M&A Agent: Generating Strategic Insight Through Role-Based Agent Simulation

Anonymous ACL submission

Abstract

Large language models (LLMs) and LLM-based agent systems have shown considerable potential in investment and stock trading. However, their use in corporate finance, particularly in strategic decision-making tasks such as merger and acquisition (M&A) evaluation, remains underexplored. Traditional analysis and machine learning methods often struggle in this domain due to the limited availability of target company data. To address this challenge, we propose M&A Agent, a multi-agent framework built on LLMs that simulates the M&A process and assesses deal value. The framework consists of two stages: M&A simulation and value evaluation. Following real-world procedures, the simulation includes financial analysis, negotiation, board decision-making, and regulatory review. Through structured interactions among agents, the system transforms static financial data into richer, more dynamic information. This simulation is then reviewed by an evaluation committee of agents, which assigns a score and provides justification. Experiments on real-world M&A cases demonstrate that our method produces significantly better deal value rankings compared to baselines, as measured by NDCG. The code is publicly available at `https://anonymous.4open.science/r/2AB73965` to support reproducibility.

1 Introduction

With the rapid advancement of large language models, their impressive capabilities in reasoning, autonomous planning, and related tasks have become increasingly evident (Ruan et al., 2023; Jin et al., 2024; Jin and Lu, 2023; Zhong et al., 2024). These developments have also had a profound impact on financial research (Li et al., 2023a; Liang et al., 2024; Wu et al., 2023b; Xie et al., 2024; Li et al., 2023b). However, research in the area of corporate finance strategic decision-making remains scarce.

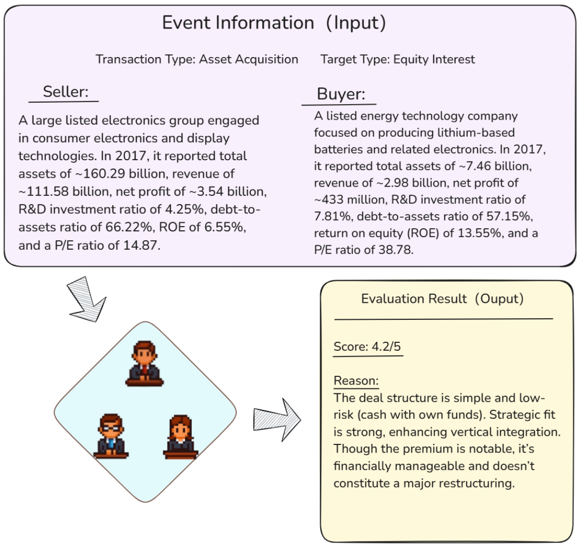

Figure 1: Overview of the M&A evaluation task. The agent processes firm-level financial data and outputs: (1) A deal score with accompanying justification; (2) Concise, financially grounded reasoning that reflects strategic considerations such as deal structure, integration potential, and financial impact.

Meanwhile, LLM-based agent system have been applied across multiple fields, including code generation (Islam et al., 2024; Huang et al., 2023; Islam et al., 2025), judicial reasoning (He et al., 2024; Yuan et al., 2024), hallucination detection (Cheng et al., 2024), drama creation (Han et al., 2024), medical necessity justification (Pandey et al., 2024), investment management (Fatemi and Hu, 2024; Li et al., 2024a), and long-document question answering (Zhao et al., 2024). These studies have demonstrated the tremendous potential of agent framework in complex decision-making tasks. We believe that LLM-based agent can provide significant support for research in corporate finance and strategic decision-making, while also inspiring new agent framework designs driven by the practical challenges in these areas. Therefore, this study focuses on the widely discussed topic of mergers and acquisitions (M&A) evaluation as the application scenario.

For years, M&A have served as a central strategy for firms aiming at rapid expansion and strategic repositioning (Jensen, 1986). By integrating operations and leveraging synergies, they are expected to improve competitive advantage in dynamic markets. Yet, a substantial body of research indicates that most M&A transactions fail to deliver the anticipated value. Empirical estimates suggest failure rates exceeding 60% or even 70%, attributed to overestimated synergies, weak integration, or cultural misalignment (Joshi et al., 2020; King et al., 2004; Dao and Bauer, 2021; Thompson and Kim, 2020; Meglio and Risberg, 2011). These outcomes highlight the inherent complexity of M&A decisions and the ongoing need for more reliable methods to evaluate deal quality.

Traditional approaches, such as regression analysis and financial ratio assessments, or more recent machine learning methods predicting post-merger performance, struggle to address several difficulties in M&A event evaluation: (1) In M&A cases involving non-public companies, historical financial data is often unavailable or must be manually collected from limited sources. This information gap reduces the likelihood that a merger achieves its intended outcomes and undermines the effectiveness of traditional and machine learning methods (Officer et al., 2009; Borochin et al., 2016; Alexandridis et al., 2024);(2) Multiple and complex performance factors – the factors influencing M&A performance include leadership negotiations, internal agreements, and regulatory approvals, which add layers of complexity to the evaluation process. Traditional models often simplify these diverse factors, failing to fully account for the intricate interactions in M&A processes such as negotiations, regulatory reviews, and internal voting, which leads to incomplete or biased assessments of potential value; (3) Lack of interpretability — machine learning methods often fail to explain the rationale behind their predictions, making it difficult to understand the basis for decision-making. In contrast, agent-based simulations can offer interpretable outputs.

To address these issues, we propose M&A Agent, a multi-agent framework for M&A evaluation. As shown in Figure 1, the framework takes financial data from both parties and outputs a deal score along with clear, reasoned justifications to enhance the framework's credibility and interpretability. The framework includes CFO agents, CEO agents, Board agents, Regulatory agents, and Evaluation Committee agents. The first four agents represent key roles in the M&A process, responsible for decision-making, providing feedback, and interacting with each other during the simulation. This allows the framework to dynamically generate information and let the Evaluation Committee agent provide an overall assessment after the simulation ends.

The CFO agent initiates the process by extracting key information from financial data and assessing the financial status of both the target and acquiring companies, providing essential data support for downstream decision-making. The CEO agent leads the negotiation, simulating strategies and responses from both sides to coordinate interests and reach an agreement. It leverages CFO insights to formulate and adapt negotiation strategies that maximize its own side's benefits. Once a preliminary deal is reached, the Board agent evaluates the outcome and votes on it. Each board member (agent) assesses the proposal based on strategic alignment and long-term interests, with their collective votes determining the final decision. Following board approval, the Regulatory agent reviews the deal for legal and compliance issues, simulating the regulatory process and offering recommendations. Finally, the Evaluation Committee agent conducts a comprehensive assessment of the entire M&A process, assigning a final score and rationale. It integrates feedback from all stages, weighing financial performance, strategic fit, and risk, ensuring the decision is both immediately viable and aligned with long-term goals.

Through this multi-agent collaboration, the framework can fully simulate the interactions and decisions of different roles in the M&A process, addressing the complexity that traditional methods cannot handle. Each agent makes decisions based on its independent role and goals, while interacting and providing feedback with other agents, ensuring that the final M&A evaluation is closer to real-world decision-making processes. Our framework not only processes existing data but also generates real-time information based on simulations, offering a more comprehensive and accurate M&A evaluation.

We summarize our contributions as follows:

- We propose a simulation–decision structure inspired by real-world M&A scenarios, where financial data serves as input and new, informative signals are generated through structured agent-based interactions. This framework moves beyond static evaluation and enables agents to actively reason, coordinate, and produce useful insights for decision-making.

- We extend agent-based frameworks beyond investment management into the domain of corporate finance and strategic decision-making, demonstrating their applicability to high-level planning tasks such as M&A evaluation.

- Our framework simulates the M&A process, including negotiation, board voting, and regulatory review, to generate additional information that is not available before a real transaction. These simulations offer plausible future scenarios rather than exact predictions, helping agents make more informed decisions under limited information, thereby improving the accuracy and credibility of the evaluation. Therefore, agent-based frameworks can meaningfully support strategic decision-making. Extensive experiments confirm the effectiveness of our approach in M&A evaluation tasks.

2 Related Work

Post-M&A forecast The event study method evaluates market expectations of M&A performance by analyzing stock market reactions before and after merger announcements (Malmendier et al., 2018). However, this method primarily captures short-term effects and may be subject to market biases (Malhotra and Chauhan, 2018). Accounting-based performance measures use financial ratios—such as profitability, efficiency, and liquidity metrics—to assess post-merger operational and financial outcomes (Moeller et al., 2005). Renneboog and Vansteenkiste (2019) highlight that horizontal mergers are associated with significantly improved post-merger ROA. In recent years, machine learning and AI-driven models have gained attention for their ability to handle complex datasets and nonlinear relationships in M&A performance prediction (Zhang et al., 2024). For example, deep learning models have been used to forecast financial trends in high-tech sector mergers (Tarasov and Dessoulavy-Śliwiński, 2024). However, traditional methods often fail to estimate deal value before the M&A event begins, while machine learning models often lack interpretability.

Multi-Agent Framework in Financial Research In financial research, multi-agent systems have been explored for their simulation capabilities. For instance, hundreds of agents have been organized into artificial societies to simulate and validate macroeconomic phenomena (Li et al., 2024b). In stock trading, agent frameworks vary widely—Yu et al. (2024), for example, designed a manager-trader framework inspired by real-world trading structures. Debate has been found to enhance agent reasoning (Qian et al., 2024; Wu et al., 2023a; Hong et al., 2024). Building on this, Xiao et al. (2025) introduced a buy-side vs. sell-side debate mechanism to improve trading performance. To improve factor discovery and understand model design rationale, Li et al. (2024d) proposed a multi-agent framework for factor mining. Building upon these prior efforts, our work aims to extend the application of multi-agent systems into corporate finance and strategic decision-making by proposing a novel agent-based framework for M&A evaluation.

3 FrameWork

Figure 2 shows an overview of our proposed framework for M&A valuation and financial forecasting, comprising two modules: the Simulation Module and the Evaluation Module. Utilizing LLMs, it processes structured data and unstructured data (e.g., financial reports) to simulate M&A decision-making and predict financial outcomes.

3.1 Simulation Module

Reflecting real-world M&A practice and supported by prior studies, our framework simulates the decision-making process with four agents: CFO, CEO, Board of Directors, Regulatory Authority(Ferris and Sainani, 2021).

Financial Feasibility Analysis The CFO agent assesses the target's financial health by analyzing structured metrics, like return on equity, and unstructured insights, such as company's business overview. It generates a feasibility report detailing synergy potential (e.g., cost savings, market expansion) and risks (e.g., high debt), providing critical inputs for negotiation and evaluation stages.

Negotiation Simulation The CEO agent leads the negotiation by adjusting terms such as acquisition price or equity share, based on the CFO's report, buyer constraints, and market dynamics (e.g., competitor bids). To simulate realistic decision-making, the negotiation is designed to always reach an

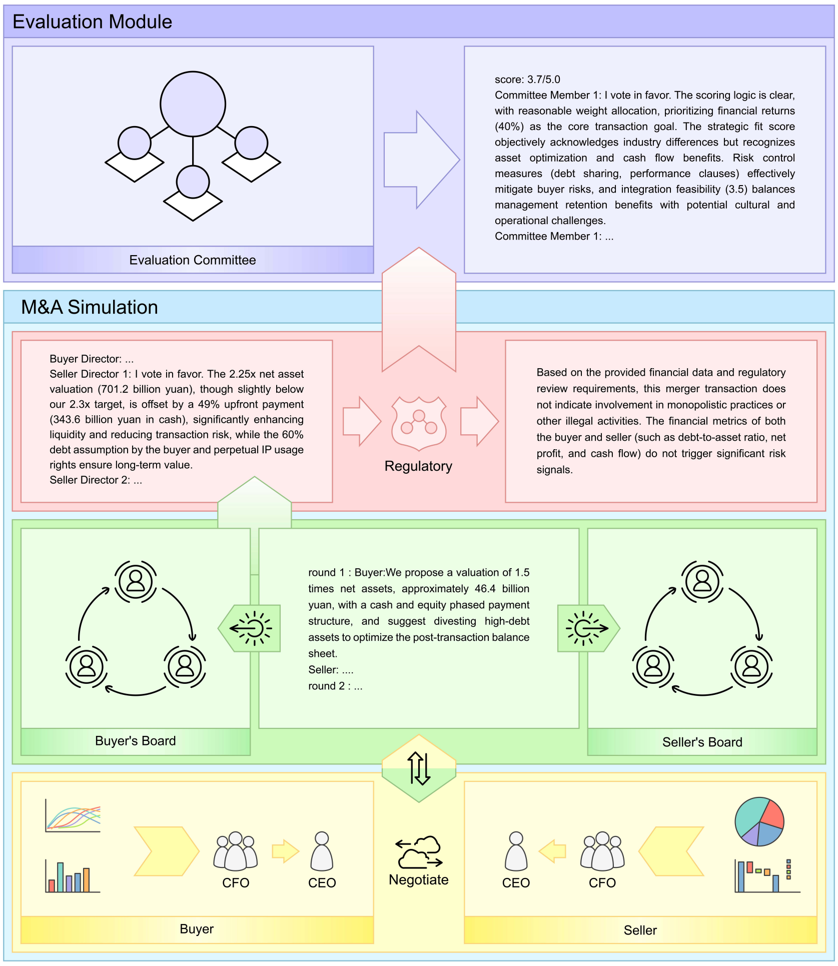

Figure 2: Our framework overview: The M&A simulation component reenacts the company acquisition process through agent interactions, using limited financial information to explore potential future outcomes of the deal. The evaluation module assesses the event based on the extracted information.

agreement, prompting the agent to adapt and compromise as needed. The LLM-based agent supports dynamic bargaining by incorporating feedback from previous exchanges, with all interactions logged to ensure transparency.

Board Approval Process The Board agent evaluates negotiation outcomes and votes on proposals based on alignment with firm goals, such as market expansion, and constraints like budget limits. LLMs analyze negotiation logs and feasibility reports to assess risks and benefits, ensuring that decisions reflect strategic priorities and governance principles. The board is designed to conduct up to three voting rounds, after which the process proceeds regardless of the outcome.

Regulatory Compliance Check The Regulatory agent reviews the deal for compliance with legal and market regulations, analyzing texts such as antitrust laws or securities filings. Identifies risks, such as monopolistic concerns, and suggests modifications,

3.2 Evaluation Module

This module, driven by the Evaluation Committee—comprising multiple agent evaluators—produces a final predictive score as the system's result, reflecting the expected growth amplitude of financial metrics. One agent proposes a score and rationale, others vote, and if consensus is not reached, voting rationales are passed to the next agent until approval. For cost-effectiveness and efficiency, the debate and voting process adopts a sparse communication topology(Li et al., 2024c). The committee autonomously designs the scoring formula based on CFO reports, CEO negotiation logs, and Board voting outcomes.

Preliminary Assessment The Evaluation Committee aggregates CFO feasibility reports, CEO negotiation logs, and Board voting results to estimate the target's growth potential. LLMs integrate financial projections, such as revenue growth, and qualitative factors, such as technological synergies, to form a preliminary assessment. This assessment guides an agent in proposing an initial score and rationale for subsequent voting.synergies, to produce a preliminary assessment that informs the design of the scoring formula.

Scoring Calculation and Voting One agent proposes an initial score and rationale, autonomously designing a formula that weights factors like profit growth, synergies, and risks based on transmitted data. Other agents vote on the score. If consensus is not reached, the voting rationale of each agent (for example, concerns about synergy weighting) is passed to the next, iterating until a score is approved, ensuring that it reflects the expected growth amplitude of financial metrics.

Output Generation The module outputs the approved predictive score as the system's result, directly indicating the expected growth amplitude of financial metrics, such as percentage increases in profits or stock returns. Assisted by the final rationale, such as key synergy drivers or risk mitigations, the score enables enterprises to select optimal targets and investors to identify high-return M&A events.

3.3 LLM Setting

To ensure a fair comparison with baseline methods, we use minimal and standardized prompts across all components in our framework. All agents are powered by the `gpt-3.5-0125`, which has approximately 175 billion parameters (OpenAI, 2023). The model is configured with a temperature of 0 to produce consistent and deterministic outputs. This controlled setup isolates the effect of our proposed components without introducing variability from prompt engineering. Detailed prompt templates are provided in the **Appendix** A.

4 Experiment

4.1 Experimental Setup

4.1.1 Data Collection

We collect a dataset of 281 real-world M&A cases from the CSMAR[1] database, spanning August 2004 to August 2019. For ease of collection and consistency, we select only cases in which both the acquirer and the target are publicly listed companies in China. This selection does not compromise the generality of our study, as we only utilize a limited set of commonly accessible financial indicators for each firm, as detailed in Appendix B. Each case includes buyer and seller profiles, target details, information available prior to the transaction—including planned payment method, funding source, restructuring type, and region type. The dataset provides diversity in transaction types and industries, with verified accuracy and completeness to support simulation and evaluation of financial outcomes. The distribution of M&A types and funding sources is illustrated in Figure 3.

[1] A widely used financial database in China.

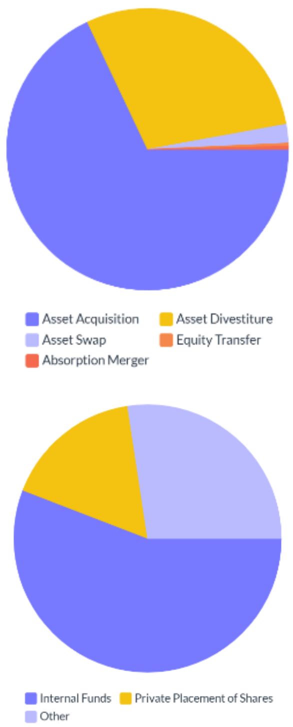

Figure 3: Distribution of Restructuring Types (Top) and Funding Sources (Bottom).

4.1.2 Evaluation Metrics

To evaluate the performance of enterprises following M&A restructuring, we select three representative financial metrics: **Return on Assets (ROA)**, **Return on Equity (ROE)**, and **Tobin's Q**. Their growth rates, calculated as the percentage change post-M&A, serve as indicators of financial and market performance, reflecting enhanced profitability and value creation.

ROA measures a company's efficiency in generating profits from its assets.A higher ROA indicates effective asset management and operational performance.

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}} \tag{1}$$

Method	ROA				ROE				Tobin's Q			
	R	MSE	MAE	NDCG	R	MSE	MAE	NDCG	R	MSE	MAE	NDCG
Vanilla	-0.072	0.025	0.129	0.573	-0.068	0.070	0.218	0.492	-0.018	57.872	6.530	**0.854**
CoT	-0.142	0.018	0.094	0.497	-0.149	0.056	0.172	0.484	0.038	53.642	6.674	0.771
CoT-SC	-0.110	0.067	0.234	0.701	-0.062	0.187	0.390	0.732	-0.098	**39,903**	**4.740**	0.544
Chain of Logic	0.058	0.012	0.084	0.831	0.044	**0.031**	**0.126**	0.755	0.035	80.710	8.473	0.725
React	-0.082	0.015	0.096	0.566	-0.115	0.043	0.168	0.555	-0.043	68.871	7.609	0.527
M&A agent	**0.254**	**0.008**	**0.077**	**0.844**	**0.333**	0.041	0.147	**0.864**	**0.095**	63.124	7.263	0.807

Table 1: Performance of each method across three financial indicators, with four evaluation metrics reported for each. Growth rate = (current year – post-merger year) / post-merger year.

where **Net Income** is profit after expenses and **Total Assets** includes all company resources.
ROE assesses profitability relative to shareholders' equity. A higher ROE reflects strong returns for investors, signaling robust financial management.

$$\text{ROE} = \frac{\text{Net Income}}{\text{Total Equity}} \tag{2}$$

where **Total Equity** is shareholders' residual value after liabilities.
Tobin's Q reflects market expectations of future growth by comparing a firm's market value to its asset replacement cost.A higher Tobin's Q suggests strong market confidence in future growth.

$$\text{Tobin's Q} = \frac{\text{Market Value of Firm}}{\text{Replacement Cost of Assets}} \tag{3}$$

where **Market Value** is typically market capitalization, and **Replacement Cost** approximates total assets.

To assess the ability to accurately predict post-M&A performance and identify high-potential M&A events, we employ four metrics: **Pearson Correlation Coefficient(R)**, **Mean Squared Error (MSE)**, **Mean Absolute Error (MAE)**, and **Normalized Discounted Cumulative Gain (NDCG)**. These metrics evaluate the precision and ranking quality of the system's predictive scores against actual financial outcomes.
Pearson Correlation Coefficient evaluates the linear relationship between predicted and actual growth rates, ranging from -1 to 1. A higher positive coefficient indicates stronger alignment.

$$R = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \tag{4}$$

MSE measures the average squared difference between predicted and actual financial metric growth rates. Lower MSE signifies better prediction reliability.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{5}$$

MAE quantifies the average absolute difference between predicted and actual financial metric growth rates. Lower MAE indicates higher predictive accuracy.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{6}$$

NDCG assesses the ranking quality of predicted M&A events by comparing the system's ranked list to actual performance. Higher NDCG reflects better identification of high-value M&A opportunities.

$$\text{DCG@}k = \sum_{i=1}^{k}\frac{\text{rel}_i}{\log_2(i+1)} \tag{7}$$

$$\text{IDCG@}k = \sum_{i=1}^{k}\frac{\text{rel}_i^{\text{(ideal)}}}{\log_2(i+1)} \tag{8}$$

$$\text{NDCG@}k = \frac{\text{DCG@}k}{\text{IDCG@}k} \tag{9}$$

where rel_i denotes the relevance score of the item at position i, $\log_2(i+1)$ is the positional discount factor, and $\text{rel}_i^{\text{(ideal)}}$ represents the relevance score under ideal ranking. The mapping from financial growth to rel_i is provided in **Appendix C**. NDCG@k quantifies ranking quality by normalizing DCG@k against IDCG@k, where 1 indicates perfect alignment with the ideal order.We use $k = 5$ in our evaluation.

4.1.3 Baselines

We compare M&A evaluation system against the following baselines:
Vanilla We use `gpt-3.5-0125` with few-shot prompting as the vanilla model.
CoT (Wei et al., 2023) improves reasoning by prompting models to break down complex problems into sequential, logical steps. Each step is addressed explicitly, building toward a final answer through structured reasoning.
CoT-SC (Wang et al., 2023) enhances reasoning by prompting models to systematically evaluate their

own responses. It involves decomposing a problem into logical steps, generating an initial answer, and then critically assessing its accuracy, completeness, and coherence. Through iterative self-reflection, COT-SC refines outputs,

Chain of Logic (Servantez et al., 2024) facilitates rule-based reasoning by breaking down complex rules into individual elements, addressing each element systematically, and combining them via a logical expression to derive the final answer.

ReAct (Yao et al., 2023) This system enables the agent to improve its actions based on the outcomes of past activities.

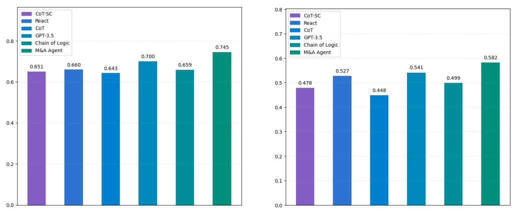

Figure 4: Left: 1,000 groups of 5 events each; selecting the top 2 events is considered correct. Right: 1,000 groups of 10 events each; selecting the top 3 events is considered correct.

4.2 Main Result

Table 1 presents the performance comparison across different methods on three financial indicators: ROA, ROE, and Tobin's Q. Our proposed M&A Agent consistently outperforms all baselines across nearly all metrics, demonstrating superior predictive accuracy and ranking quality. It achieves the highest correlation scores (R), lowest error rates (MSE and MAE), and the best NDCG values in most cases. For example, in the ROE prediction task, the M&A agent achieves an R of 0.333 and an NDCG of 0.864—substantially higher than the second-best method. Similar advantages are observed on ROA and Tobin's Q, confirming the robustness and effectiveness of our agent in capturing complex patterns in M&A events. These results highlight the value of incorporating simulation-based decision processes into financial event modeling. Performance on Tobin's Q, while still competitive, is less stable across systems, as the indicator's complexity makes its trend harder to capture.

Performance on Tobin's Q, while still competitive, is less consistent across systems. This is likely due to the abstract and multifactorial nature of Tobin's Q, which incorporates both market valuations and asset structures. Such complexity makes it inherently difficult to capture meaningful trends. Consequently, scoring results on this metric tend to exhibit greater randomness, with less clear separation between methods compared to ROA and ROE.

4.3 Ablation Result

Model	R	MSE	MAE	NDCG
w/o board	0.096	0.285	0.438	0.732
w/o negotiation	0.076	0.390	0.538	0.677
w/o regulatory	0.091	0.315	0.477	0.722
complete agent	**0.228**	**0.277**	**0.435**	**0.838**

Table 2: Results of the ablation study. Each metric represents the average performance across three financial indicators. Due to the large difference in scale between Tobin's Q and the other indicators, we use the geometric mean to compute the overall MAE and MSE. Detailed results for each individual indicator are provided in **Appendix D**.

As shown in Table 2, all three components of our proposed framework—board, negotiation, and regulatory—significantly enhance the agent's ability to assess M&A events, with the negotiation module demonstrating the most pronounced effect.

How it works We believe that the M&A simulations carried out by these components provide the agent with meaningful information, rather than purely fictitious or unreliable projections. While the simulated outcomes may not exactly predict future developments, they approximate the complexity of real-world scenarios. Importantly, this process generates additional decision signals—such as negotiation outcomes, internal voting patterns, and regulatory opinions—that are not present in the original financial data but significantly enrich the context for evaluation. The effectiveness of these simulated signals is evidenced by the improved performance of the agent in assessing M&A value, as confirmed by the results of the ablation study.

The exceptional contribution of the negotiation module, in particular, can be attributed to the richness and completeness of the information it offers. In our simulation process, we enforced successful negotiation as a requirement, which often led to multiple rounds of negotiation in the simulated events. In contrast, the board module did not require a majority approval, and the regulatory module merely provided advisory opinions. This variation in the quantity and quality of information across components is clearly reflected in their respective impacts on the overall framework performance.

4.4 Discussion and Analysis

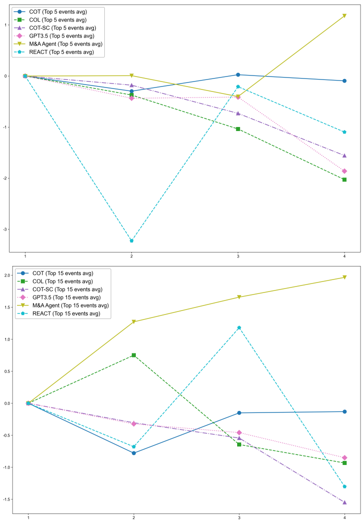

Figure 5: Average financial growth rates of top-scoring events by system. X-axis: years after merger; Y-axis: average growth rate of financial indicators. Top: top 5 events; bottom: top 15. Growth rate = (current year – first post-merger year) / first post-merger year.

To further evaluate the applicability and robustness of our framework in identifying high-potential M&A targets and assessing the value of M&A events, we conducted a robustness test using randomized event combinations. Specifically, we randomly grouped all collected M&A events into m combinations, each containing n events. For each group, if the top k events ranked by our framework also exhibited the highest average financial growth, the group was considered correctly identified. The accuracy comparison results are shown in Figure 4.

The results indicate that the M&A Agent's evaluations generally align with actual event value. However, when the number of candidate events becomes large, it becomes increasingly difficult to precisely distinguish the highest-potential events. We argue, however, that this does not undermine the framework's overall effectiveness.

To further illustrate this, we compare the average financial growth rates of the top k events selected by each system in Figure 5. Here, financial growth is computed relative to the first post-merger year, which serves as a baseline. This approach minimizes the influence of short-term fluctuations and emphasizes long-term performance trends, while also serving to test the robustness of our method under different growth rate calculation methods. Results show that the top five events selected by the M&A Agent achieve the highest compound growth rate four years after the merger, and the top fifteen events reach their performance peak as early as the second year post-merger—outperforming all other systems. These findings strongly suggest that the M&A Agent consistently assigns higher scores to genuinely high-value, high-potential events. This observation is also consistent with our NDCG evaluation results.

These experimental results align with our expectations. The newly generated information from the simulation proves to be effective. Based on this, we argue that when agents operate with the same information and under the same environmental conditions as humans, their behaviors tend to converge with human decision-making patterns. This behavioral similarity reinforces our confidence in the validity of using agent-based simulations for strategic tasks. We believe that large language models and intelligent agents possess the capability to assist decision-makers in strategy formulation and even anticipate possible future developments.

5 Conclusion

This work introduces a multi-agent framework that simulate real-world processes in M&A decisions. A key feature of the framework is its ability to generate new information through structured simulation, rather than relying solely on existing data. This expands the scope of language model applications from static prediction to dynamic scenario reasoning. The approach offers potential value for both AI research on decision-making and empirical studies of complex economic events.

Ethics Statement

In this work, we ensured ethical compliance by exclusively using publicly available academic datasets, strictly avoiding any private or sensitive data. The licenses and terms of use for all datasets and artifacts were reviewed and respected. Our use of these existing artifacts aligns with their intended purposes as specified by their licenses. Furthermore, the framework we developed is intended solely for academic research and complies with the original access conditions. Any derivative data generated for research purposes is used strictly within the scope of academic research and not beyond.

AI tools were employed solely to enhance the clarity and correctness of the writing, without contributing to the generation of content or ideas, thereby maintaining the originality and integrity of the research.

Limitations

We acknowledge several limitations of our current research:

- While the agent-based system is capable of capturing the trend changes in financial indicators such as ROA and ROE, it still struggles to assess variations in more abstract market-based metrics like Tobin's Q. These metrics reflect investors' expectations and market perceptions, which are influenced by broader macroeconomic conditions and intangible factors such as market sentiment, brand value, or innovation potential—elements that are difficult to simulate purely through agent interactions.

- The current framework models agent behavior using general decision rules rather than incorporating more nuanced individual traits such as personality, risk preference, or strategic style—particularly for CEO agents. However, adding such complexity would lead to significantly increased modeling costs and longer simulation times, which may not be practical at scale.

- Our framework is intended for research purposes and may not capture all real-world variables affecting M&A outcomes. If used beyond its intended scope, such as in high-stakes financial decision-making without expert validation, the system's recommendations could be misinterpreted or overtrusted, leading to suboptimal decisions.

References

George Alexandridis, Nikolaos Antypas, and Vicky Y Lee. 2024. Do boutique investment banks have the midas touch? evidence from m&as. *European Financial Management*, 30(1):634–672.

Paul Borochin, Chinmoy Ghosh, and Di Huang. 2016. Is there an information asymmetry discount? new insights from mergers and acquisitions. Technical report, working paper, University of Connecticut, Storrs.

Xiaoxue Cheng, Junyi Li, Wayne Xin Zhao, Hongzhi Zhang, Fuzheng Zhang, Di Zhang, Kun Gai, and Ji-Rong Wen. 2024. Small agent can also rock! empowering small language models as hallucination detector. *Preprint*, arXiv:2406.11277.

Mai Anh Dao and Florian Bauer. 2021. Human integration following m&a: Synthesizing different m&a research streams. *Human Resource Management Review*, 31(3):100746.

Sorouralsadat Fatemi and Yuheng Hu. 2024. Finvision: A multi-agent framework for stock market prediction. *ArXiv*, abs/2411.08899.

Stephen P Ferris and Sushil Sainani. 2021. Do cfos matter? evidence from the m&a process. *Journal of Corporate Finance*, 67:101856.

Senyu Han, Lu Chen, Li-Min Lin, Zhengshan Xu, and Kai Yu. 2024. IBSEN: Director-actor agent collaboration for controllable and interactive drama script generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1607–1619, Bangkok, Thailand. Association for Computational Linguistics.

Zhitao He, Pengfei Cao, Chenhao Wang, Zhuoran Jin, Yubo Chen, Jiexin Xu, Huaijun Li, Xiaojian Jiang, Kang Liu, and Jun Zhao. 2024. Agentscourt: Building judicial decision-making agents with court debate simulation and legal knowledge augmentation. *Preprint*, arXiv:2403.02959.

Sirui Hong, Mingchen Zhuge, Jiaqi 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. 2024. Metagpt: Meta programming for a multi-agent collaborative framework. *Preprint*, arXiv:2308.00352.

Dong Huang, Qingwen Bu, Jie M. Zhang, Michael Luck, and Heming Cui. 2023. Agentcoder: Multi-agent-based code generation with iterative testing and optimisation. *ArXiv*, abs/2312.13010.

Md. Ashraful Islam, Mohammed Eunus Ali, and Md Rizwan Parvez. 2024. Mapcoder: Multi-agent code generation for competitive problem solving. *Preprint*, arXiv:2405.11403.

Md. Ashraful Islam, Mohammed Eunus Ali, and Md. Rizwan Parvez. 2025. Codesim: Multi-agent code generation and problem solving through simulation-driven planning and debugging. *ArXiv*, abs/2502.05664.

Michael C Jensen. 1986. Agency costs of free cash flow, corporate finance, and takeovers. *The American economic review*, 76(2):323–329.

Mingyu Jin, Qinkai Yu, Dong Shu, Haiyan Zhao, Wenyue Hua, Yanda Meng, Yongfeng Zhang, and Mengnan Du. 2024. The impact of reasoning step length on large language models. *Preprint*, arXiv:2401.04925.

Ziqi Jin and Wei Lu. 2023. Tab-cot: Zero-shot tabular chain of thought. In *Annual Meeting of the Association for Computational Linguistics*.

Mahendra Joshi, Carol Sanchez, and Paul Mudde. 2020. Improving the m&a success rate: Identity may be the key. *Journal of Business Strategy*, 41(1):50–57.

David R King, Dan R Dalton, Catherine M Daily, and Jeffrey G Covin. 2004. Meta-analyses of post-acquisition performance: Indications of unidentified moderators. *Strategic management journal*, 25(2):187–200.

Haohang Li, Yupeng Cao, Yangyang Yu, Shashidhar Reddy Javaji, Zhiyang Deng, Yueru He, Yuechen Jiang, Zining Zhu, Koduvayur Subbalakshmi, Guojun Xiong, Jimin Huang, Lingfei Qian, Xueqing Peng, Qianqian Xie, and Jordan W. Suchow. 2024a. Investorbench: A benchmark for financial decision-making tasks with llm-based agent. *ArXiv*, abs/2412.18174.

Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024b. Econagent: Large language model-empowered agents for simulating macroeconomic activities. *Preprint*, arXiv:2310.10436.

Yinheng Li, Shaofei Wang, Han Ding, and Hang Chen. 2023a. Large language models in finance: A survey. ICAIF '23, page 374–382, New York, NY, USA. Association for Computing Machinery.

Yinheng Li, Shaofei Wang, Han Ding, and Hang Chen. 2023b. Large language models in finance: A survey. In *Proceedings of the Fourth ACM International Conference on AI in Finance*, ICAIF '23, page 374–382, New York, NY, USA. Association for Computing Machinery.

Yunxuan Li, Yibing Du, Jiageng Zhang, Le Hou, Peter Grabowski, Yeqing Li, and Eugene Ie. 2024c. Improving multi-agent debate with sparse communication topology. *Preprint*, arXiv:2406.11776.

Zhiwei Li, Ran Song, Caihong Sun, Wei Xu, Zhengtao Yu, and Ji-Rong Wen. 2024d. Can large language models mine interpretable financial factors more effectively? a neural-symbolic factor mining agent model. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3891–3902, Bangkok, Thailand. Association for Computational Linguistics.

Yixuan Liang, Yuncong Liu, Boyu Zhang, Christina Dan Wang, and Hongyang Yang. 2024. Fingpt: Enhancing sentiment-based stock movement prediction with dissemination-aware and context-enriched llms. *Preprint*, arXiv:2412.10823.

Amarjeet Kaur Malhotra and Ajay Kumar Chauhan. 2018. Market response of mergers and acquisitions'announcement and its predictive ability in forecasting post merger long term performance of the acquiring company. *Academy of Accounting and Financial Studies Journal*, 22(2):1–16.

Ulrike Malmendier, Enrico Moretti, and Florian S Peters. 2018. Winning by losing: Evidence on the long-run effects of mergers. *The Review of Financial Studies*, 31(8):3212–3264.

Olimpia Meglio and Annette Risberg. 2011. The (mis) measurement of m&a performance—a systematic narrative literature review. *Scandinavian journal of management*, 27(4):418–433.

Sara B Moeller, Frederik P Schlingemann, and René M Stulz. 2005. Wealth destruction on a massive scale? a study of acquiring-firm returns in the recent merger wave. *The journal of finance*, 60(2):757–782.

Micah S. Officer, Annette B. Poulsen, and Mike Stegemoller. 2009. Target-firm information asymmetry and acquirer returns. *Review of Finance*, 13(3):467–493.

OpenAI. 2023. Chatgpt.

Himanshu Gautam Pandey, Akhil Amod, and Shivang Kumar. 2024. Advancing healthcare automation: Multi-agent system for medical necessity justification. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 39–49, Bangkok, Thailand. Association for Computational Linguistics.

Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2024. Chatdev: Communicative agents for software development. *Preprint*, arXiv:2307.07924.

Luc Renneboog and Cara Vansteenkiste. 2019. Failure and success in mergers and acquisitions. *Journal of Corporate Finance*, 58:650–699.

Jingqing Ruan, YiHong Chen, Bin Zhang, Zhiwei Xu, Tianpeng Bao, du qing, shi shiwei, Hangyu Mao, Xingyu Zeng, and Rui Zhao. 2023. TPTU: Task planning and tool usage of large language model-based AI agents. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*.

Sergio Servantez, Joe Barrow, Kristian Hammond, and Rajiv Jain. 2024. Chain of logic: Rule-based reasoning with large language models. *Preprint*, arXiv:2402.10400.

Stanislav Tarasov and Bartłomiej Dessoulavy-Śliwiński. 2024. Post-merger financial performance–a study of high-tech companies in the united states using artificial neural networks. *Metody Ilościowe w Badaniach Ekonomicznych*, 25(2):70–85.

Ephraim Kwashie Thompson and Changki Kim. 2020. Post-m&a performance and failure: Implications of time until deal completion. *Sustainability*, 12(7):2999.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models. *Preprint*, arXiv:2203.11171.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. Chain-of-thought prompting elicits reasoning in large language models. *Preprint*, arXiv:2201.11903.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. 2023a. Autogen: Enabling next-gen llm applications via multi-agent conversation. *Preprint*, arXiv:2308.08155.

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

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. 2025. Tradingagents: Multi-agents llm financial trading framework. *Preprint*, arXiv:2412.20138.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, Yijing Xu, Haoqiang Kang, Ziyan Kuang, Chenhan Yuan, Kailai Yang, Zheheng Luo, Tianlin Zhang, Zhiwei Liu, Guojun Xiong, and 15 others. 2024. Finben: A holistic financial benchmark for large language models. In *Advances in Neural Information Processing Systems*, volume 37, pages 95716–95743. Curran Associates, Inc.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. *Preprint*, arXiv:2210.03629.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Rong Liu, Zhenyu Cui, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. 2024. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Preprint*, arXiv:2407.06567.

Weikang Yuan, Junjie Cao, Zhuoren Jiang, Yangyang Kang, Jun Lin, Kaisong Song, tianqianjin lin, Pengwei Yan, Changlong Sun, and Xiaozhong Liu. 2024. Can large language models grasp legal theories? enhance legal reasoning with insights from multi-agent collaboration. *Preprint*, arXiv:2410.02507.

Haodong Zhang, Yanli Pu, Shuaiqi Zheng, and Lin Li. 2024. Ai-driven mamp;a target selection and synergy prediction: A machine learning-based approach. *Journal of Artificial Intelligence General science (JAIGS) ISSN:3006-4023*, 6(1):359–377.

Jun Zhao, Can Zu, Xu Hao, Yi Lu, Wei He, Yiwen Ding, Tao Gui, Qi Zhang, and Xuanjing Huang. 2024. LONGAGENT: Achieving question answering for 128k-token-long documents through multi-agent collaboration. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 16310–16324, Miami, Florida, USA. Association for Computational Linguistics.

Qihuang Zhong, Kang Wang, Ziyang Xu, Juhua Liu, Liang Ding, Bo Du, and Dacheng Tao. 2024. Achieving >97% on gsm8k: Deeply understanding the problems makes llms better reasoners. *ArXiv*, abs/2404.14963.

A Prompt

A.1 CFO Agent

System Prompt:

You are an experienced financial analyst acting on behalf of your company. The company is preparing for an M&A negotiation. Based on the financial data of your own company and the buyer or seller, along with the information about the deal target, please provide your recommendation. Keep your response as concise as possible.

User Prompt:

Please analyze the financial condition of the following companies and provide your recommendation:
Buyer Company Name: `{buyer_company_name}`
Financial Data: `{buyer_financial_data}`
Seller Company Name: `{seller_company_name}`
Financial Data: `{seller_financial_data}`
Target: `{target_details}`

A.2 CEO Agent

System Prompt:

You are an experienced M&A negotiation expert, acting on behalf of `{buyer_company_name}` or `{seller_company_name}`. Your task is to conduct negotiations with the opposing party to maximize your company's interests.

User Prompt:

The CFO has provided the following financial analysis:
`{cfo_report}`
Multiple negotiation rounds have taken place. The conversation history is provided below:
`{Round1}`
`{Round2}`

A.3 Board Member Agent

System Prompt:

You are an experienced corporate board member.

Method	ROA				ROE				Tobin's Q			
	R	MSE	MAE	NDCG	R	MSE	MAE	NDCG	R	MSE	MAE	NDCG
M&A Agent	**0.254**	**0.008**	**0.077**	**0.844**	**0.333**	0.041	0.147	**0.864**	**0.095**	**63.124**	**7.263**	**0.807**
w/o negotiation	0.036	0.015	0.106	0.648	0.065	0.042	0.158	0.592	0.126	5.894	9.292	0.791
w/o board	0.133	0.009	0.082	0.813	0.152	**0.033**	**0.124**	0.821	0.004	76.224	8.296	0.563
w/o regulatory	0.109	0.010	0.090	0.771	0.099	0.039	0.147	0.789	0.065	77.502	8.215	0.608

Table 3: Performance of ablation variants across three financial indicators.

Your role is to vote on the outcome of an M&A negotiation.

User Prompt:

You are `{member_name}`, a board member representing the `{company_side}`. Please vote (approve or reject) on the following negotiation outcome and explain your reasoning.

Negotiation Summary: `{negotiation_result}`

A.4 Evaluation Committee Agent

Proposal-Creation Agent

System Prompt:

You are `{member_name}`, a member of the evaluation committee and an experienced expert in M&A assessment. Based on the information provided, conduct a comprehensive evaluation of the transaction and assign an overall score (1–5) reflecting the potential for post-merger growth in metrics such as ROE, ROA, and Tobin's Q.

User Prompt:

Please conduct a comprehensive evaluation of the following M&A transaction. You may define your own evaluation criteria and scoring formula. Provide a total score from 1 to 5 (5 being the highest), along with detailed reasoning.

Buyer: `{buyer_name}`

Seller: `{seller_name}`

Negotiation Process: `{Negotiation Result}`

Board Voting Results: `{Boder Vote Result}`

Regulatory Review Result: `{review_content}`

Committee Vote History:

`{Committee Vote History}`

You may consider dimensions such as strategic fit, financial benefits, integration feasibility, and risk level, but you are free to choose your own factors and assign weights accordingly. Clearly explain your formula, weight choices, and scoring rationale.

Please output the result in JSON format including the following fields: `total_score`, `scoring_formula`, `weight_explanation`, and `evaluation_reason`.

Voting Agent

System Prompt:

You are a member of the evaluation committee. Your task is to vote on another member's proposed score for an M&A transaction, based on whether you believe it accurately reflects the buyer's potential for post-merger financial growth.

User Prompt:

Please vote to approve or reject the following evaluation proposal:

Buyer: `{buyer_name}`

Seller: `{seller_name}`

Proposed by: `{member_name}`

Total Score: `{total_score}`

Scoring Formula: `{scoring_formula}`

Weight Explanation: `{weight_explanation}`

Evaluation Reason: `{evaluation_reason}`

Please cast your vote (approve or reject) and provide a brief explanation.

B Data Deatail

The dataset contains both financial and event-level information. Financial variables include accounting period, total assets, total liabilities, total equity, operating revenue, net profit, R&D investment ratio, operating cash flow, investment cash flow, financing cash flow, debt-to-assets ratio, return on equity (ROE), and price-earnings ratio (P/E). Event-related variables comprise source of funding, merger type, restructuring type, regional classification (domestic or cross-regional), cross-province and cross-city indicators, related-party status, major restructuring status, and involvement of intellectual property. Additionally, the dataset includes company profiles such as business scope, all of which are recorded in Chinese.

C NDCG

The mapping between financial growth percentiles and NDCG scores is shown in table 4.

Mapped Score	**Percentile Range**
5.0	Top 12.5%
4.5	75.0–87.5%
4.0	62.5–75.0%
3.5	50.0–62.5%
3.0	37.5–50.0%
2.5	25.0–37.5%
2.0	12.5–25.0%
1.0	Bottom 12.5%

Table 4: Mapping of NDCG Evaluation Scores Based on Percentile Rankings of Financial Growth Rates.

D Ablation Detail

Detailed results are presented in Table 3.