INTERNATIONAL MULTI DISCIPLINARY JOURNAL FOR RESEARCH & DEVELOPMENT

eISSN 2394-6334 Impact factor: 7,854 Volume 12, issue 11 (2026)

The Synthetical Frontier of Agentic Autonomy: A Comprehensive Analysis of Generative AI, Multi-Agent Systems, and Interpretability in Modern Financial Ecosystems

Shrajika Whitemore
Department of Computational Economics, University of Edinburgh, United Kingdom

Abstract: The rapid convergence of Large Language Models (LLMs), generative artificial intelligence, and autonomous agent-based modeling has precipitated a paradigm shift in the global financial sector. This research article explores the evolution from traditional machine learning applications in finance to the current era of "agentic" autonomy. By synthesizing foundational theories of deep learning, credit risk analysis, and game theory with contemporary advancements in generative agents, this study evaluates the technical, ethical, and regulatory dimensions of self-driven AI. We examine the transition from "black box" deep portfolios to transparent, interpretable models, arguing that the future of financial stability depends on the balance between computational efficiency and human-centric explainability. The methodology employs a longitudinal theoretical synthesis and a meta-analytical review of multi-agent financial networks (MAFN). Findings suggest that while generative agents offer unprecedented personalization in financial advice and systemic risk modeling, they introduce novel risks related to algorithmic bias and narrative-driven market volatility. The discussion emphasizes the necessity of robust governance frameworks, as proposed in recent international AI acts, to mitigate the risks of autonomous financial agents. This article provides a definitive roadmap for the integration of hybrid artificial intelligence within the banking and investment sectors, ensuring that the pursuit of financial autonomy does not compromise systemic integrity or consumer trust.

Keywords: Agentic AI, Generative Finance, Multi-Agent Systems, Financial Regulation, Interpretable Machine Learning, Systemic Risk.

Introduction

The integration of artificial intelligence into the financial domain is not a contemporary novelty but rather the culmination of decades of iterative progress. From the early experiments in heuristic programming to the current sophistication of generative pre-trained transformers, the objective has remained consistent: the optimization of decision-making under uncertainty. As noted by Samuel (2019) in his seminal work on checkers, the essence of machine learning lies in the ability of a system to improve its performance through experience without being explicitly programmed for every contingency. In the context of modern finance, this "experience" is derived from petabytes of high-frequency trading data, credit histories, and global economic indicators.

However, the nature of this integration has fundamentally shifted. Early applications focused on narrow tasks, such as bio-inspired credit risk analysis using support vector machines (Yu et al., 2008). These models provided a significant leap over linear regression techniques but remained largely reactive. Today, we witness the rise of agentic AI-systems characterized by self-driven autonomy and the ability to interact within complex, multi-agent environments. This evolution is deeply intertwined with the concept of narrative economics. Shiller (2019) posits that popular stories and "viral" narratives drive major economic events. In an era where AI agents can both consume and generate these narratives at scale, the boundary between market analysis and market manipulation becomes increasingly blurred.

The transition to deep learning for finance, particularly in the construction of deep portfolios, allowed for the extraction of non-linear features that traditional models missed (Heaton et al., 2017). Yet, these advancements came at a cost: the "black box" problem. As financial institutions began deploying deep learning at scale, the lack of interpretability emerged as a critical vulnerability. The debate between using complex, opaque models versus inherently interpretable ones has become a central theme in academic discourse. Rudin (2019) argues vehemently against the practice of explaining black box models for high-stakes decisions, suggesting that the risks of "explanation error" are too high in fields like finance and law. Instead, she advocates for the design of models that are transparent by construction.

This research addresses the gap between the high-performance capabilities of generative AI and the stringent requirements of financial regulation and ethical fairness. While machine learning offers enterprises a path to enhanced efficiency and competitive advantage (Lee & Shin, 2020), the challenges of algorithm selection and bias remain pervasive. Mehrabi et al. (2021) highlight that bias can enter the machine learning pipeline at multiple stages, from data collection to model deployment, potentially leading to discriminatory outcomes in credit lending and insurance.

Furthermore, the emergence of generative AI agents as personalized financial advisors (Takayanagi et al., 2024) introduces a new layer of complexity to the fiduciary relationship. Can a machine be a "trusted" advisor? Chia (2019) explores this by comparing the trustworthiness of robo-advisers to human counterparts, noting that while machines lack human bias, they also lack human empathy and the ability to navigate moral nuances. As generative AI begins to power agent-based modeling and simulation (Gao et al., 2024), we are entering a phase where the entire financial system can be simulated as a "living" laboratory of autonomous agents. This paper explores these themes in depth, providing a comprehensive analysis of the risks and rewards of the agentic frontier in finance.

Methodology

The methodology for this research is rooted in a multi-dimensional theoretical synthesis that bridges the gap between traditional stochastic modeling and modern agentic AI. To understand the impact of self-driven AI on financial autonomy, we employ a comparative framework that evaluates three distinct technological epochs: the era of predictive heuristics, the rise of deep learning, and the current state of generative agentic systems.

The core of our analysis involves the Multi-Agent Financial Network (MAFN) approach, as pioneered by Markose (2013). This methodology allows for the modeling of systemic risk as an emergent property of the interactions between heterogeneous agents. Unlike traditional representative agent models, which assume all market participants behave identically and rationally, the MAFN approach accounts for diversity in strategies, information sets, and risk appetites. We extend this framework by incorporating the capabilities of Large Language Models (LLMs) to represent agent behavior. As Lu et al. (2024) demonstrate, LLMs can simulate complex human-like reasoning, allowing for more realistic agent-based models that react to narrative shifts and "sentiment" in a way that purely mathematical agents cannot.

To assess the interpretability and fairness of these systems, we utilize the theoretical foundations of game theory and explainable AI (XAI). Specifically, the Shapley value, originally a concept from n-person games (Shapley, 1953), is evaluated as a mechanism for feature attribution in complex financial models. This provides a rigorous mathematical basis for determining which variables-such as income, debt-to-equity ratios, or even social media sentiment-are driving an AI agent's decision. This is complemented by the guidelines provided by Molnar (2020) on making black box models explainable, contrasting post-hoc explanations with the "interpretable by design" philosophy advocated by Rudin (2019).

The research also incorporates a regulatory analysis, focusing on the EU AI Act and its implications for algorithmic trading (Azzutti, 2024). This involve a systematic review of legal frameworks that govern the use of AI in banking, particularly regarding the fiduciary duties of generative AI agents (Caspi et al., 2023). By synthesizing these diverse methodological strands-computational, ethical, and legal-we construct a holistic view of the financial AI landscape.

We also examine the role of hybrid artificial intelligence (de la Mata et al., 2024). This approach combines the strengths of symbolic AI (rule-based systems) with sub-symbolic AI (neural networks). In the banking sector, this hybridity is essential for maintaining compliance with rigid regulatory standards while benefiting from the pattern-recognition capabilities of deep learning. Our methodology scrutinizes how these hybrid systems manage the "exploration-exploitation" trade-off in portfolio management and risk assessment.

Finally, we analyze the concept of financial autonomy through the lens of enhanced customer engagement (Bhat & Krishnan, 2025). This involves evaluating the technological architecture of agentic AI-how these systems move from being simple "chatbots" to autonomous entities capable of executing trades, rebalancing portfolios, and negotiating terms on behalf of their human users. The methodology concludes with a critical assessment of systemic financial risk indicators within an agent-based framework (Mazzocchetti et al., 2020), ensuring that the micro-level autonomy of agents is reconciled with macro-level financial stability.

Results

The descriptive analysis of the current financial AI landscape reveals several critical findings. First and foremost, the shift toward agentic AI is no longer a theoretical possibility but a rapidly manifesting reality. Agentic AI, characterized by its ability to set its own sub-goals to achieve a high-level objective, represents a significant leap from the "passive" AI of the previous decade. Bhat and Krishnan (2025) identify that this self-driven nature is the key to achieving true financial autonomy for consumers, allowing for "hyper-personalized" financial paths that adapt in real-time to market volatility.

In the realm of credit risk, our analysis of bio-inspired computational intelligence and support vector machines (Yu et al., 2008) shows that while these models are effective at classification, they often fail to capture the cascading nature of systemic risk. This is where agent-based risk management (Theobald, 2015) provides superior results. By simulating the financial market as an ecosystem of interacting agents, we can identify "tipping points" that traditional models overlook. For example, our synthesis of the findings by Mazzocchetti et al. (2020) suggests that securitized assets, when managed by autonomous agents, can create hidden feedback loops that amplify systemic fragility.

The results regarding generative AI in financial advice (Takayanagi et al., 2024) are particularly telling. While LLMs demonstrate a high degree of "knowledge" regarding financial products, their effectiveness as personalized advisors is contingent upon their ability to handle "out-of-distribution" events-scenarios that were not present in their training data. Furthermore, the issue of trust remains a significant hurdle. Chia (2019) found that users are more likely to trust a robo-advisor for objective, data-driven tasks (like tax optimization) but still prefer human intervention for emotionally charged decisions (like retirement planning during a market crash).

A significant portion of our results focuses on the "Deep Portfolio" theory proposed by Heaton et al. (2017). Their work demonstrates that deep learning can successfully navigate the "factor zoo" of financial markets, identifying underlying drivers of return that are invisible to linear models. However, our analysis suggests that when these deep portfolios are integrated into an agentic framework, they must be governed by strict interpretability constraints. If an agent rebalances a multi-million dollar portfolio based on a "hidden layer" correlation, the lack of transparency can lead to regulatory non-compliance and investor panic.

In terms of fairness, our findings align with the survey by Mehrabi et al. (2021). We found that without explicit "de-biasing" interventions, agentic AI in finance tends to replicate historical inequities. For instance, if a generative agent is trained on historical mortgage data that reflects redlining practices, the agent will "learn" to avoid certain demographics, even if those demographics are currently creditworthy. This underscores the importance of "fairness-aware" machine learning in the banking sector.

The regulatory results indicate a growing consensus that "black box" algorithms in high-frequency trading and lending must be phased out or heavily supplemented by explainable components. Azzutti (2024) highlights that the EU AI Act will likely categorize many financial AI applications as "high-risk," requiring rigorous documentation, human oversight, and "auditability." This aligns with the move toward hybrid artificial intelligence (de la Mata et al., 2024), where neural networks provide the predictive power, and rule-based systems provide the "guardrails" for compliance.

Finally, the use of LLMs in agent-based modeling (Gao et al., 2024) has proven to be a game-changer for economic simulation. We found that LLM-powered agents can simulate "irrational" human behaviors-such as panic buying or narrative-driven speculation-more accurately than traditional rational-choice models. This allows central banks and regulators to "stress test" the financial system against the viral spread of harmful economic narratives, as theorized by Shiller (2019).

Discussion

The transition to agentic AI in finance represents a double-edged sword. On one hand, the promise of self-driven AI offers a level of efficiency and personalization that was previously unimaginable. On the other hand, the removal of the "human-in-the-loop" creates profound challenges for accountability, transparency, and systemic stability.

One of the most intense points of discussion is the tension between model complexity and interpretability. As Rudin (2019) correctly points out, the stakes in finance are too high to rely on "explanations" of models that are fundamentally opaque. If a model decides to deny a loan or sell off an asset class, the justification must be rooted in the model's inherent logic, not a post-hoc approximation like those generated by LIME or SHAP (Molnar, 2020). However, the counter-argument, often championed by proponents of deep learning (Heaton et al., 2017), is that simpler, interpretable models might lack the predictive accuracy necessary to navigate modern, complex markets. This creates a "performance-transparency trade-off" that financial institutions must navigate.

The role of narrative economics (Shiller, 2019) in the age of generative AI cannot be overstated. We are moving toward an era where AI agents are not just reacting to the news but are actively participating in the creation of the economic "story." If a generative AI agent, serving as a financial advisor, begins to promote a specific narrative about a stock or a cryptocurrency, and that narrative goes viral, the resulting market movement could be a self-fulfilling prophecy. This necessitates a new form of "narrative regulation," where the influence of AI on market sentiment is monitored as closely as its influence on trade execution.

The systemic risk implications of agentic AI are also a major concern. Markose (2013) and Mazzocchetti et al. (2020) have shown that multi-agent systems can exhibit "herding" behavior, where agents following similar algorithms all try to exit a position simultaneously, leading to a liquidity crisis. In an agentic future, where agents have the autonomy to evolve their own strategies, the risk of "algorithmic collusion"-where agents inadvertently learn to coordinate their actions to the detriment of the market-becomes a real possibility. This is why agent-based risk management (Theobald, 2015) must become a standard tool for regulators.

The ethical dimension of AI fairness (Mehrabi et al., 2021) also requires a shift in perspective. It is not enough to ensure that the data is clean; we must also ensure that the objectives of the agents are fair. In a competitive financial environment, an agent programmed solely to maximize profit will naturally seek out any advantage, including those that might be considered unethical or biased. Therefore, the "utility function" of financial agents must be designed to include social and ethical constraints, a task that is as much a philosophical challenge as it is a technical one.

Furthermore, the rise of "robo-advisers" and generative AI agents (Chia, 2019; Caspi et al., 2023) forces us to redefine the concept of fiduciary duty. If an AI agent provides advice that leads to a financial loss, who is responsible? Is it the developer of the LLM, the financial institution that deployed the agent, or the user who followed the advice? The current legal framework is ill-equipped to handle the "distributed agency" of LLM-based systems. As Azzutti (2024) suggests, the solution likely lies in a combination of strict liability for AI providers and enhanced transparency requirements that allow users to understand the risks involved.

The concept of hybrid artificial intelligence (de la Mata et al., 2024) offers a potential middle ground. By combining the intuitive, pattern-matching capabilities of LLMs with the logical, rule-following capabilities of traditional AI, we can create systems that are both powerful and predictable. For example, a generative agent could suggest a novel investment strategy, but that strategy would then be "vetted" by a symbolic AI system that checks it against regulatory requirements and risk limits. This "checks and balances" approach within the AI architecture itself could be the key to safe agentic autonomy.

Finally, we must consider the future of work and the "human element" in finance. As agents become more capable, the role of the human financial professional will shift from execution to oversight. Humans will become the "agents of agents," responsible for setting the high-level goals and ethical boundaries within which the AI operates. This requires a new set of skills, focusing on AI literacy, ethical judgment, and complex system management.

Conclusion

The evolution from machine learning to agentic AI marks a turning point in the history of finance. We have moved from models that help us understand the world to agents that can act upon it. This research has demonstrated that while the potential for enhanced efficiency, personalization, and risk management is vast, the challenges are equally significant.

The "black box" nature of deep learning remains a primary obstacle, and the movement toward interpretable models and hybrid AI is not just a technical preference but a regulatory necessity. The findings emphasize that systemic risk in the 21st century will be driven by the interactions of autonomous agents and the viral narratives they generate. Therefore, the tools we use to monitor and regulate the financial system must be as sophisticated as the agents they oversee.

We conclude that the successful integration of agentic AI into finance requires a three-pillared approach:

1. Interpretable Autonomy: Prioritizing models that are transparent by design or governed by rigorous explainability frameworks to ensure accountability.

2. Systemic Resilience: Utilizing agent-based modeling to simulate and mitigate the emergent risks of algorithmic herding and narrative-driven volatility.

3. Proactive Governance: Implementing regulatory frameworks, such as the EU AI Act, that define the ethical boundaries and fiduciary responsibilities of autonomous financial entities.

The path forward is one of "trusted autonomy," where the power of self-driven AI is harnessed within a framework of human values and systemic stability. As we stand on the precipice of this new frontier, the goal is not to replace human judgment but to augment it with agents that are as ethical and transparent as they are intelligent.

References

1. Azzutti, A. (2024). AI governance in algorithmic trading: some regulatory insights from the EU AI act. SSRN.

2. A. K. Bhat and G. Krishnan, "A Review of Agentic Artificial Intelligence: Power of Self-Driven AI in the Future of Financial Autonomy and Enhanced Customer Engagement," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2025, pp. 1160-1165, doi: 10.1109/ICSCDS65426.2025.11167368.

3. Caspi, I., Felber, S. S., & Gillis, T. B. (2023). Generative AI and the future of financial advice regulation.

4. Chia, H. (2019). In machines we trust: are robo-advisers more trustworthy than human financial advisers? Law Technology and Humans, 1, 129-141.

5. de la Mata, D. C., de Blanes Sebastián, M. G., & Camperos, M. C. (2024). Hybrid artificial intelligence: application in the banking sector. Revista de Ciencias Sociales, 30(3), 22-36.

6. Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., & Zhou, Z. (2024). Large language models empowered agent-based modeling and simulation: a survey and perspectives. Humanities and Social Sciences Communications, 11(1), 1-24.

7. Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. Applied Stochastic Models in Business and Industry, 33(1), 3-12.

8. Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons, 63(2), 157-170.

9. López de Prado, M. (2018). Advances in financial machine learning. John Wiley & Sons.

10. Lu, Y., Aleta, A., Du, C., Shi, L., & Moreno, Y. (2024). LLMs and generative agent-based models for complex systems research. Physics of Life Reviews, 51, 283-293.

11. Markose, S. M. (2013). Systemic risk analytics: a data-driven multi-agent financial network (MAFN) approach. Journal of Banking Regulation, 14(3–4), 285-305.

12. Mazzocchetti, A., Lauretta, E., Raberto, M., Teglio, A., & Cincotti, S. (2020). Systemic financial risk indicators and securitised assets: an agent-based framework. Journal of Economic Interaction and Coordination, 15(1), 9-47.

13. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35.

14. Molnar, C. (2020). Interpretable machine learning: A guide for making black box models explainable. Independently published.

15. Rudin, C. (2019). Stop explaining black box models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

16. Samuel, A. L. (2019). Some studies in machine learning using the game of checkers. Reprinted from IBM Journal of Research and Development, 1959.

17. Shapley, L. S. (1953). A value for n-person games. In Contributions to the theory of games (Vol. 2, pp. 307-317). Princeton University Press.

18. Shiller, R. J. (2019). Narrative economics: How stories go viral and drive major economic events. Princeton

University Press.

19. Takayanagi, T., Izumi, K., Sanz-Cruzado, J., McCreadie, R., & Ounis, I. (2024). Are generative AI agents effective personalized financial advisors?

20. Theobald, T. (2015). Agent-based risk management-a regulatory approach to financial markets. Journal of Economic Studies, 42(5), 780-820.

21. Yu, L., Wang, S., Lai, K. K., & Zhou, L. (2008). Bio-inspired credit risk analysis: computational intelligence with support vector machines. Springer, Berlin/Heidelberg, Germany, 197-222.