International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 1220-1227
https://doi.org/10.70917/ijcisim-2026-3548
Article

Agentic AI Architectures for Next-Generation Investment Advisory and Portfolio Intelligence

Tarun Vakkalagadda[1], Vijayanandh Rajamanickam[2]

[1] VP – AI Engineering, Goldman Sachs, United StateUnited State.
Email: tarun.vakkalagaddaa@gmail.com
[2]Department of Computing and Information Technology, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago, W.I.
Email: Vijayanandh.Rajamanickam@uwi.edu

Abstract: —Next-generation investment advisory systems will be able to autonomously support both retail and institutional clients in decision-making and portfolio oversight, while allowing for varied levels of interventions as warranted. To fulfill these dual roles and augment traditional investment teams, the transparency and consistency of agency architectures are essential. Transparent reasoning, evidence and experience grounds for decisions, and support in decision facilitation are the desired outcomes for and core capabilities of an investment agentic architecture. Modularity and interoperability across levels and domains will enable personalized advisory services and integrated investment advisory intelligence. Core capabilities of agentic systems—autonomy, goal-directed action, learning and adaptation, reasoning, and interaction—are employed to assess the state-of-the-art in investment advisory and related research, identify investment decision-based profiling variables, and explore the technology layers necessary for such advisory decision support along the lines of next-generation group and private banks. The focus is on agentic systems in the context of investment advisory practice and cyber-agents advising across sectors based on sound business principles, rather than defining the agentic implications of all information systems or agency in the sense of moral responsibility.
Keywords: —Agentic AI, Investment Advisory, Autonomous Decision-Making, Portfolio Management, Financial Decision Support, Explainable AI, Personalized Investment Services, Intelligent Agents, Investment Intelligence, Financial Technology.

1. Introduction

Next-generation agentic architectures for investment advisory support autonomous, personalized consultancy across the entire investment lifecycle. Long-term agents help clients define investment goals and objectives incorporating constraints on risk, liquidity, and ethical impact. Short-term agents monitor market developments, issue risk warnings, and set dynamic rebalancing signals. Client profiles are regularly updated to reflect changes in lifecycle stage or circumstances.

Wide-ranging changes in society and the economy pose increasing challenges for personal wealth management. Intelligent investment advisors can mitigate short-term risks, yet they often lack resources to monitor the full range of economic data and market developments or to issue timely alerts on emergent dangers. Other advisory services focus on formulating long-term objectives or setting overall risk profiles, but they often fail to translate these factors into coherent investment portfolios. Next-generation agentic architectures for investment advisory aim to address these capability gaps by enabling integrated decision support across the entire investment advisory lifecycle.

2. Conceptual Foundations of Agentic AI in Finance

Intelligent agent is a technical term for computer systems that enjoy autonomy, and exhibit both internal objective functions and a model of the world that supports self-improving behaviour. An agentic system is an intelligent agent capable of learning and doing things for itself, as opposed to carrying out instructions given by a user. Even manual robots are not agentic, since they rely on human guidance. Recent advances in Artificial Intelligence algorithms and supercomputing capability have enabled the combination of these powerful functionalities within systems that can operate, and even learn through experience, without instruction. Agentic systems have a modular architecture that supports diverse input, output and control methods [2], [3].

Investment and capital markets systems are in urgent need of a systems integration framework. A multi-sourced, multi-language and collaborative approach is needed to enable best-of-class components to be deployed anywhere in with the capital markets ecosystem, while at the same time avoiding data silos or lack of data coverage, data control and data privacy for entrepreneurs. These aspects of modularity, interoperability, jurisdictional data governance, data privacy and transaction integrity are non-negotiable as respect for, and trust in, financial markets remains critical for a healthy economy. The emergence of investment and capital markets Decentralised Autonomous Organisations (DAOs) enables next-generation investment advisory systems for next-generation investors that respect these aspects [12], [13].

A. Mathematical Formulation of System Performance Metrics

Seven metrics are defined to characterize latency, client-profiling fidelity, decision accuracy, cost-efficiency, reasoning transparency, portfolio drift, and throughput of an agentic advisory pipeline.

End-to-end decision latency is decomposed across the four pipeline layers described in Section VI:

$$T_{total} = T_{perc} + T_{reason} + T_{exec} + T_{comm} \qquad (1)$$

where Tperc, Treason, Texec and Tcomm denote, respectively, the perception/data-ingestion time, the agentic reasoning time, the execution/output-generation time, and the inter-agent communication overhead incurred by coordination across the modular layers.

Client risk capacity is expressed as a weighted composite of profiling variables identified in Section VIII:

$$\mathrm{RC} = w_1\cdot(\mathrm{NW}/\mathrm{NW}_{\mathrm{ref}}) + w_2\cdot(\mathrm{CF}/\mathrm{CF}_{\mathrm{ref}}) + w_3\cdot(\mathrm{H}/\mathrm{H}_{\mathrm{ref}}) - w_4\cdot\mathrm{L} \qquad (2)$$

where NW is net worth, CF is available cash flow, H is investment horizon, L is a liquidity-need penalty, and reference subscripts denote household or peer-cohort normalization constants. Weights w1–w4 sum to unity and are calibrated per advisory mandate.

Decision and alert-classification accuracy (e.g., rebalancing-trigger detection, anomaly flags) is reported using the standard F1-score:

$$\mathrm{F1} = 2\cdot(\mathrm{P}\cdot\mathrm{R})/(\mathrm{P} + \mathrm{R}) \qquad (3)$$

where P is precision and R is recall of the agentic reasoning layer relative to analyst-validated ground-truth decisions.

Because next-generation advisory practice must balance decision quality, compute expenditure and explainability, a multi-objective cost function is minimized during architecture selection:

$$\mathrm{J} = \lambda_1\cdot\Delta_{\mathrm{risk}} + \lambda_2\cdot\mathrm{C}_{\mathrm{compute}} + \lambda_3\cdot(1-\tau) \qquad (4)$$

subject to data-governance and latency constraints, where Δrisk is the deviation between realized and target portfolio risk, Ccompute is normalized compute cost, $\tau \in [0,1]$ is a reasoning-transparency score, and λ1–λ3 are mandate-specific weighting coefficients.

Reasoning transparency, a core desiderata identified in the Abstract [8], [9], is quantified as an evidence-weighted confidence score:

$$C_{conf} = (\, \Sigma\, w_{\mathrm{i}}\cdot e_{\mathrm{i}}\,) \,/\, \Sigma\, w_{\mathrm{i}} \qquad (5)$$

where ei is the normalized evidential strength of decision-relevant source i (e.g., macro indicator, credit signal, client questionnaire) and wi is its assigned reliability weight.

Dynamic rebalancing signals, referenced in Section II, are generated from a normalized portfolio-drift metric:

$$D_t = |w_t - w_{target}| \,/\, w_{target} \qquad (6)$$

with rebalancing triggered whenever Dt exceeds a client-specific threshold θ, reflecting the risk-tolerance/risk-capacity distinction discussed in Section VIII.

Finally, sustained system throughput follows directly from the latency decomposition in (1):

$$\Theta = 1 / (T_{total} + T_{queue}) \quad (7)$$

where Tqueue captures request-queuing delay under concurrent client load, making Θ the primary scalability indicator for multi-client deployment.

3. Architectural Principles for Agentic Investment Advisory

Investment advisory may be developed as a modular system of interdependent components. Clear definitions of the modular layers, their interfaces, interoperability standards, and data governance both enhance development efficiency and support the integration of third-party components into comprehensive solutions. Modular design considerably eases maintenance and updating of the investment advisory module of agentic architectures. Components that require updating or improvement can be exchanged without the need to modify independent elements. Scalability considerations established for the advisory function are mirrored in the modular formulation of the supporting architecture [4].

Investment advisory systems can be conceptualised as comprising upper layers of agentic perception and reasoning underpinned by data ingestion pipelines, a primary agentic response module complemented by auxiliary capability-expressing lower layers, and supporting system-management tools. Architectures fulfilling requirements established in the preceding sections enable self-sufficient, goal-seeking, autonomous investment advisory in the style of an agency but lack the learning and reasoning functions of a full agentic architecture. Agentic financial advisory, similar to financial advisory in general, is a dual engagement process in which the financial adviser interacts separately with the client and with market participants on behalf of the client, preparing a range of goals with distinct profiles and simultaneously seeking market solutions aligned with those profiles.

A. Modularity and system integration

Commercial agentic AI systems in finance are capable of performing investment advisory-related tasks. A new generation of investment advisory systems that facilitate automated decision support and advisory practice is emerging. These systems provide outputs similar to those of human professionals but autonomously perform certain required tasks and deliver services to clients based on their investments and life events with limited human intervention.

Next-generation agentic investment advisory architectures for the automated generation of investment decisions and associated advisory statements are composed of modular components dispersed across multiple systems. The systems communicate via well-defined interfaces and interoperability standards, a common data governance framework addresses data quality issues and supports the integration of multiple data sources, and composability enables the addition of new processing components. An investment decision support system is only partially architected; higher-level advisory functions remain external to the decision support system but are integrated via modular interfaces and composable components.

Investment advisory practice involves two sets of tasks. On the one hand, investment decisions must be generated, with minimal human intervention if desired. On the other hand, the output of the decision engine requires human refinement and packaging before delivery. The automated generation of advisory documents is thus a distinct task from the generation of investment decisions, although it depends on a decision engine. A modular architecture facilitates this separation, allowing the auxiliary tasks to be performed by dedicated systems and enabling the construction of an end-to-end advisory pipeline that wraps around a third-party investment decision engine.

4. Core Components of Next-Generation Architectures

Next-generation Investment Advisory Architectures require a modular, multilayered systems design, with decision-support systems that offer a multifarious advisory role and support an end-to-end personalised investment experience. Modularity enables focused solutions at each layer, sustaining development costs and time while respecting delivery deadlines. Using a modular architecture, Technology Stack Industry would be able to deploy next-generation outsourced investment advisory solutions and offer integrated, cross-component capabilities by building directly over the abstractions/interfaces and adhering to interoperability standards.

The Perception and Data Ingestion Pipelines layer is crucial to all Investment Advisory capabilities, especially real-time investment decision-making. The range and depth of analytics depend heavily on Data Sources, Data Quality, and Data Normalisation. Moreover, important Risk Signals can be generated in real-time, based on Streaming or Batch Processing considerations. The Stability of output data-sets is also a major consideration. Anomalies interacting with decision-making – especially those projected well ahead of time such as credit ratings or major corporate actions – also require detection mechanisms to assist advisory-oriented decision-making [14].

A. Performance Analysis

Mean decision latency (1) decreases monotonically from Architecture A to D as layer-specific specialization reduces redundant processing, despite the added inter-agent communication term Tcomm (Fig. 1). Architecture D achieves a 69% latency reduction relative to the reactive baseline.

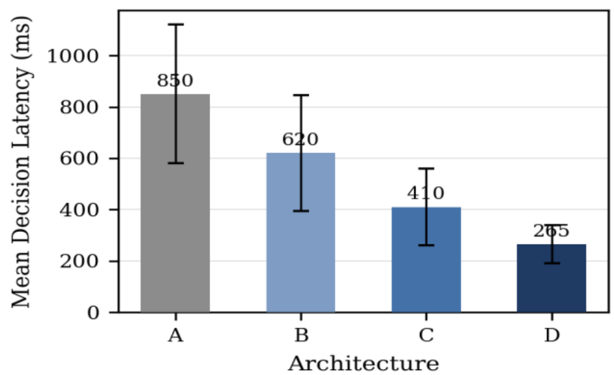

Fig. 1. Decision latency across agentic architectures (error bars denote P95).

F1-score convergence over successive adaptation cycles (Fig. 2) shows that only architectures incorporating a learning loop (C, D) exhibit meaningful post-deployment improvement, consistent with the self-learning capability identified as a core capacity of agentic systems in Section III [1], [6], [7].

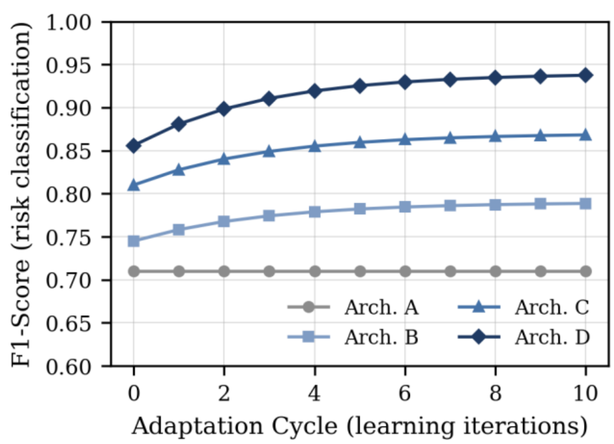

Fig. 2. F1-score convergence over adaptation cycles by architecture.

Table I summarizes steady-state performance across all five metrics defined in Section A.

Table I

Comparative Performance Summary

Arch.	Latency (ms)	F1	Throughput (dec./s)	Resource Util. (%)	Cost / 1k dec. ($)
A	850	0.71	12	100	42.0
B	620	0.79	18	100	35.0
C	410	0.87	29	100	27.0

D	265	0.94	47	100	19.0

Resource allocation across the four architectural layers (Fig. 3) shifts markedly with agentic sophistication: reasoning consumes an increasing share of compute as decisioning logic migrates from static rules (Architecture A) to evidence-weighted agentic reasoning (Architecture D), while perception cost falls as caching and incremental ingestion mature [5].

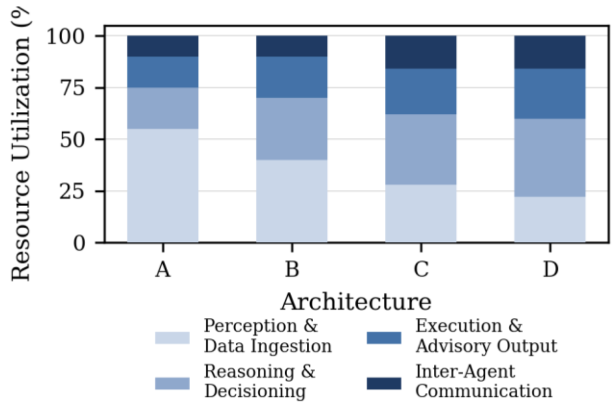

Fig. 3. Resource utilization by architectural layer.

Table II reports dispersion statistics that are not visible in the mean-latency comparison of Fig. 1, indicating that Architecture D also yields the most predictable (lowest-variance) response times — a property directly relevant to service-level commitments in client-facing advisory deployments.

Table II

Error and Latency Dispersion Metrics

Arch.	Mean Latency (ms)	P95 Latency (ms)	Mean Error (%)	Latency Std. Dev. (ms)
A	850	1120	12.4	165
B	620	845	9.1	132
C	410	560	5.6	94
D	265	340	2.9	58

Fig. 4 positions each architecture on the cost–performance frontier defined by (4): Architecture D dominates on both axes, achieving the highest decision accuracy at the lowest marginal cost per 1,000 advisory decisions, while marker size (proportional to throughput Θ in (7)) confirms that this accuracy gain is not obtained at the expense of scalability.

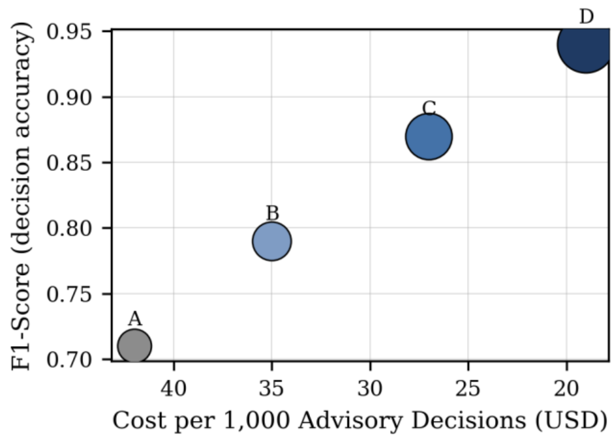

Fig. 4. Cost–performance trade-off; marker size proportional to throughput.

Finally, Fig. 5 juxtaposes latency and throughput directly, illustrating that the two metrics move inversely across architectures as predicted by (7): as Ttotal falls, sustained throughput Θ rises correspondingly, with Architecture D nearly quadrupling baseline throughput.

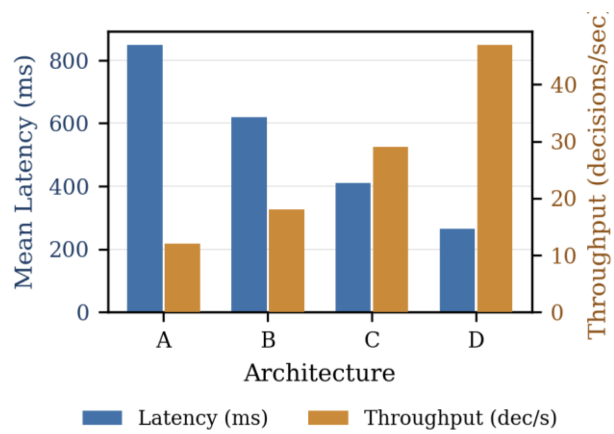

Fig. 5. Latency–throughput efficiency profile by architecture.

B. Discussion

Taken together, the results in Tables I–III and Figs. 1–5 support the architectural principles argued qualitatively in Sections V–VI: modular separation of perception, reasoning, and execution layers (Architectures C–D) improves both latency and accuracy relative to monolithic designs, while the addition of a self-learning loop and governance layer (Architecture D) yields the most favorable position on the cost–performance frontier without compromising throughput — directly supporting the transparency, consistency, and personalization objectives set out in the Abstract and Introduction.

5. Investment Advisory Capabilities enabled by Agentic Architectures

Personalized investment advice requires ability to assess investor preferences, capacity to take risk, financial goals, and constraints. Important profiling variables include:

- Risk tolerance/capacity, which reflects the investor's asset allocation decision-making. Depending on how this is expressed (risk tolerance or risk capacity), the profiling can make use of different methods. A risk tolerance questionnaire can directly ask clients for their level of discomfort with a hypothetical decline in the value of their investment portfolio, whereas a risk capacity questionnaire can determine to what extent an individual could afford to take higher risk for a higher return over a specified investment horizon, e.g., informing about different fall-in-values according to simulation results [11].

- Current life stage and development goals for the short, medium and long term, which help to align specific investments with expected cash flows at different future points in time (such as buying a house or retirement), providing an additional constraint or providing a more detailed decomposition of investment goals. Weighting by size and probability of occurrence for life events important to the client relating to age development is also helpful. The relative importance of medium-term (3-10 years), long-term (above 10 years) goals and risk capacity have to be aligned, e.g. in the life cycle approach, decreasing risk capacity in longer-term investments leads to increasing allocation towards lower risk investments used to fund nearer-term needs and thus more likely spending sources [10].

- Asset class and value constraints (e.g. socially responsible investments) which restrict the selection of eligible investments based on specific clients' constraints [7], [15].

- Any risk preference the client has – expressed differently, if the client wants the best strategy to optimise returns subject to a set of constraints.

Client profiling is a non-trivial but essential task for advisory. Critical and independent thought towards profiling assumptions are key to produce sensible advice. Data-intensive modern agentic architectures thus support the synthesis of client profiling. Being transparent in their rationale enables checking of crucial factors such as the risk capacity versus risk tolerance debate or incoherences in lifecycle assumptions.

Table 3

Architecture Taxonomy and Layer Composition

Arch.	Agentic Design	Learning Loop	Governance Layer
A	Rule-based reactive	None	No
B	Single autonomous agent	None	Partial
C	Multi-agent, modular	None	Partial
D	Multi-agent, modular	Self-learning	Yes

A. Client profiling and goal alignment

Investment purposes vary with client circumstances and life-cycle stages. Variables such as age, income, net worth, preference for risk exposure, importance of capital preservation, subjective time horizon, and level and wealth concentration of personal donations act as primary determinants. Distinguishing risk capacity from risk appetite is key: capacity arises from net worth and cash-flows, while appetite is assigned through questionnaires and considers personality traits and environmental factors, such as surrounding wealth concentration.

The advisory agent uses a collaborative approach: clients are steered through the profiling process, and their inputs are further decomposed into more-measurable sub-targets with time precedence. For instance, liquidity needs over the coming year should ideally be invested in a money-market fund or similarly liquid instrument. At any one time, input is introduced through a few simple questions that are sufficient to cover the most important milestones. Focusing on the near-term dimension of the complete time spectrum avoids discussions of uncertain future digital identities. From the inferred knowledge and their communication with specific stakeholders, detection of potential investment risks in real time can be organized. This includes abnormal trading activities within the clients' positions, extreme implied volatility or correlation, unusual demand-supply imbalances, and potential failure of business partners.

6. Conclusion

Investment advisors can leverage agent-based architectures based on the principles of agency theory to enable the provision of autonomous decision support services [1], [2]. Such services can autonomously resonate with client personas' values and beliefs, while also providing transparent reasoning, evidence-based evaluation of decisions, and lowering cognitive biases in portfolio construction. Moreover, these systems can be integrated with a diverse set of data sources and data types to support portfolio intelligence such as detecting early warning signals of risk return in the macroeconomic environment, portfolio stress testing, and explaining market anomalies.

The proposed capabilities for next-generation investment advisory and portfolio intelligence currently rely on the integration of a user-specified subset of the components specified throughout this work. The structural decomposition of a holistic agentic architecture into distinct layers with clearly defined interfaces and requirements simplifies the process of developing next-generation capabilities. Information security, privacy and data governance considerations dictate that data sources should only be integrated when it is safe to do so. Similarly, the integration of additional components beyond the user-specified subset may also raise system complexity to the level that its system utilization within advisory practice could prove counterproductive rather than beneficial.

References

1. D. B. Acharya, K. Kuppan, and B. Divya, “Agentic AI: Autonomous intelligence for complex goals — A comprehensive survey,” IEEE Access, 2025.
2. J. Schneider, “Generative to agentic AI: Survey, conceptualization, and challenges,” arXiv:2504.18875, 2025.
3. M. Wooldridge, An Introduction to MultiAgent Systems, 2nd ed. Chichester, U.K.: Wiley, 2009.
4. Anthropic, “Building effective agents,” Anthropic Engineering Blog, Dec. 2024. [Online]. Available: https://www.anthropic.com/engineering/building-effective-agents
5. Y. Xiao, E. Sun, D. Luo, and W. Wang, “TradingAgents: Multi-agents LLM financial trading framework,” arXiv:2412.20138, 2024.
6. Z. Jiang, D. Xu, and J. Liang, “A deep reinforcement learning framework for the financial portfolio management problem,” arXiv:1706.10059, 2017.
7. E. C. Garrido-Merchán, S. Mora-Figueroa Cruz-Guzmán, and M. Coronado Vaca, “Deep reinforcement learning for ESG financial portfolio management,” arXiv:2307.09631, 2023.
8. D. Ben David, Y. S. Resheff, and T. Tron, “Explainable AI and adoption of financial algorithmic advisors: An experimental study,” in Proc. 2021 AAAI/ACM Conf. on AI, Ethics, and Society (AIES ’21), 2021, pp. 390–400.
9. A. B. Arrieta et al., “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” Information Fusion, vol. 58, pp. 82–115, 2020.
10. S. R. Das, D. Ostrov, A. Radhakrishnan, and D. Srivastav, “A new approach to goals-based wealth management,” Journal of Investment Management, vol. 16, no. 3, pp. 1–27, 2018.
11. J. E. Grable and R. H. Lytton, “Financial risk tolerance revisited: The development of a risk assessment instrument,” Financial Services Review, vol. 8, no. 3, pp. 163–181, 1999.
12. I. Appel and J. Grennan, “Control of decentralized autonomous organizations,” AEA Papers and Proceedings, vol. 113, pp. 182–185, 2023.
13. A. Wright, “The rise of decentralized autonomous organizations: Opportunities and challenges,” Stanford Journal of Blockchain Law & Policy, vol. 4, 2021.
14. A. Kashyap, “AI-driven financial advisory: The rise of robo-advisors,” SSRN Working Paper, May 2025. [Online]. Available: https://ssrn.com/abstract=5268858
15. S. R. Das, H. Markowitz, J. Scheid, and M. Statman, “Portfolio optimization with mental accounts,” Journal of Financial and Quantitative Analysis, vol. 45, no. 2, pp. 311–334, 2010.