Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2023. Metagpt: Meta programming for a multi-agent collaborative framework. *Preprint*, arXiv:2308.00352.

Takashi Kanamura, Lasse Homann, and Marcel Prokopczuk. 2021. Pricing analysis of wind power derivatives for renewable energy risk management. *Applied Energy*, 304:117827.

Lyes Koliai. 2016. Extreme risk modeling: An evt–pair-copulas approach for financial stress tests. *Journal of Banking Finance*, 70:1–22.

Gang Kou, Xiangrui Chao, Yi Peng, Fawaz E Alsaadi, Enrique Herrera Viedma, et al. 2019. Machine learning methods for systemic risk analysis in financial sectors.

Xiangyu Li, Xinjie Shen, Yawen Zeng, Xiaofen Xing, and Jin Xu. 2024. Finreport: Explainable stock earnings forecasting via news factor analyzing model. *Preprint*, arXiv:2403.02647.

Xiangyu Li, Yawen Zeng, Xiaofen Xing, Jin Xu, and Xiangmin Xu. 2025. Hedgeagents: A balanced-aware multi-agent financial trading system. *Preprint*, arXiv:2502.13165.

Harald A Mieg. 2022. Volatility as a transmitter of systemic risk: Is there a structural risk in finance? *Risk Analysis*, 42(9):1952–1964.

Tobias J Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. 2012. Time series momentum. *Journal of financial economics*, 104(2):228–250.

OpenAI. 2023. text-embedding-3-large. Available at: `https://openai.com/index/new-embedding-models-and-api-updates/`.

OpenAI, Josh Achiam, et al. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Keyu Pan and Yawen Zeng. 2023. Do llms possess a personality? making the mbti test an amazing evaluation for large language models. *Preprint*, arXiv:2307.16180.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *Preprint*, arXiv:2304.03442.

Molei Qin, Shuo Sun, Wentao Zhang, Haochong Xia, Xinrun Wang, and Bo An. 2023. Earnhft: Efficient hierarchical reinforcement learning for high frequency trading. *arXiv preprint arXiv:2309.12891*.

Theodore R Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L Griffiths. 2023. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427*.

Shuo Sun, Molei Qin, Xinrun Wang, and Bo An. 2023a. Prudex-compass: Towards systematic evaluation of reinforcement learning in financial markets. *Preprint*, arXiv:2302.00586.

Shuo Sun, Molei Qin, wentao zhang, Haochong Xia, Chuqiao Zong, Jie Ying, Yonggang Xie, Lingxuan Zhao, Xinrun Wang, and Bo An. 2023b. Trademaster: A holistic quantitative trading platform empowered by reinforcement learning. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Zhicheng Wang, Biwei Huang, Shikui Tu, Kun Zhang, and Lei Xu. 2021a. Deeptrader: a deep reinforcement learning approach for risk-return balanced portfolio management with market conditions embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 643–650.

Zhicheng Wang, Biwei Huang, Shikui Tu, Kun Zhang, and Lei Xu. 2021b. Deeptrader: A deep reinforcement learning approach for risk-return balanced portfolio management with market conditions embedding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):643–650.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

Yiqi Wu, Xiaodan Hu, Ziming Fu, Siling Zhou, and Jiangong Li. 2024. Gpt-4o: Visual perception performance of multimodal large language models in piglet activity understanding. *Preprint*, arXiv:2406.09781.

An Yang, Baosong Yang, Binyuan Hui, et al. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023a. Fingpt: Open-source financial large language models. *arXiv preprint arXiv:2306.06031*.

Hui Yang, Sifu Yue, and Yunzhong He. 2023b. Auto-gpt for online decision making: Benchmarks and additional opinions. *Preprint*, arXiv:2306.02224.

Xiao Yang, Weiqing Liu, Dong Zhou, Jiang Bian, and Tie-Yan Liu. 2020. Qlib: An ai-oriented quantitative investment platform. *Preprint*, arXiv:2009.11189.

Jianfeng Yu and Yu Yuan. 2011. Investor sentiment and the mean–variance relation. *Journal of Financial Economics*, 100(2):367–381.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. 2023. Finmem: A performance-enhanced llm trading agent with

layered memory and character design. *Preprint*, arXiv:2311.13743.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, Longtao Zheng, Xinrun Wang, and Bo An. 2024. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. *Preprint*, arXiv:2402.18485.

A Overview of Appendix

We have nearly 15 pages of this appendix, comprising the following subsections for the convenience of readers:

More details about of our framework

- **Definitions of Single Agent**: This section provides comprehensive instructions on tools, memory, and other details.
- **Prompt Templates for Various Tasks**: This section details the prompt templates used for various tasks within our framework.
- **Profiles of Agents**: A thorough exposition presenting detailed profiles of each agent.
- **Construction of the Strategy Pool**: This section elaborates on the methodology employed for constructing the strategy pool, which is pivotal for evaluating and refining trading strategies within the QuantAgents framework.

More details about of our setting

- **PRUDEX Evaluation Benchmark**: An evaluation benchmark assessing performance across multiple dimensions.
- **Details of Dataset Setup**: Includes details of our datasets.
- **Details of Evaluation Metrics**: Includes the calculation of associated metrics.
- **Details of Baselines**: Comprehensive descriptions of our competitors.

More additional experiments

- **Experiment of Ablation Study**: Additional experiments focusing on ablation study.
- **Single-Asset Performance Comparison**: This section presents a performance comparison for a single-asset scenario, focusing on Apple Inc. (AAPL) stock from 2021 to 2023, to evaluate the effectiveness of QuantAgents against baseline models.
- **Empirical Evaluation of QuantAgents in Live Trading**: This section describes the performance of QuantAgents in real-world trading scenarios within the Chinese market, covering Q1-Q3 of 2024.

B Definitions of Single Agent

In this section, we will provide a comprehensive overview of the composition and execution process of a single agent, designed to simulate the human decision-making process in investments.Each agent comprises a range of financial analysis tools, along with definitions for action, memory, profile, reflection and the execution workflow.

B.1 Tool

The tool module $\mathcal{T}$ encompasses a comprehensive suite of technical and analytical tools for investment decision-making, including:

- t_1: **Technical Indicator Analysis**, providing analysis of traditional technical indicators such as moving averages, relative strength index, and others;
- t_2: **Sentiment Analysis from Social Media**, gauging market sentiment through the analysis of social media platforms;
- t_3: **Algorithmic Trading Strategies**, employing algorithms to identify trading opportunities and execute trades;
- t_4: **Regulatory Change Impact Analysis**, assessing the potential impact of regulatory changes on the market;
- t_5: **Economic Indicator Forecasting**, predicting future economic conditions by analyzing leading economic indicators;
- t_6: **Corporate Earnings Analysis**, scrutinizing financial reports to evaluate corporate performance;
- t_7: **NASDAQ-100 Index Component Tracking**, monitoring the performance of individual components within the NASDAQ-100 Index;
- t_8: **Sector Performance Evaluation**, assessing the performance of different industry sectors for sector-specific investment decisions;
- t_9: **Risk-Adjusted Return Analysis**, measuring the return of an investment in relation to its risk;

- t_{10}: **Portfolio Diversification Tools**, aiding in the strategic distribution of investments across various asset classes;
- t_{11}: **Central Bank Policy Analysis**, interpreting the implications of central bank policies on currency values and economic conditions;
- t_{12}: **Global Macroeconomic Trend Analysis**, examining large-scale economic trends and their impact on global markets;
- t_{13}: **Currency Pair Correlation Matrix**, studying the correlation between different currency pairs for informed trading decisions;
- t_{14}: **Interest Rate Differential Analysis**, analyzing the effects of interest rate differentials on currency exchange rates;
- t_{15}: **Asset Allocation Optimization**, strategically allocating investments to maximize returns and minimize risk;
- t_{16}: **Risk Management Frameworks**, employing frameworks to identify, assess, and mitigate investment risks;
- t_{17}: **Portfolio Stress Testing**, simulating the impact of extreme market conditions on the portfolio to evaluate its resilience;
- t_{18}: **Derivatives Strategy Formulation**, creating strategies involving derivatives to hedge risks and enhance returns;
- t_{19}: **Fund Performance Evaluation**, measuring and assessing the performance of investment funds against benchmarks and objectives;
- t_{20}: **FinReport**, generating detailed financial reports to provide insights into company and market performance;
- t_{21}: **Trend Forecasting**, predicting future market trends based on historical data and predictive analytics;
- t_{22}: **Volatility Assessment Tool**, analyzing market volatility to better inform investment decisions;
- t_{23}: **Simulation Optimization Toolkit**, optimizing trading strategies through simulation techniques;
- t_{24}: **Strategy Analysis Suite**, providing comprehensive analysis of investment strategies for performance evaluation;
- t_{25}: **RiskAnalyzer Toolkit**, assessing and quantifying various risk factors in the investment portfolio;
- t_{26}: **Risk Score Assessment Tool**, calculating a risk score to guide investment decisions based on the overall risk profile.

B.2 Action

The specific actions $\mathcal{A}$ that agent can execute include:

- a_1: **Buy/Sell/Hold the current assets**, making decisions on whether to acquire new assets, divest existing ones, or maintain the current position;
- a_2: **Adjust the quantity and price of securities to be bought or sold**, fine-tuning the volume and pricing strategy for securities transactions;
- a_3: **Set or modify trading stop-loss, take-profit, and other trading strategy conditions**, implementing or revising parameters for automated trading strategies to manage risk and lock in profits;
- a_4: **Adjust the risk exposure of the investment portfolio**, allocating budget weights and modifying the portfolio to achieve the desired level of risk exposure;
- a_5: **Execute Asset Allocation**, strategically distributing investment capital across various asset classes to optimize the portfolio's risk and return profile;
- a_6: **Initiate Risk Assessment Protocols**, beginning the process of evaluating potential risks and determining the appropriate measures to mitigate them;
- a_7: **Authorize Capital Deployment**, approving the use of funds for investment opportunities in line with the asset allocation strategy;
- a_8: **Enforce Compliance with Regulatory Standards**, ensuring that all investment activities adhere to the legal and regulatory framework governing financial markets.