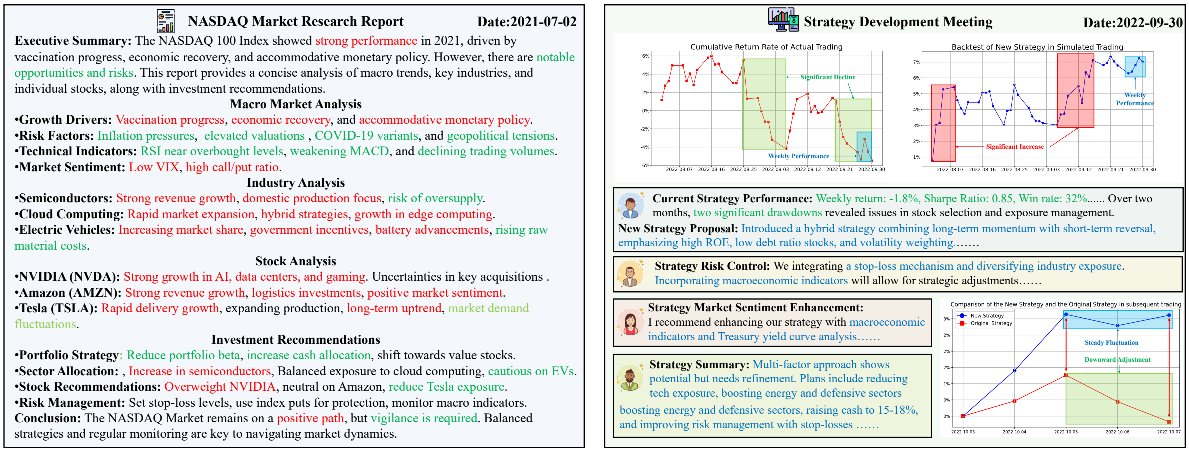
4.5 Ablation Study

4.5.1 Effectiveness of Each Meeting

We conducted an ablation study to evaluate the contribution of each meeting module in QuantAgents. Table 3 presents the performance metrics for different combinations of Market Analysis Meeting (MAM), Strategy Development Meeting (SDM), and Risk Assessment Meeting (RAM). We have the following observations: 1) MAM significantly enhances profitability and portfolio diversity, as evidenced by its high ARR (40.89%) and ENT (1.73) when used alone. Its exclusion from two-meeting combinations reduces ARR and ENT, highlighting its critical role in trend identification and diversification. 2) SDM enhances risk-adjusted returns and portfolio efficiency, with the highest single-meeting SR (1.93) and CR (6.27). Its inclusion in two-meeting setups consistently improves these metrics. The SDM-MAM combination achieves the highest two-meeting SR (2.86) and CR (9.14), demonstrating SDM's effectiveness in strategy formulation. 3) RAM demonstrates strength in risk management and volatility reduction. Despite having the lowest standalone ARR (35.53%), it achieves the lowest single-meeting MDD (21.73%) and VoL (1.34%). In two-meeting configurations,RAM improves risk metrics, notably reducing MDD when combined with SDM. 4) The synergistic effect of all three meetings is evident in our QuantAgents, which outperforms all partial combinations across all metrics. It achieves the highest ARR (58.68%), SR (3.11), and ENT (2.97), while maintaining the lowest MDD (16.86%) and VoL (1.23%).

4.5.2 Effectiveness of LLM Backbone

To evaluate the performance of different LLMs as the backbone, we selected 6 representative



NASDAQ Market Research Report **Date:2021-07-02**

Executive Summary: The NASDAQ 100 Index showed strong performance in 2021, driven by vaccination progress, economic recovery, and accommodative monetary policy. However, there are notable opportunities and risks. This report provides a concise analysis of macro trends, key industries, and individual stocks, along with investment recommendations.

Macro Market Analysis

•**Growth Drivers:** Vaccination progress, economic recovery, and accommodative monetary policy.
•**Risk Factors:** Inflation pressures, elevated valuations , COVID-19 variants, and geopolitical tensions.
•**Technical Indicators:** RSI near overbought levels, weakening MACD, and declining trading volumes.
•**Market Sentiment:** Low VIX, high call/put ratio.

Industry Analysis

•**Semiconductors:** Strong revenue growth, domestic production focus, risk of oversupply.
•**Cloud Computing:** Rapid market expansion, hybrid strategies, growth in edge computing.
•**Electric Vehicles:** Increasing market share, government incentives, battery advancements, rising raw material costs.

Stock Analysis

•**NVIDIA (NVDA):** Strong growth in AI, data centers, and gaming. Uncertainties in key acquisitions .
•**Amazon (AMZN):** Strong revenue growth, logistics investments, positive market sentiment.
•**Tesla (TSLA):** Rapid delivery growth, expanding production, long-term uptrend, market demand fluctuations.

Investment Recommendations

•**Portfolio Strategy**: Reduce portfolio beta, increase cash allocation, shift towards value stocks.
•**Sector Allocation:** , Increase in semiconductors, Balanced exposure to cloud computing, cautious on EVs.
•**Stock Recommendations:** Overweight NVIDIA, neutral on Amazon, reduce Tesla exposure.
•**Risk Management:** Set stop-loss levels, use index puts for protection, monitor macro indicators.
Conclusion: The NASDAQ Market remains on a positive path, but vigilance is required. Balanced strategies and regular monitoring are key to navigating market dynamics.

Strategy Development Meeting **Date:2022-09-30**

Current Strategy Performance: Weekly return: -1.8%, Sharpe Ratio: 0.85, Win rate: 32%...... Over two months, two significant drawdowns revealed issues in stock selection and exposure management.
New Strategy Proposal: Introduced a hybrid strategy combining long-term momentum with short-term reversal, emphasizing high ROE, low debt ratio stocks, and volatility weighting……

Strategy Risk Control: We integrating a stop-loss mechanism and diversifying industry exposure. Incorporating macroeconomic indicators will allow for strategic adjustments……

Strategy Market Sentiment Enhancement:
I recommend enhancing our strategy with macroeconomic indicators and Treasury yield curve analysis……

Strategy Summary: Multi-factor approach shows potential but needs refinement. Plans include reducing tech exposure, boosting energy and defensive sectors boosting energy and defensive sectors, raising cash to 15-18%, and improving risk management with stop-losses ……

(a)

(b)

Figure 6: Visualizations of Market Research Report and Strategy Development Meeting.

models, including ChatGLM3-6B(GLM et al., 2024), Llama-2-13b-chat(Touvron et al., 2023), Qwen2-72B-Instruct(Yang et al., 2024), GPT-4-1106-preview(OpenAI et al., 2024), Claude 3.5 Sonnet(Anthropic, 2024), and GPT-4o-2024-05-13(Wu et al., 2024). Each of these models serves as the brain of QuantAgents, as shown in Figure 5. We have the following observations: 1) QuantAgents achieves consistent performance across diverse LLM backbones, showcasing its adaptability through a robust multi-agent architecture. 2) Larger models, such as Qwen2-72B-Instruct, outperform smaller ones like Llama-2-13b-chat, likely due to their superior capacity to handle complex financial data and detect subtle market patterns. 3) Closed-source models like Claude 3.5 Sonnet outperform open-source ones, such as Qwen2-72B-Instruct, likely due to proprietary training data and advanced fine-tuning techniques enhancing their understanding of financial contexts. Therefore, we select GPT-4o as the core of QuantAgents. Notably, our system has accumulated a total cost of $180 over the three years, averaging only $0.17 per day!

4.6 Visualization

4.6.1 Market Analysis Meeting

Figure 6 (a) showcases the visualization of market research report dated 2021-07-02, generated after the market analysis meeting. The report encapsulates key market insights in a structured layout, progressing from an executive summary highlighting the NASDAQ 100 Index's strong 2021 performance to specific investment recommendations.

4.6.2 Strategy Development Meeting

Figure 6 (b) presents the visualization of a strategy development meeting held on 2022-09-30. The upper left shows the current strategy's performance, marked by a -1.8% weekly return and significant drawdowns due to issues in stock selection and exposure management. The upper right shows a six-month backtest of the new strategy, yielding 35.3% return and a 1.85 Sharpe Ratio. Dave suggested mitigating risks through stop-loss mechanisms and diversified industry exposure. Emily recommended incorporating macroeconomic indicators and adjusting positions in overvalued tech stocks. Otto summarized the strategy, emphasizing risk management and dynamic adjustments.

4.7 Real-World Investment Performance

We evaluated its live trading performance in the A-stock and HK-stock markets from Q3 2024 to Q1 2025. Figure 11 shows cumulative returns, QuantAgents delivered superior returns of 111.87% (Sharpe Ratio: 2.02, Win Rate: 61.23%) in A-stocks and 97.69% (Sharpe Ratio: 1.76, Win Rate: 59.71%) in HK-stocks, highlighting exceptional profitability and risk management across diverse market conditions.

5 Conclusions

In this paper, we present a sophisticated multi-agent financial system, QuantAgents that incorporates simulated trading, configured similarly to that of human quant traders. Furthermore, our system encourages agents to receive feedback on their performance in the real market and their predictive accuracy in simulated trading. Compared to

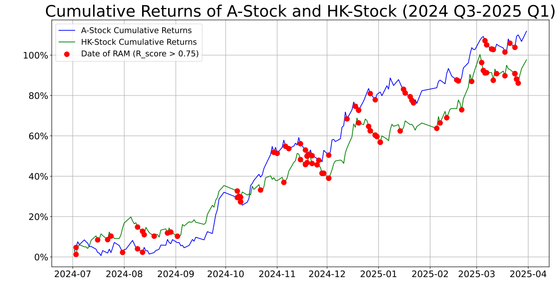


Figure 7: Cumulative Returns of QuantAgents during live trading (24Q3-25Q1). RAM were held 36 times in the A-stock and 46 times in the HK stock market.

baselines and variations in meeting structures, our framework demonstrated strong performance across all metrics, resulting in an impressive overall return of nearly 300%.

Limitations

We still have the following limitations: 1) In terms of generalization, our QuantAgents is modular and flexible, designed to adapt to various market scenarios. We will explore generalization further in future work. 2) Although backtesting may be affected by potential LLM information leakage, QuantAgents' effectiveness is proven through impeccable live trading performance in A-stock and HK-stock markets over three quarters. We plan to validate it across global markets and report performance periodically.

Ethical Impact

We respect intellectual property rights and comply with relevant laws and regulations. The documents in our dataset are publicly available, and we have taken careful measures to ensure that the documents in our dataset do not contain any personal sensitive information. In addition, our work is only for research purposes, not for commercial purposes.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (62372187), in part by the National Key Research and Development Program of China (2022YFC3601005) and in part by the Guangdong Provincial Key Laboratory of Human Digital Twin (2022B1212010004). The authors are highly grateful to the anonymous reviewers for their careful reading and insightful comments.

References

Carol Alexander. 2008. *Market Risk Analysis, Volume III: Pricing, Hedging, and Trading Financial Instruments*. Wiley.

Anthropic. 2024. Claude 3.5 sonnet. Available on https://claude.ai/. Accessed: 2024-07-15.

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

Tolga Buz and Gerard de Melo. 2023. Democratization of retail trading: Can reddit's wallstreetbets outperform investment bank analysts? *arXiv preprint*.

Kinjal Chaudhari and Ankit Thakkar. 2023. Data fusion with factored quantization for stock trend prediction using neural networks. *Information Processing & Management*, 60(3):103293.

Peng Chen, Pi Bu, Yingyao Wang, Xinyi Wang, Ziming Wang, Jie Guo, Yingxiu Zhao, Qi Zhu, Jun Song, Siran Yang, Jiamang Wang, and Bo Zheng. 2025. Combatvla: An efficient vision-language-action model for combat tasks in 3d action role-playing games. *Preprint*, arXiv:2503.09527.

Louis R. Eeckhoudt and Roger J. A. Laeven. 2018. Dual moments and risk attitudes. *Preprint*, arXiv:1612.03347.

Eugene F. Fama and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.

Abhijeet Gaikwad, Viet_Dung Doan, Mireille Bossy, Françoise Baude, and Frédéric Abergel. 2012. Superquant financial benchmark suite for performance analysis of grid middlewares. In *Modeling, Simulation and Optimization of Complex Processes*, pages 103–113, Berlin, Heidelberg. Springer Berlin Heidelberg.

Team GLM, Aohan Zeng, et al. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *Preprint*, arXiv:2406.12793.

Jihao Gu, Qihang Ai, Yingyao Wang, Pi Bu, Jingxuan Xing, Zekun Zhu, Wei Jiang, Ziming Wang, Yingxiu Zhao, Ming-Liang Zhang, Jun Song, Yuning Jiang, and Bo Zheng. 2025. Mobile-r1: Towards interactive reinforcement learning for vlm-based mobile agent via task-level rewards. *Preprint*, arXiv:2506.20332.

Yunsoo Ha and Juliane Mueller. 2024. Adaptive sampling bi-fidelity stochastic trust region method for derivative-free stochastic optimization. *Preprint*, arXiv:2408.04625.

Tuomas Haarnoja, Aurick Zhou, Kristian Hartikainen, George Tucker, Sehoon Ha, Jie Tan, Vikash Kumar, Henry Zhu, Abhishek Gupta, Pieter Abbeel, et al. 2018. Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*.