International Journal of Advanced Engineering, Management and Science (IJAEMS)
Peer-Reviewed Journal
ISSN: 2454-1311 | Vol-12, Issue-2; Mar-Apr, 2026
Journal Home Page: https://ijaems.com/
DOI: https://dx.doi.org/10.22161/ijaems.122.8

Policy Convergence and Explainable Stability in Orchestrated Multi-Agent LLM Trading Frameworks

Shourya Dewansh

Student, Wentworth Institute of Technology, Boston, MA

Received: 05 Feb 2026; Received in revised form: 06 Mar 2026; Accepted:09 Mar 2026; Available online: 14 Mar 2026

Abstract *— This article examines orchestrated multi-agent large language model trading frameworks through the paired lenses of policy convergence and explainable stability. The topic has gained urgency because recent financial agent systems increasingly rely on distributed deliberation, memory, critique, and risk filtering, while evaluation practice still privileges return metrics over decision coherence and rationale persistence. The aim is to formulate an analytical framework for assessing whether specialized agents move toward a stable executable policy and whether the accompanying explanation remains consistent under workflow pressure, market variability, and iterative revision. The study draws on recent work in explainable AI for finance, explainable reinforcement learning, multi-agent reinforcement learning, quantitative trading, and LLM-based financial agents. Comparative source analysis, conceptual structuring, analytical synthesis, and cross-framework interpretation were applied. The analytical section identifies four determinants of stability: evidence discipline, deliberative topology, risk-gating logic, and adaptive memory control. The resulting framework supports research design, model auditing, and architecture selection for high-stakes AI trading systems.*

Keywords *— **multi-agent LLMs, algorithmic trading, policy convergence, explainable stability, financial AI, agent orchestration, explainable reinforcement learning, trading system evaluation, risk-aware automation, decision traceability***

I. INTRODUCTION

The current stage of financial AI research is marked by a visible shift from isolated predictive models toward coordinated agent assemblies that divide analytical labor across market interpretation, risk screening, memory handling, and decision synthesis. That shift has intensified a methodological problem: performance figures alone do not reveal whether a trading policy emerges from disciplined inter-agent agreement or from unstable narrative negotiation. In finance, where execution errors are magnified by volatility, leverage, and latency, a system's internal coherence matters alongside profitability. Recent surveys on financial explainability and reinforcement learning point to the same unresolved tension: high-capacity models often produce strong outputs while leaving decision pathways insufficiently transparent for audit, intervention, or post-trade diagnosis.

The purpose of the article is to develop an analytical model for evaluating policy convergence and explainable stability in orchestrated multi-agent LLM trading frameworks. Three tasks structure the study:

1) to differentiate policy convergence from raw prediction agreement and links it to executable coordination under market constraints.
2) to identify the architectural conditions under which explanation traces remain stable across

debate, revision, memory updates, and risk control.

3) to formulate an evaluation scheme suitable for comparing recent trading-oriented multi-agent systems without reducing assessment to return statistics.

The novelty lies in joining two lines of inquiry that are often treated separately: multi-agent coordination on one side and explainable financial AI on the other. The article argues that stable execution in LLM trading systems depends less on the presence of many specialized agents than on the disciplined relation among evidence intake, deliberation order, gating rules, and adaptive revision.

II. MATERIALS AND METHODS

The source base combines ten recent publications that jointly cover explainability, reinforcement learning, financial decision systems, and orchestrated LLM agents. P.-D. Arsenault, S. Wang, and J.-M. Patenaude [1] systematize explainable AI approaches for financial time-series forecasting and distinguish interpretability from explanation design. H.S. Jung and H. Lee [2] present a zero-shot multi-agent trading architecture with explicit rationale generation and backtested interpretability checks. X. Li, Y. Zeng, X. Xing, J. Xu, and X. Xu [3] analyze a simulated-trading multi-agent system in which forward-looking evaluation and meeting-based coordination structure decision formation. Z. Ning and L. Xie [4] provide a broad survey of multi-agent reinforcement learning, focusing on coordination problems, benchmark environments, and implementation burdens. L. Saulières [5] reviews explainable reinforcement learning and proposes a taxonomy centered on explanation targets and delivery modes. S. Sun, R. Wang, and B. An [6] synthesizes reinforcement learning research for quantitative trading and clarifies the relation among objectives, environments, and action spaces. C.-A. Wang, S.-H. Huang, C.-T. Chen, and Y.-T. Fang [7] studied reinforcement learning in finance under concept drift and showed why stability cannot be detached from environmental regime change. Y. Xiao, E. Sun, D. Luo, and W. Wang [8] introduce a trading-firm-like multi-agent architecture with analyst debate, risk management, and manager approval. W.J. Yeo, W. Van Der Heever, R. Mao, E. Cambria, R. Satapathy, and G. Mengaldo [9] review explainable AI for finance from the perspectives of transparency requirements, method families, and sector-specific constraints. Y. Yu, Z. Yao, H. Li, Z. Deng, Y. Cao, Z. Chen, J.W. Suchow, R. Liu, Z. Cui, Z. Xu, D. Zhang, K. Subbalakshmi, G. Xiong, Y. He, J. Huang, D. Li, and Q. Xie [10] formulate an LLM multi-agent financial framework with conceptual verbal reinforcement and hierarchical communication.

The study applies comparative analysis, source analysis, conceptual structuring, analytical synthesis, and cross-framework interpretation. A comparative analysis was used to identify recurring coordination patterns across trading-oriented agent systems. Source analysis was used to track how recent literature defines explainability, reinforcement learning stability, and multi-agent decision structure. Conceptual structuring was used to derive the paired constructs of policy convergence and explainable stability. Analytical synthesis connected these constructs to financial execution logic, memory design, and risk-screening procedures. Cross-framework interpretation aligned architecture choices with likely stability outcomes in trading environments.

III. RESULTS

A review of recent multi-agent financial systems shows that policy convergence should not be reduced to simple agreement among agents. In trading architectures, convergence has a narrower and more operational meaning: divergent local interpretations must be transformed into a single executable policy that remains aligned with risk constraints, evidence quality, and market timing. A multi-agent system may display verbal consensus while still producing unstable trade selection if its agreement rests on weak evidence routing, excessive prompt sensitivity, or unfiltered memory carryover. For that reason, convergence is better understood as a property of the orchestration layer rather than of individual model outputs. The trading literature already points in this direction. Financial reinforcement learning surveys treat environment design, reward shaping, and action specification as structural determinants of agent behavior [6]. In

contrast, MARL surveys show that coordination quality depends on communication form, information sharing, and task decomposition [4]. In LLM-based trading systems, these determinants reappear in natural-language form through meetings, debates, supervisory agents, and memory-mediated revision [3; 8; 10].

Within this analytical frame, explainable stability refers to the persistence of a decision rationale across iterative refinement. A rationale is stable when the system preserves the same causal story about evidence, risk, and action after critique, summarization, and managerial synthesis. It becomes unstable when later stages alter the justification while preserving the final action, or preserve the justification while changing the action without a traceable reason. Financial explainability surveys show why this distinction matters. In high-risk domains, explanation does not function as decorative transparency; it serves auditability, trust calibration, and intervention support [1; 9]. Explainable reinforcement learning reaches a similar conclusion from another direction by distinguishing what is being explained from how that explanation is delivered [5]. Applied to trading, this means that an intelligible rationale must retain semantic continuity from analyst signals to final order logic. Recent work on explainable multi-agent trading [2] makes this issue explicit by evaluating rationale quality alongside market outcomes.

Recent agent frameworks permit a more precise decomposition of convergence mechanics. TradingAgents organizes analysts, adversarial researchers, traders, a risk management team, and a manager into a staged pipeline where evidence is first diversified, then contested, and only afterward converted into execution [8]. FinCon uses a manager–analyst hierarchy with conceptual, verbal reinforcement, meaning that prior reasoning episodes are transformed into reusable belief updates [10]. QuantAgents adds simulated trading and repeated meetings, shifting the system away from purely retrospective reflection toward anticipatory policy formation [3]. Jung and Lee's explainable zero-shot framework places a meta-agent above heterogeneous specialist agents, using rationale synthesis as the bridge between modalities and action [2]. Taken together, these systems suggest that convergence improves when three conditions are met: evidence streams remain functionally differentiated, contradiction is processed before action selection, and a dedicated gate compresses deliberation into a bounded policy. Where one of these conditions is absent, systems drift toward either parallel monologues or premature consensus.

The literature further indicates that convergence depends on the sequence in which disagreement is resolved. If debate begins before evidence normalization, agents argue from incomparable informational bases. If risk control enters too late, the system may optimize narrative plausibility rather than tradability. If memory updates occur before contradiction resolution, unstable beliefs harden into future prompts. In this respect, policy convergence in financial LLM systems resembles a constrained reduction process: heterogeneous signals are progressively compressed through specialist interpretation, adversarial testing, risk adjudication, and managerial synthesis. The value of orchestration lies in how it orders these reductions. QuantAgents formalizes this through meetings dedicated to market analysis, strategy development, and risk alerts [3]. TradingAgents formalizes it through specialized analyst and researcher teams, followed by approval from traders and managers [8]. FinCon formalizes this through hierarchical communication and verbal reinforcement that selectively propagate revised beliefs [10]. Stable policy formation emerges when this reduction path remains explicit and reversible.

A second determinant concerns the relation between explanation and memory. LLM trading systems rely heavily on summaries, reflections, and episodic records. Those mechanisms preserve strategic continuity, yet they introduce a latent instability: a compressed memory trace may survive longer than the evidence that initially justified it. Once that occurs, future decisions inherit an explanation artifact rather than a verified analytical premise. FinCon addresses this issue by selectively propagating conceptual belief updates [10]. QuantAgents addresses it through several memory types connected to meetings and tools [3]. From the standpoint of explainable reinforcement learning, such designs raise a direct question: Does memory store observations, evaluations, or decisions already interpreted by prior agents [5]? The answer matters

because explainable stability deteriorates when the system cannot separate factual retention from argumentative residue. A stable architecture, therefore, requires memory typing, update thresholds, and clear boundaries between raw evidence and post-debate abstraction.

A third determinant lies in risk-gating logic. Financial decision systems differ from generic multi-agent settings because convergence toward a policy is not sufficient; the policy must remain admissible under exposure limits, portfolio discipline, and regime sensitivity. TradingAgents places a risk management team before managerial approval [8]. FinCon embeds risk control into episodic self-critique and belief revision [10]. QuantAgents gives risk control its own analyst and meeting space [3]. This recurrent design choice suggests that stable convergence in trading is inseparable from veto structure. A framework without a dedicated gate may still converge, yet it converges toward an action language rather than an execution policy. In financial terms, the difference is substantial: one system ends with a persuasive narrative, while the other ends with a controllable order hypothesis. Explainable stability rises when the rationale for accepting or blocking a trade is recorded at the same level of granularity as the trade thesis itself.

Environmental variability adds another layer of instability. Trading systems do not operate in stationary settings; evidence distributions shift, sentiment regimes break, and action values decay. Reinforcement learning literature on quantitative trading has long treated non-stationarity as a structural problem [6]. Recent work on concept drift in financial RL sharpens this point by distinguishing between gradual and sudden drift and linking adaptation to sentiment-aware restructuring, curriculum learning, and knowledge distillation [7]. For orchestrated LLM trading frameworks, the implication is clear: a policy may appear convergent within one regime while remaining fragile across adjacent regimes. Explainable stability, therefore, requires more than coherent debate at a single decision point. It requires the explanation schema to be resilient to changing market distributions. If the system revises actions after regime change, the rationale for revision must remain legible; if it retains the prior action, the rationale for persistence must remain equally readable. Stability without adaptive transparency degenerates into rigid consistency.

These observations support an analytical distinction between local convergence and systemic convergence. Local convergence refers to agreement within one deliberation cycle. Systemic convergence refers to the architecture's repeated ability to transform heterogeneous evidence into bounded decisions without explanation collapse. The latter is the more demanding property and the one that matters for deployable trading systems. Surveys of explainable AI in finance warn that transparency mechanisms often remain external to the model's logic [1; 9]. Multi-agent financial systems partially remedy that weakness by exposing intermediate reasoning stages through agent interactions [2; 3; 8; 10]. Yet exposure alone does not ensure stability. A transcript of unstable reasoning remains unstable. What matters is whether the architecture constrains how evidence enters, how disagreement is handled, how risk vetoes are applied, and how memory is rewritten.

On this basis, policy convergence in orchestrated multi-agent LLM trading frameworks may be defined as the architecture's capacity to reduce heterogeneous market interpretations into a single risk-bounded, executable decision through explicit, reviewable coordination steps. Explainable stability may be defined as the persistence of a coherent, auditable rationale across those coordination steps and across subsequent adaptation episodes. The conjunction of these two properties forms the most defensible analytical criterion for contemporary LLM trading systems. A framework lacking convergence remains operationally noisy; a framework lacking explainable stability remains epistemically fragile.

To consolidate the analytical observations discussed above, Figure 1 presents the proposed model of policy convergence and explainable stability in orchestrated multi-agent LLM trading frameworks. The scheme captures the ordered transition from heterogeneous market inputs to a bounded policy proposal and then to post-decision monitoring. Its logic emphasizes that stability is produced through a structured sequence rather than through isolated agent competence. In this sequence, specialized interpretation, contradiction processing, risk filtering, and explanation tracing form an integrated chain in

which each stage constrains the next one and preserves the auditability of the final decision.

Fig. 1. Analytical model of policy convergence and explainable stability in orchestrated multi-agent LLM trading frameworks (adapted from [3; 8; 10])

The model shows that convergence becomes analytically meaningful only when the final policy remains connected to the reasoning path that produced it and when subsequent revisions under regime change remain traceable. Viewed in this way, the architecture of the system determines whether coordination yields a governable trading policy or merely a temporary narrative agreement. This interpretation prepares the transition to the discussion section, where the comparative implications of recent frameworks are examined with respect to coordination design, explanation integrity, and sources of structural instability.

IV. DISCUSSION

The analytical reconstruction above suggests that recent multi-agent financial systems are converging toward a shared architectural grammar, though they still differ in how they manage disagreement, memory, and adaptation. The comparison in Table 1 condenses those differences into the dimensions most relevant for the present study: coordination form, convergence mechanism, explanation strategy, and the primary source of instability. Table 1 is derived from the ten-source corpus and focuses on architecture-level implications rather than reported benchmark superiority.

Table 1. Comparative architecture patterns related to convergence and explainable stability (synthesized from [1–10])

Primary coordination form	Main convergence mechanism	Explanation mechanism	Likely instability source
Meta-agent synthesis across modality specialists	Central integration of heterogeneous signals	Natural-language rationale generated by the meta-agent	Explanation drift between modality-level evidence and final synthesis
Meeting-based coordination with analyst specialization	Repeated deliberation plus simulated trading feedback	Explicit meeting structure and multi-memory workflow	Memory contamination across meetings; unstable transfer from simulation to execution
Trading-firm hierarchy with analyst/researcher/trader/manager chain	Adversarial research followed by risk filtering and approval	Intermediate team outputs expose deliberation stages	Premature consensus if adversarial evidence is weakly grounded
Manager–analyst hierarchy with	Selective propagation of revised beliefs	Verbal reinforcement preserves belief	Overcompression of belief updates into future prompts

conceptual verbal reinforcement		updates in textual form	
RL-oriented financial decision systems	Reward discipline and adaptation under market change	Post hoc interpretation of policy behavior	Regime drift and reward-policy mismatch
Explainability and XRL taxonomies	No direct execution mechanism; conceptual support for evaluation	Formal separation of targets, methods, and reliability of explanations	External explanations detached from internal decision logic

Table 1 shows that convergence and explainable stability are related, but they do not collapse into a single property. Architectures with stronger supervisory compression tend to produce clearer final policies, though they risk hiding unresolved disagreement inside summary layers. Architectures with richer deliberation expose more of the reasoning path, though they may preserve conflict without resolving it into an admissible order. Survey-based sources [1; 5; 9] help explain why both tendencies remain incomplete: explanation quality depends on alignment between what is described and where that information enters the decision process. When a trading framework generates narratives after policy selection, transparency remains cosmetic. When explanation artifacts participate in memory and veto logic, transparency becomes operational.

A second issue concerns evaluation design. Existing trading studies still privilege financial output metrics, whereas the present analysis argues for a broader matrix centered on decision integrity. Table 2 translates this argument into an applied evaluation scheme suitable for architecture review, internal audit, or pre-deployment stress testing. Each criterion targets a failure point that emerges repeatedly across recent studies: unstable evidence routing, summary distortion, unmanaged regime change, and untraceable veto logic.

Table 2. Proposed evaluation matrix for analytical assessment of orchestrated LLM trading systems (analytically derived from [1–10])

Evaluation dimension	What should be inspected	Failure signature	Practical implication
Evidence discipline	Traceability from source signal to agent claim	Claims survive after source inconsistency is detected	Weak auditability and inflated confidence
Deliberative convergence	Reduction of disagreement across coordination rounds	Final action appears without a resolved contradiction	Action selection rests on narrative compression rather than agreement
Risk-gating integrity	Visibility of vetoes, overrides, and exposure checks	Trade approval lacks explicit risk justification	Execution remains difficult to govern in regulated settings
Memory hygiene	Separation of raw evidence, summaries, and learned beliefs	Old summaries override current evidence	Historical residue distorts present decisions
Regime resilience	Stability of rationale under drift and market shifts	The same explanation persists despite a clear regime break	False consistency hides adaptation failure
Explanation continuity	Semantic alignment between analyst-level reasoning and final order logic	Final rationale rewrites earlier evidence without justification	Post-trade analysis loses causal coherence

The matrix in Table 2 clarifies the main limitation of a purely analytical article: without controlled experiments, it cannot rank architectures by measurable superiority. Yet such a format remains productive because recent literature already reveals repeated structural tensions that merit conceptual consolidation. The absence of a unified vocabulary has slowed rigorous comparison between LLM agent frameworks and reinforcement-learning-based trading systems. The paired constructs proposed here address that gap by linking execution discipline to explanation persistence. In the professional field of AI-driven trading, the practical implication is straightforward. An architecture review should examine where agreement is produced, how it is bounded, what kind of memory is retained, and whether explanations survive adaptation pressure. Systems that satisfy these conditions are better positioned for internal governance, model risk review, and human override design.

V. CONCLUSION

The analysis showed that convergence in multi-agent LLM trading frameworks refers to reducing heterogeneous interpretations into a single risk-bounded, executable decision through explicit coordination steps. The article found four such conditions: disciplined evidence routing, ordered contradiction processing, explicit risk gating, and controlled memory revision. In response, the article proposed an analytical matrix that integrates convergence, explanation continuity, regime resilience, and auditability within a single assessment framework. The main inference is that deployable trading intelligence depends on the joint presence of convergent policy formation and stable explanation traces. If one property is absent, the system either remains operationally noisy or becomes difficult to govern after deployment.

REFERENCES

[1] Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2025). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. *ACM Computing Surveys, 57*(10), Article 265, 1–37. https://doi.org/10.1145/3729531

[2] Jung, H. S., & Lee, H. (2026). Explainable zero-shot trading using multi-agent LLM architecture: A backtested approach for Bitcoin price. *Information Processing & Management, 63*(2, Part B), 104466. https://doi.org/10.1016/j.ipm.2025.104466

[3] Li, X., Zeng, Y., Xing, X., Xu, J., & Xu, X. (2025). QuantAgents: Towards multi-agent financial system via simulated trading. In *Findings of the Association for Computational Linguistics: EMNLP 2025* (pp. 17438–17464). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-emnlp.945

[4] Ning, Z., & Xie, L. (2024). A survey on multi-agent reinforcement learning and its application. *Journal of Automation and Intelligence, 3*(2), 73–91. https://doi.org/10.1016/j.jai.2024.02.003

[5] Saulières, L. (2025). *A survey of explainable reinforcement learning: Targets, methods and needs*. arXiv. https://doi.org/10.48550/arXiv.2507.12599

[6] Sun, S., Wang, R., & An, B. (2023). Reinforcement learning for quantitative trading. *ACM Transactions on Intelligent Systems and Technology, 14*(3), Article 44, 1–29. https://doi.org/10.1145/3582560

[7] Wang, C.-A., Huang, S.-H., Chen, C.-T., & Fang, Y.-T. (2026). Financial reinforcement learning under concept drift based on knowledge distillation and curriculum learning. *Decision Support Systems, 203*, 114624. https://doi.org/10.1016/j.dss.2026.114624

[8] Xiao, Y., Sun, E., Luo, D., & Wang, W. (2024). *TradingAgents: Multi-agents LLM financial trading framework*. arXiv. https://doi.org/10.48550/arXiv.2412.20138

[9] Yeo, W. J., Van Der Heever, W., Mao, R., et al. (2025). A comprehensive review on financial explainable AI. *Artificial Intelligence Review, 58*, Article 189. https://doi.org/10.1007/s10462-024-11077-7

[10] Yu, Y., Yao, Z., Li, H., Deng, Z., Cao, Y., Chen, Z., Suchow, J. W., Liu, R., Cui, Z., Xu, Z., Zhang, D., Subbalakshmi, K., Xiong, G., He, Y., Huang, J., Li, D., & Xie, Q. (2024). FinCon: A synthesized LLM multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Advances in Neural Information Processing Systems, 37*, 137010–137045. https://doi.org/10.48550/arXiv.2407.06567