Journal of Industrial Engineering and Applied Science
Journal Home: http://jieas.suaspress.org/ | CODEN: JIEAAE
Vol. 3, No. 6, 2025 | ISSN 3005-6071 (Print) | ISSN 3005-608X (Online)

Development of AI Multi-Agent Frameworks for Financial Services

LI, Chen [1*]

[1] The University of Texas at Dallas, USA

**** LI, Chen is the corresponding author, E-mail: chenli.stephen@gmail.com***

Abstract: Large-language-model-based multi-agent architectures and distributed AI components are rapidly reshaping financial services. They enable autonomous decision-making, collaboration in problem-solving, and automation of complex workflows across risk management, trading, compliance, fraud detection, and customer interaction. However, existing frameworks face significant tensions between scalability, regulatory requirements, and real-time performance in highly dynamic markets. This paper revisits the current landscape of AI multi-agent frameworks for finance and proposes a refined perspective emphasizing framework architecture, quantitative evaluation, and governance. We briefly reference related developments in small-sample prompt-based classification and hybrid edge–cloud frameworks.

Keywords: Large Language Models (LLMs), Multi-agent Architectures, Financial Services, Hybrid Edge–cloud Frameworks.

Disciplines: Artificial Intelligence Technology. **Subjects:** Machine Learning.

DOI: https://doi.org/10.70393/6a69656173.333337 **ARK:** https://n2t.net/ark:/40704/JIEAS.v3n6a01

1 INTRODUCTION

The evolution of large language models (LLMs) and agentic AI has enabled systems capable of reasoning and interacting with complex financial environments. Multi-agent systems combine specialized agents into coordinated workflows that can enhance efficiency in investment analysis, risk management, and fraud detection[1]. Yet, their deployment must address regulatory constraints, transparency, and computational limitations. The combination of multi-agent reasoning and hybrid infrastructure has shown promise across domains including finance and mobile AI engineering.

2 LITERATURE REVIEW AND RELATED WORK

In recent years, the financial technology landscape has undergone a significant transformation driven by advances in artificial intelligence and distributed decision-making frameworks. Traditional monolithic models are increasingly replaced by multi-agent architectures, which offer modularity, adaptability, and resilience under dynamic market conditions. By enabling autonomous yet cooperative behaviors among intelligent agents, these architectures facilitate complex analytical workflows, real-time prediction, and continuous optimization across various financial domains.

The following subsections provide an overview of representative frameworks, domain-specific implementations, and hybrid edge–cloud extensions that collectively define the emerging paradigm of AI-driven financial analytics[2].

2.1 Overview of Multi-Agent Systems in Financial Analytics

Recent literature highlights a growing trend toward employing multi-agent systems (MAS) to address the complexity and dynamism of financial analytics. These systems enable distributed intelligence where autonomous agents collaborate to execute tasks such as market prediction, portfolio management, and risk monitoring. Contemporary frameworks—including LangChain, CrewAI, and Semantic Kernel—facilitate coordination among heterogeneous agents through standardized APIs, modular toolkits, and context-sharing protocols. Their extensible designs allow seamless integration with large language models (LLMs), external databases, and task-specific APIs, supporting both analytical reasoning and decision-making in financial environments.

2.2 Domain-Specific Implementations

Beyond general-purpose frameworks, several domain-specific MAS architectures have emerged to meet the stringent demands of financial operations. Systems such as FinRobot and Fincon exemplify this trend by embedding financial logic and compliance-aware workflows into agent coordination mechanisms. These frameworks specialize in trading automation, portfolio optimization, and risk-adjusted

Journal of Industrial Engineering and Applied Science
Journal Home: http://jieas.suaspress.org/ | CODEN: JIEAAE
Vol. 3, No. 6, 2025 | ISSN 3005-6071 (Print) | ISSN 3005-608X (Online)

asset allocation, where multi-agent collaboration ensures scalability and real-time responsiveness. Their designs emphasize not only the interoperability between analytical and execution agents but also the ability to adapt to market volatility and regulatory constraints.

2.3 Hybrid Edge–Cloud Architectures

Emerging hybrid systems, represented by SolidGPT, extend MAS capabilities into high-security and latency-sensitive domains through adaptive routing between edge and cloud environments. SolidGPT introduces a hierarchical control structure in which lightweight agents operate on the edge for time-critical inference, while computationally intensive tasks—such as reinforcement learning, financial forecasting, or multi-factor correlation analysis — are offloaded to the cloud[2]. This architecture enhances both data privacy and operational resilience, making it suitable for applications such as fraud detection, algorithmic trading, and compliance auditing within regulated financial institutions[3].

2.4 Architectural Composition

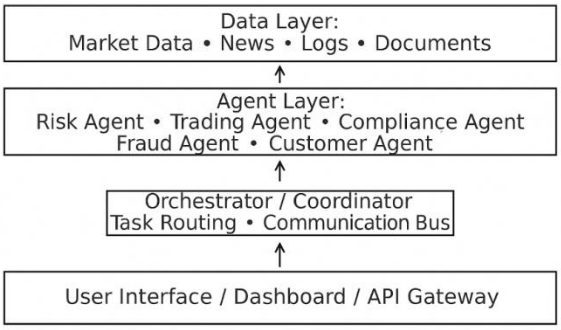

Figure 1. AI Multi-Agent Architecture for Financial Services

Figure 1 illustrates the proposed AI Multi-Agent Architecture for Financial Services, which is organized into four distinct layers:

Data Layer: Responsible for acquiring, cleansing, and normalizing structured and unstructured financial data from multiple sources, including market feeds, transactional logs, and regulatory repositories.

Agent Layer: Comprises specialized agents for prediction, reasoning, trading, and evaluation. Each agent possesses a defined capability and communicates via a shared protocol, enabling modular and explainable task execution.

Central Orchestrator: Functions as the coordination core, managing agent communication, workload distribution, and adaptive routing between local and cloud environments based on task criticality and latency requirements.

User Interface Layer: Provides a secure, interactive environment for human analysts, offering visual analytics dashboards, simulation controls, and model interpretability features.

This layered organization underscores the framework's scalability, adaptability, and alignment with the evolving needs of financial institutions operating under real-time constraints and stringent compliance standards.

3 AI MULTI-AGENT FRAMEWORKS IN FINANCIAL SYSTEMS

The rapid evolution of artificial intelligence and distributed computing has led to the emergence of multi-agent frameworks as a core paradigm in financial technology. These frameworks integrate multiple autonomous agents that cooperate to analyze markets, execute trades, manage risks, and ensure regulatory compliance.

3.1 Application Areas of AI Multi-Agent Frameworks in Finance

Applications of AI multi-agent systems in financial services are expanding rapidly across both front-office and back-office operations. Key domains include investment analysis, portfolio management, risk assessment, fraud detection, and compliance automation. Each domain benefits from the decentralized intelligence and adaptive coordination enabled by multi-agent frameworks.

For instance, in investment analysis, predictive agents analyze multi-modal data such as market sentiment, price trends, and macroeconomic indicators to generate real-time forecasts. Portfolio management agents then use these insights to rebalance asset allocations based on dynamically evolving market conditions. In parallel, risk assessment agents continuously evaluate exposure profiles, liquidity risks, and leverage ratios, while compliance agents ensure adherence to jurisdiction-specific regulations through rule-based validation and anomaly detection.

The distributed nature of multi-agent teams enables domain specialization and cooperative reasoning. Each agent is optimized for a subset of market indicators or operational constraints — such as volatility tracking, regulatory monitoring, or transaction verification — allowing for a heterogeneous yet coordinated intelligence network. This modular configuration closely mirrors the human-in-the-loop collaboration strategies demonstrated in SolidGPT, where human experts interact with autonomous agents through structured feedback loops[2]. Such coordination ensures that algorithmic decisions remain transparent, auditable, and aligned with organizational objectives.

In addition, the integration of hybrid edge – cloud infrastructures has further strengthened the applicability of multi-agent frameworks in financial contexts. Edge nodes located within financial institutions or trading platforms can perform low-latency inference and sensitive data processing locally, while high-capacity cloud nodes handle computationally demanding workloads such as deep learning

Journal of Industrial Engineering and Applied Science
Journal Home: http://jieas.suaspress.org/ | CODEN: JIEAAE
Vol. 3, No. 6, 2025 | ISSN 3005-6071 (Print) | ISSN 3005-608X (Online)

SUAS Press

retraining or long-horizon simulations.

As shown in Figure 2, a Markov Decision Process (MDP)-based router serves as the adaptive control mechanism that dynamically allocates computational tasks between edge and cloud environments. The router evaluates contextual parameters — including latency, energy budget, data confidentiality, and accuracy requirements — to determine the optimal execution path. This context-aware task routing enables financial AI systems to maintain a fine balance between performance efficiency, data privacy, and operational cost.

By leveraging this hybrid design, financial institutions can achieve real-time responsiveness at the edge without compromising the scalability and learning capacity of cloud intelligence, marking a significant advancement in the evolution of distributed financial analytics.

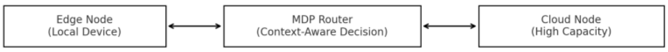

FIGURE 2. HYBRID EDGE–CLOUD ROUTING MECHANISM

An MDP-based router dynamically allocates computational tasks between edge and cloud nodes based on runtime constraints such as latency, energy budget, security sensitivity, and accuracy thresholds.

3.2 ARCHITECTURES AND COMPARATIVE ANALYSIS

Typical financial multi-agent frameworks are structured into multiple interdependent layers — data ingestion, agent reasoning, orchestration, and user interaction—each serving a distinct functional role in the overall system pipeline.
The data ingestion layer is responsible for continuously collecting, cleansing, and transforming high-frequency market data, transactional records, and regulatory feeds into standardized formats suitable for analytical processing. This foundation ensures that downstream agents operate on reliable and synchronized information streams.

The agent reasoning layer contains specialized agents trained for distinct tasks such as risk evaluation, trading strategy generation, and compliance verification. Each agent functions autonomously within its domain, leveraging both historical and real-time data to produce task-specific outputs. The modular design of this layer allows new agents to be introduced or deprecated dynamically without disrupting the entire system, thereby improving scalability and maintainability.

At the core lies the orchestration layer, which governs inter-agent communication, task allocation, and adaptive workload scheduling. The orchestrator serves as a coordination hub, ensuring that each agent's inference and decision-making align with system-level goals. It enables cross-agent context sharing—where insights from one domain (e.g., risk analysis) inform actions in another (e.g., trading or compliance)—thus fostering synergistic reasoning.
The orchestration layer also manages adaptive routing strategies, a defining feature of hybrid financial AI frameworks. Through intelligent load balancing across edge and cloud resources, the system dynamically adjusts computational routes based on factors such as latency, energy consumption, and security requirements. This adaptive approach ensures both high performance and regulatory compliance, a capability inspired by hybrid AI systems like SolidGPT and other edge – cloud coordination architectures[2,3,4].

The user interaction layer integrates visual analytics dashboards, decision audit trails, and feedback interfaces for human analysts. It provides transparency by visualizing how multiple agents collaborate to reach a decision, thereby enhancing interpretability and human oversight in mission-critical financial contexts.

To evaluate system-wide behavior, financial multi-agent frameworks rely on a set of quantitative performance indicators, as illustrated in Figure 3.

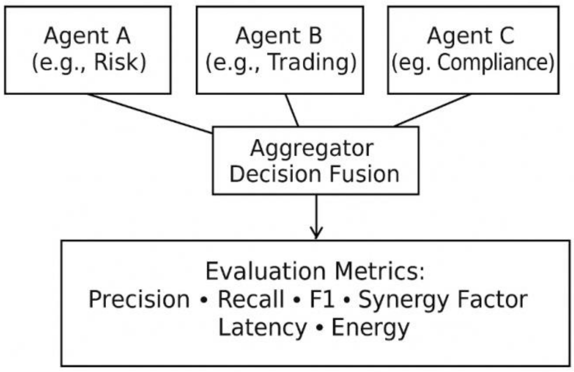

FIGURE 3. AGENT COLLABORATION AND EVALUATION METRICS

Multiple specialized agents (e.g., risk, trading, compliance) feed their outputs into an aggregator that performs decision fusion. The aggregated results are then assessed through standard classification metrics such as precision, recall, and F1-score, alongside composite measures like the synergy factor, which quantifies cooperative performance gains. Additional operational metrics—including latency and energy efficiency—capture the real-time and sustainability aspects of financial AI systems. Together, these indicators provide a comprehensive evaluation of both individual agent performance and emergent system-level intelligence, forming the analytical foundation for comparative assessment across architectures.

3.3 QUANTITATIVE EVALUATION OF FINANCIAL AI AGENTS

Evaluation metrics for AI multi-agent frameworks include precision, recall, F1-score, and synergy factors

Journal of Industrial Engineering and Applied Science
Journal Home: http://jieas.suaspress.org/ | CODEN: JIEAAE
Vol. 3, No. 6, 2025 | ISSN 3005-6071 (Print) | ISSN 3005-608X (Online)

representing collective agent performance. The synergy factor measures the gain from agent cooperation relative to individual results. Additional indicators include latency, energy usage, and compliance accuracy, metrics that have been validated in hybrid AI experiments.

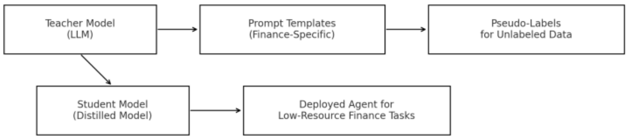

FIGURE 4. PROMPT-DISTILLATION INTEGRATION FOR LOW-RESOURCE FINANCE TASKS

A large teacher model and finance-specific prompts generate pseudo-labels, which are distilled into a smaller student model and deployed in data-scarce environments[4,5].

4 CHALLENGES AND FUTURE DIRECTIONS

Despite the growing maturity of AI multi-agent frameworks in financial systems, several technical and governance challenges remain before large-scale and trustworthy deployment can be achieved. The most critical issues involve data quality, model interpretability, and regulatory compliance, all of which directly affect the reliability and transparency of financial decision-making[5][6]. Financial data often contain noise, inconsistencies, and biases arising from heterogeneous sources such as trading logs, customer records, and third-party feeds. Maintaining data integrity and ensuring proper lineage tracking are thus prerequisites for any multi-agent analytics system[7].

Another major challenge concerns interpretability and explainability. Financial institutions operate in tightly regulated environments where decisions must be auditable and understandable to human supervisors. However, multi-agent systems—especially those built on deep learning and large language models — tend to behave as black boxes, making it difficult to trace causal reasoning or justify collective outcomes. Developing explainable coordination mechanisms, such as agent-level transparency modules and interpretable policy routing, will be key to fostering trust in these systems.

Furthermore, the adversarial robustness of financial AI remains an open problem. Agents exposed to manipulated inputs, spoofed transactions, or coordinated attacks may propagate erroneous behaviors across the system, leading to cascading risks. Designing robust training pipelines, anomaly detectors, and consensus-based defense protocols is therefore essential to safeguarding multi-agent infrastructures from both internal errors and external threats[8].

From a methodological perspective, the adaptability of AI agents under limited or shifting data conditions continues to pose difficulties. In many financial contexts—especially in emerging markets or specialized asset classes—labeled data are scarce or unevenly distributed. Techniques such as prompt distillation, context fusion, and few-shot adaptation, inspired by recent advances in small-sample natural language processing, offer promising avenues for improving generalization in low-resource environments[4]. These methods allow smaller agents to inherit domain-specific reasoning skills from larger foundation models, enabling continuous learning with minimal data overhead.

Finally, future research should explore human-in-the-loop governance models that balance automation with expert oversight. Integrating human expertise into agent orchestration, as demonstrated by hybrid frameworks like SolidGPT, can enhance both accountability and adaptability [2,8]. Over the next decade, the convergence of edge–cloud hybridization, interpretable AI, and regulatory-aware orchestration is expected to define the next generation of intelligent, trustworthy financial multi-agent ecosystems [9-11].

5 CONCLUSION

AI multi-agent frameworks are fundamentally transforming financial operations by enabling distributed reasoning, adaptive coordination, and real-time automation across diverse analytical and transactional processes. Through collaborative agent architectures, financial institutions can dynamically integrate predictive modeling, risk control, and regulatory compliance within a unified decision-making ecosystem.

Moreover, hybrid approaches that integrate edge and cloud computing provide an effective balance between data privacy, computational performance, and scalability, ensuring that latency-sensitive tasks are processed locally while resource-intensive analytics are executed in the cloud. Drawing insights from SolidGPT and prompt-based data-efficient models, future frameworks are expected to achieve greater interpretability, resilience, and adaptability in mission-critical financial environments. These developments point toward a new generation of transparent, context-aware, and self-optimizing financial AI ecosystems capable of maintaining regulatory alignment while continuously learning from real-world market dynamics.

ACKNOWLEDGMENTS

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

LI, Chen

The University of Texas at Dallas, USA.

REFERENCES

[1] Qi, R., & Hu, L. (2025). Stock Price Prediction of Apple Inc. Based on LSTM Model: An Application of Artificial Intelligence in Individual Stock Analysis. European Journal of AI, Computing & Informatics, 1(3), 1-9.

[2] Hu, L., Wu, Q., & Qi, R. (2025). Empowering smart app development with SolidGPT: an edge–cloud hybrid AI agent framework. Advances in Engineering Innovation, 16(7), 86-92.

[3] Hu, L. (2025). Hybrid Edge-AI Framework for Intelligent Mobile Applications: Leveraging Large Language Models for On-device Contextual Assistance and Code-Aware Automation. Journal of Industrial Engineering and Applied Science, 3(3), 10-22.

[4] Hu, L. (2025). Topic Classification of Small Sample News Based on Prompt Engineering. Applied and Computational Engineering, 170, 101-107.

[5] Qi, R. (2025). Financial News Sentiment Analysis and Market Sentiment Prediction Based on Large Language Models. Journal of Computer, Signal, and System Research, 2(6), 22-31.

[6] Qi, R. (2025). An Empirical Study on Credit Risk Assessment Using Machine Learning: Evidence from the Kaggle Credit Card Fraud Detection Dataset. Journal of Computer, Signal, and System Research, 2(5), 48-64.

[7] Qi, R. (2025). Loan Default Prediction and Feature Importance Analysis Based on the XGBoost Model. European Journal of Business, Economics & Management, 1(2), 141-149.

[8] Joshi, S. (2025). Architectures and Challenges of AI Multi-Agent Frameworks for Financial Services. Current Journal of Applied Science and Technology, Vol. 44, Issue 6.

[9] Ante, L. (2024). Autonomous AI Agents in Decentralized Finance: Market Dynamics, Application Areas, and Theoretical Implications. Application Areas, and Theoretical Implications (December 14, 2024).

[10] Hettiarachchi, I. (2025). THE RISE OF GENERATIVE AI AGENTS IN FINANCE: OPERATIONAL DISRUPTION AND STRATEGIC EVOLUTION. International Journal of Engineering Technology Research & Management, 447.

[11] Agnihotram, G., Sarkar, J., Kasthuri, M., & Balasubramaniam, V. (2025). Agentic AI for Risk Assessment Controllers in BFSI: A Technical Framework for Autonomous Risk Mitigation. Journal of Banking and Financial Dynamics, 9(10), 1-8.