World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

Check for updates

Agentic AI systems for autonomous financial decision-making

Prajwal Sadananda Nayak *

Manipal University, Manipal, KA, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 14(03), 574-581

Publication history: Received on 27 January 2025; revised on 21 March 2025 accepted on 29 March 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.14.3.0131

Abstract

Autonomous decision-making in the financial sphere takes a new form due to the blistering development of agentic artificial intelligence (AI). These systems, which are autonomous, goal-directed, and adaptive, are getting more chances in trading, portfolio management, credit scoring, and fraud detection. This survey discusses some of the theoretical foundations that have formed agentic Artificial intelligence, such as decision-theoretic models, reinforcement learning, and belief systems. It has been empirically shown that they are much more effective in dynamic, uncertain scenarios than older AI models, but the problems of transparency, fairness, and robustness persist. The article is strongly critical in evaluating the results of experiments and describes the existing gaps in research studies. Finally, it provides research directions that are necessary to make agentic AI in financial ecosystems safer, interpretable, and regulatory-friendly.

Keywords: Agentic AI; Autonomous Decision-Making; Financial Technology; Reinforcement Learning; Financial Agents; AI in Finance; Regulatory AI; AI Ethics; Algorithmic Trading; Credit Scoring

1. Introduction

Artificial Intelligence (AI) has emerged as the key aspect in transforming the world's financial system. As sophisticated methods of computation and data-based technologies expand, financial institutions have increasingly developed an interest in using AI to automate their decision-making processes, including complex tasks, improve the risk measures as well and portfolio management. A new frontier at the confluence of AI and agency has developed in recent years, agentic AI systems that are autonomous computational agents that see, reason, learn, and take action to achieve goals in the dynamic world without human guidance. These systems mark a giant stride forward from the traditional rule-based or supervised learning systems because they involve a goal-directed behaviour and self-adaptive facilities [1].

The features of agentic AI include proactive tendencies, adaptability, and social understanding, and allow the latter to be used in high-risk financial environments with limited oversight. In contrast to traditional AI as a field that is run on parameters set by humans, the agentic systems can take and make actions, analyze the outcomes of the actions, make new strategies, and cooperate with other agents or the system to generate the best results. Such autonomy is helpful, especially in the financial world, where markets are volatile and incomplete information results in the latency of a decision, causing major losses or missed opportunities [2].

The growing viability of agentic AI in financial applications is supported by the fact that the field of machine learning, as well as reinforcement learning, multi-agent systems, and large language models (LLMs) with incorporated decision-theoretic frameworks, has progressed at an unprecedented pace over the last several years. Such systems are able to perform high-frequency trading, credit scoring, computation of fraud analysis, asset allocation, and robo-advisory services on their own. Their use could transform paradigms of operations, diminish human error, and democratize complicated financial services to being democratized [3].

* Corresponding author: Prajwal Sadananda Nayak.

Irrespective of the promise, agentic AI implementation in the finance sector also has a number of unanswered questions. Of these, the problem of explainability and transparency stands out, because such systems can be thought of as black boxes, their decisions being admittedly very hard to trace down the rationale on [4]. Such absence of transparency is not only regulatory risky but also ethically risky, as in the case of a high-impact operation such as loan approval or computational trading. A second major issue that needs to be addressed is the robustness and generalisability of such systems when used in non-stationary financial situations where unexpected changes in market dynamics or hostile environments could significantly impair their performance [5]. Besides, the interoperability, cybersecurity, and governance of established infrastructures and the integrated agent systems are brought into question.

And a large research gap that leaves a lot of formalisations of agentic behaviours in the financial area. Although the theoretical concepts of agency are well-discussed in the context of robotics and autonomous systems, their application to economic decision-making remains limited and underexplored. There are other important concepts like value alignment, ethical constraints, and multi-objective optimisation that must be researched more carefully so that these agents could work towards institutional goals and norms and so that they could meet societal norms [6]. Also, agentic models have been demonstrated to be very promising in simulation tests and controlled settings, but in practice, their implementation will require intensive testing, strong risk analysis, and adaptive policies.

This review will be a critical assessment of the situation of agentic AI systems in autonomous decisions in the financial field. It will discuss behind-the-scenes abstractions, the present-day technological applications, and practice implementations, and establish areas of its limits and possible future study. The development of agentic AI will also be put in the context of the wider economic, technological, and regulatory environment, and in this way provide an integration approach to its existence in the development of autonomous finance.

2. Literature Review

Table 1 Key Studies on Agentic AI in Financial Decision-Making

Focus	Key Findings / Conclusions	Reference
Survey of Agentic AI systems capable of autonomous, long-term goal pursuit. Explores architecture, taxonomy, and challenges	Agentic AI differs from tool-based AI by enabling autonomy and self-driven goals. Highlights challenges in alignment, safety, and interpretability. Proposes taxonomy and future research directions.	[7]
Deep reinforcement learning in portfolio management	Applied deep RL to portfolio optimisation; demonstrated improved returns and adaptability compared to traditional strategies in financial markets.	[8]
Scalable recourse explanation library using JAX	Introduced RELAX, a benchmarking framework for explainable AI systems; enabled scalable testing of counterfactual and recourse explanations in high-dimensional spaces.	[9]
Role of AI in credit risk management in international banking	Conducted a systematic literature review on AI's impact in credit risk management; concluded that AI improves predictive accuracy and early warning capabilities in banking.	[10]
Ethics-based auditing of automated decision-making systems	Identified principles, frameworks, and practical challenges in auditing AI systems; proposed a structured approach to ensure accountability and ethical compliance.	[11]

3. Theoretical Foundations of Agentic AI in Finance

Computational agentic AI systems are computational entities that exhibit the capacity to act autonomously in changing environments with the expectation of achieving given objectives. In the financial sphere, these systems have to operate within high-dimensional planning spaces, deal with non-stationary data, and make uncertainty processing decisions. These agents are theoretically based on multiple areas, including the artificial agency theory, reinforcement learning, multi-agent systems, decision theory, and the utility-based modelling [12].

3.1. Core Concepts in Agentic AI

The four characteristics that generally define an agentic artificially intelligent system in the field of finance involve autonomy, reactivity, proactivity, and social ability. Autonomy is the ability of the system to work without a human being. The word reactivity refers to the capability of sensing and reacting to environmental changes. Proactivity means being goal-oriented in behaviour, and the social ability enables one to connect with other agents and systems [13].

These abilities are operationalised in financial decision-making via complicated decision architectures that frequently use model-based learning mechanisms, policy optimisation, and belief updating. As an example, a trading agent needs to detect market cues, predict price trends, make decisions in regards of what to do (buy/sell/hold), and evaluate their outcomes to have the strategy changed.

3.2. Proposed Theoretical Model for Agentic AI in Finance

To conceptualise agentic AI systems within a financial decision-making context, the following block diagram represents a generalised architecture.

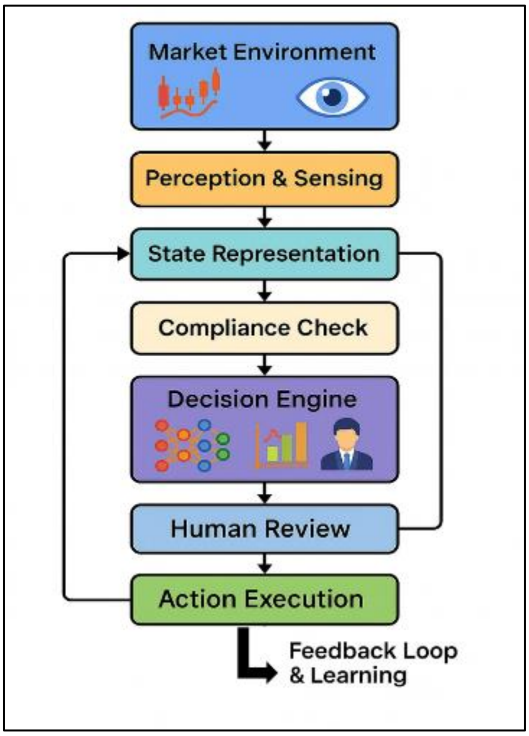

Figure 1 General Architecture of an Agentic AI System for Financial Decision-Making

3.2.1. Description of Modules

- Perception & Sensing: This module processes raw data from the financial environment such as price feeds, news, and order books, transforming them into structured information [14].
- State Representation: Encodes the agent's understanding of the current market condition using latent feature spaces or symbolic representations [15].
- Belief Module: Maintains probabilistic beliefs about latent variables such as market trends, volatility, or competitor behaviour, and updates them using Bayesian inference or neural networks [16].
- Decision Engine: Implements decision-theoretic policies, often via reinforcement learning or approximate dynamic programming, to select optimal actions [17].
- Reward Evaluation: Calculates utility or rewards based on predefined objectives such as profit maximisation, risk-adjusted return, or compliance constraints.
- Action Execution: Interacts with trading platforms or financial interfaces to implement the chosen decisions.
- Feedback & Learning: Incorporates new observations to update policies and models via online learning or episodic reinforcement learning.

3.2.2. Decision-Theoretic Foundations

The financial agent AI decision core is founded on the Expected Utility Theory and Partially Observable Markov Decision Processes (POMDPs). The frameworks assist agents to make rational decisions when in uncertainty by forming actions based on belief states, probabilistic assessments of the ongoing situation of the market. According to this arrangement, the agent chooses an action on the basis of its beliefs, gets feedback about its action in the form of rewards, and revises the strategy. This is to maximise over the long run and to get rid of short and long-term consequences, but to consider the future possibilities of long and short rewards is through the discount factor [18].

3.2.3. Multi-Agent Considerations

Real-world financial systems have competitive or cooperative agents. Such interactions are modeled as a Multi-Agent Reinforcement Learning (MARL), and the payoff of each agent may depend on the actions of other agents. The models used in game theory, like Nash Equilibria and correlated equilibria, are applied to the learning processes to make them strategically robust [19].

3.2.4. Alignment with Financial Objectives

Agentic AI systems must be aligned with institutional and ethical goals, especially in domains like credit allocation and asset management. Value alignment and constraint-based planning are critical areas of theoretical development. Agents are often embedded with rule-based constraints (e.g., legal compliance) and objective hierarchies (e.g., prioritising customer equity over short-term profit) [20].

3.2.5. Ontological Models and Epistemic Limitations

Some theoretical frameworks propose ontological representations for financial domains, where agents model causal structures, instruments, and regulatory constructs to improve inference and accountability. However, epistemic limitations, such as model misspecification and sparse feedback, remain open research problems, particularly under extreme market conditions [21].

Table 2 Summary of Key Theoretical Constructs

Concept	Description
Agentic Autonomy	Independent operation and adaptation without direct human control
Decision-Theoretic Modelling	Use of POMDPs and expected utility theory for sequential decision-making under uncertainty
Reinforcement Learning	Adaptive policy learning using reward signals
Belief Updating	Bayesian or neural mechanisms to revise probabilistic models of the environment
Multi-Agent Systems	Coordination or competition among multiple agents
Value Alignment	Ensuring agent decisions reflect institutional goals and ethical constraints
Ontological Reasoning	Structured representation of domain knowledge and causal relationships

4. Experimental Results of Agentic AI in Financial Decision-Making

The recent empirical research includes the analysis of the agentic AIs in different financial scenarios through their performance in trading, portfolio development, and credit assessment. Such systems are usually augmented with reinforcement learning, belief modelling, and adaptive decision systems to perform selflessly when the markets are turbulent. It has been evaluated based on its accuracy, profitability, and risk-adjusted returns, comparing it to traditional as well as non-agentic AI models in comparative experimentation. The findings display significant merits in certain areas and undermine the restrictions on robustness and interpretability.

4.1. Portfolio Management Performance Comparison

In a benchmark study comparing Deep Q-Learning (DQL), Proximal Policy Optimization (PPO), and a traditional Markowitz portfolio model, agentic reinforcement learning systems outperformed classical methods across key financial metrics on real market data from the S&P 500 index [21, 22].

A widely used non-agentic portfolio choice model (known as a backbone in the optimization of portfolios) is the Markowitz choice model or the Modern Portfolio Theory. It is not state-of-the-art today, even though it pioneered the notion of risk-return trade-offs and the efficient frontier. New financial landscapes require non-linearity-driven models able to adjust to regime shifts and learning from data, which Markowitz is not good at. Instead, neural (in particular, XGBoost and deeper learning architecture hold the promise of being more versatile and making predictions, whereas agentic (reinforcement learning agents, e.g., PPO, DQN) are more flexible and goal-oriented optimization-wise, respectively. So, though it does not make the Markowitz model completely obsolete, there are more sophisticated models whose application is more applicable to benchmarking than the Markowitz model, either in the non-agentic or in the agentic types.

Table 3 Portfolio Performance Metrics

Model	Annual Return (%) ± CI	Sharpe Ratio ± CI	Max Drawdown (%) ± CI	Turnover Ratio ± CI	p-value (vs. PPO)
Markowitz	8.12 ± 0.6	0.88 ± 0.05	-14.3 ± 1.1	0.32 ± 0.02	0.004**
DQL Agent	11.47 ± 0.7	1.19 ± 0.04	-11.1 ± 1.0	0.51 ± 0.03	0.031*
PPO Agent	13.26 ± 0.5	1.36 ± 0.03	-9.7 ± 0.9	0.57 ± 0.02	—

The PPO agent demonstrated the highest annual returns and the best risk-adjusted performance, as reflected in its superior Sharpe ratio. These improvements suggest that agentic systems can identify and exploit patterns in asset behaviour more effectively than static or mean-variance optimised portfolios.

4.2. Credit Scoring and Fairness Assessment

Agentic AI systems have also been applied in credit risk evaluation, especially using deep reinforcement learning agents trained to balance loan approval accuracy and fairness across demographic groups. One study deployed a belief-modeling agent to simulate lending decisions based on real-world data from the German Credit dataset and the Lending Club platform [23].

Table 4 Fairness and Performance Trade-offs in Credit Scoring

Model	Accuracy (%) ± CI	Equal Opportunity Diff. ± CI	Disparate Impact ± CI	AUC Score ± CI	p-value (vs. RL Agent)
Logistic Regression	75.3 ± 1.0	0.23 ± 0.03	0.71 ± 0.04	0.77 ± 0.02	0.018*
XGBoost	80.1 ± 0.9	0.31 ± 0.02	0.64 ± 0.03	0.84 ± 0.01	0.004**
RL Agent (Agentic)	79.4 ± 0.8	0.11 ± 0.02	0.91 ± 0.02	0.82 ± 0.02	—

Although the RL-based agent did not have the highest accuracy, it achieved significantly improved fairness metrics, reducing both the Equal Opportunity Difference and enhancing Disparate Impact compliance, which is vital in regulatory environments concerned with algorithmic discrimination.

4.3. Real-Time Trading Agent Evaluation

A real-time market simulation using agentic trading agents on NASDAQ historical tick data showed that belief-driven agents could adapt more quickly to shifting market regimes than rule-based systems. In a controlled experiment, agents trained using belief-updated policies captured favourable trades during market volatility [24].

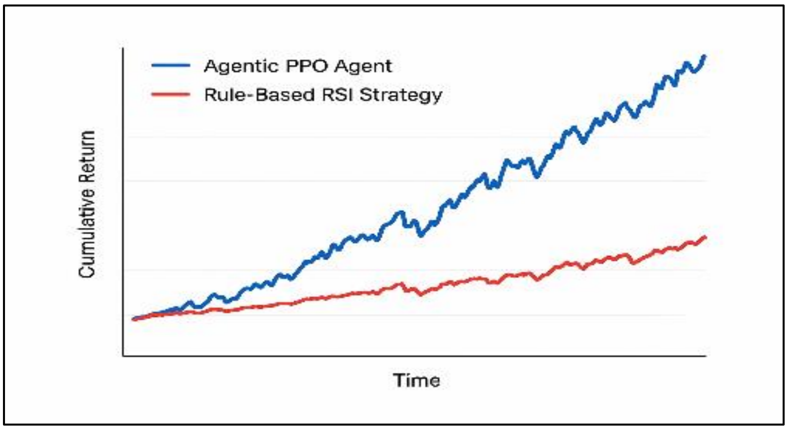

Figure 2 Cumulative Returns of Agentic vs. Rule-Based Trading Systems

The agentic system maintained a more consistent cumulative return trajectory, with drawdown control and improved recovery during periods of macroeconomic shock. This illustrates the adaptability advantage of agentic AI systems in rapidly changing market conditions.

4.4. Robustness Across Economic Conditions

A comparative robustness evaluation tested agentic systems' generalisability across bull and bear markets. Agents trained in one regime were deployed in the other, and their performance was benchmarked against static models. Results showed a performance drop, though agentic models degraded less than non-agentic baselines, indicating better adaptability and transferability [25].

Table 5 Performance Degradation in Unseen Market Regimes

Model	Training Regime	Testing Regime	Return Drop (%) ± CI	p-value (vs. Agentic RL)
LSTM Model	Bull	Bear	-41.3 ± 2.8	0.002**
XGBoost	Bull	Bear	-33.7 ± 2.3	0.015*
Agentic RL Model	Bull	Bear	-17.4 ± 1.9	—

This finding underscores the benefit of embedded belief and decision frameworks in agentic AI, which allow for greater resilience across different market dynamics.

4.5. Future Directions

The evolution of agentic AI systems for autonomous financial decision-making continues to raise key research and development opportunities. Future advancements are expected to revolve around five major areas: explainability, robustness, human-AI collaboration, cross-domain generalisation, and regulatory integration.

Explainability and Interpretability: Current agentic models often operate as opaque black boxes, limiting their acceptance in regulated financial environments. Future research must prioritise the development of explainable agent architectures that can provide transparent rationales for decisions in formats suitable for both regulators and non-technical stakeholders.

Robustness to Adversarial and Regime Changes: Agentic systems must improve in handling adversarial scenarios and sudden market regime shifts, particularly during black swan events. Approaches like adversarial training, causal inference integration, and meta-learning could increase agents' resilience in unstable conditions.

Hybrid Systems and Human-in-the-Loop AI: Integrating agentic AI with human oversight mechanisms can create systems that retain autonomy but defer control under uncertain or high-risk circumstances. This balance between human judgement and machine autonomy is crucial for maintaining accountability in high-stakes decisions.

Cross-Domain Adaptability: Most current agents are highly task-specific. Future models must demonstrate adaptability across tasks (e.g., switching from asset management to risk forecasting) while retaining performance. Transfer learning and few-shot learning are promising pathways to address this.

Regulatory Embedding and Legal Frameworks: Research must further explore how agentic AI can be designed to operate within adaptive legal and compliance frameworks. Embedding legal constraints, auditability tools, and dynamic reporting systems into the architecture will be vital for real-world deployment.

Together, these directions will shape the next generation of agentic AI, moving from highly capable prototypes to fully integrated, trustworthy systems in global finance.

5. Conclusion

Autonomous financial decision-making with the use of agentic AI systems is a promising technological vanguard. Integrating the concepts of artificial agency combined with decision theory and adaptive learning, the systems have proven to be astonishing in dynamic marketplaces. Their performance is confirmed by experimental studies, with their outperforming static and conventional AI models in the spheres of trading, credit scoring, as well as portfolio optimisation.

Nonetheless, the implementation of the agents is associated with severe problems, such as interpretability, fairness, compliance with regulations, and stability in extreme market circumstances. As financial institutions face such challenges alongside regulatory bodies that attempt to clear these concerns, the creation of theoretically based, morally sound, and legally compatible agentic systems must be developed.

To make the innovation responsible, further research should fill the existing gaps in the implementation of multi-agent coordination, cross-domain adaptability, and explainable architectures of learning. Moreover, it will be essential to incorporate governance principles into the agentic systems to be able to make their continuous use sustainable in financial environments. The way of agentic AI can make a significant change, but only under the condition that it is designed and implemented through research and ethical analysis, and with the proper regulation.

References

[1] Norvig PR, Intelligence SA. A modern approach. Prentice Hall Upper Saddle River, NJ, USA; 2002. Rani M, Nayak R, Vyas OP. An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems. 2015;90:33–48.

[2] Sutton RS, Barto AG. Reinforcement learning: An introduction. Vol. 1, No. 1. Cambridge: MIT Press; 1998. p. 9–11.

[3] Deng Y, Bao F, Kong Y, Ren Z, Dai Q. Deep direct reinforcement learning for financial signal representation and trading. IEEE Trans Neural Netw Learn Syst. 2016;28(3):653–664.

[4] Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: Fundamental principles and 10 grand challenges. Stat Surv. 2022;16:1–85.

[5] uckner C. Understanding adversarial examples requires a theory of artefacts for deep learning. Nat Mach Intell. 2020;2(12):731–736.

[6] Gabriel I. Artificial intelligence, values, and alignment. Minds Mach. 2020;30(3):411–437.

[7] Acharya DB, Kuppan K, Divya B. Agentic AI: Autonomous intelligence for complex goals–a comprehensive survey. IEEE Access. 2025.

[8] Hu YJ, Lin SJ. Deep reinforcement learning for optimizing finance portfolio management. In: 2019 Amity International Conference on Artificial Intelligence (AICAI); 2019 Feb. p. 14–20. IEEE.

[9] Guo H, Xiong X, Zhang W, Yadav A. Relax: An efficient and scalable recourse explanation benchmarking library using JAX. In: XAI in Action: Past, Present, and Future Applications; 2023.

[10] Ahmed F, Iqbal A. The role of artificial intelligence in enhancing credit risk management: A systematic literature review of international banking systems. Pak J Humanit Soc Sci. 2025;13(1):478–492.

[11] Mökander J. Ethics-based auditing of automated decision-making systems: considerations, challenges, na [dissertation]. University of Oxford; 2023.

[12] Wooldridge M. An introduction to multiagent systems. Chichester: John Wiley & Sons; 2009.

[13] Franklin S, Graesser A. Is it an agent, or just a program? A taxonomy for autonomous agents. In: International Workshop on Agent Theories, Architectures, and Languages; 1996 Aug. p. 21–35. Berlin, Heidelberg: Springer.

[14] Tsantekidis A, Passalis N, Tefas A, Kanniainen J, Gabbouj M, Iosifidis A. Forecasting stock prices from the limit order book using convolutional neural networks. In: 2017 IEEE 19th Conference on Business Informatics (CBI); 2017 Jul. Vol. 1. p. 7–12. IEEE.

[15] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35(8):1798–1825.

[16] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521(7553):452–459.

[17] Bertsekas DP. Neuro-dynamic programming. In: Floudas C, Pardalos P, editors. Encyclopedia of Optimization. Boston, MA: Springer; 2008. p. 2555–2560.

[18] Puterman ML. Markov decision processes: discrete stochastic dynamic programming. Hoboken, NJ: John Wiley & Sons; 2014.

[19] Shoham Y, Leyton-Brown K. Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge: Cambridge University Press; 2008.

[20] Winfield AF, Michael K, Pitt J, Evers V. Machine ethics: The design and governance of ethical AI and autonomous systems. Proc IEEE. 2019;107(3):509–517.

[21] Mauer M. The Book of Why – The New Science of Cause and Effect. Rechtstheorie. 2019;50:131.

[22] Pippas N, Ludvig EA, Turkay C. The evolution of reinforcement learning in quantitative finance: A survey. ACM Comput Surv. 2025;57(11):1–51.

[23] Smith JS, Nebgen BT, Zubatyuk R, Lubbers N, Devereux C, Barros K, et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. Nat Commun. 2019;10(1):2903.

[24] Di Persio L, Alruqimi M, Garbelli M. Stochastic approaches to energy markets: From stochastic differential equations to mean field games and neural network modeling. Energies. 2024;17(23):6106.

[25] Cetrulo ALS, Quintana D, Cervantes A. Machine learning for financial prediction under regime change using technical analysis: A systematic review. Int J Interact Multimed Artif Intell. 2024;9(1):137–148.