A Framework for Provable Consistency and Governance in Multi-Agent Generative AI Systems for Finance Management Reporting

Sushil Kumar Nagur

Research Scholar, Presidency University, Hyderabad, India

Corresponding Author Email: sushilnagur@gmail.com

Abstract

Can we rely on an Automated System for Reports or have Reports be produced by an Experienced Financial Professional with Many Years of Financial Knowledge? This question is the basis of the rationale for conducting this review. Financial Management Reporting (FMR) is composed of many categories such as; Profit/Loss Statements, Variance Analysis, Budget versus Actual, Executive Dashboards, etc, and contains a very basic implied principle; each number you see in a FMR can be traced back to at least one source (organisational ledger), and each narrative used within a FMR must be based upon verifiable data. In addition to continuing to automate many areas of FMR via the use of LLM's (Large Language Models) and multi-agent AI systems, there are also indications that they employ probabilistic processes when reasoning. The findings of this review can be used as further support for the continued use of Automated Systems in the preparation of Financial Management Reports. This means the results produced by these systems can be hallucinatory in nature and rely upon a foundation of assumptions that are created on a fluid basis and cannot be independently verified. After reviewing 20 recent research studies across five interrelated research fields — Generative AI in Finance, Multi-Agent Collaboration, Explainable AI (XAI), Formal Verification, and AI Governance — we conducted this same review but used this body of literature to identify each study's relevance with respect to the integrity of management reporting. What we discovered in all respects was relatively concerning; however, it did not surprise us due to the pre-existing knowledge that the primary focus of the research community has been on developing ways to automate FMR reporting as quickly as possible and much less research has been conducted concerning methods of verifying the accuracy of report results. Furthermore, many Governance Models continue to rely on human review for validating data rather than using automated mathematical algorithms. Based on these gaps, we have developed a "Provable Consistency" framework consisting of three components: a Multi-Agent Audit Layer; an audit *trail linking all outputs to their source data using XAI; and an industry-agnostic governance protocol supporting the reporting processes. We also discuss potential target Q1/Q2 journals for this area of research related to our findings.*

***Keywords**: Provable Consistency, Multi-Agent Systems, Generative AI, Finance Management Reporting, Explainable AI, LLM Hallucinations, Formal Verification, Deterministic Anchoring, Governance Frameworks, Executive Trust*

1. Introduction

The answer provided by 4 out of 5 CFOs when asked what is most important for establishing traceability and trust within managerial reporting is "traceability". Each quarter, as part of the financial close, it is expected that Finance personnel will extract multiple forms of data (general ledgers, cost centres, departmental types of expenditures) from a variety of systems (Operations, HR, Payroll, etc.) to create reports such as Profit and Loss Statements, Variance Reports, Cashflownoids, etc. for the Board of Directors. The unwritten rule is that any number that appears at the end of any report must be traceable backward through journals, journal entries and the original transaction being documented. If that chain is broken the report loses its validity.

We have seen the emergence of Generative A.I., particularly LLMs, into this area of business. Companies are currently piloting A.I. in various applications: producing drafts of managerial narratives, flagging anomalies within large data sets and constructing real-time executive summaries (Aldasoro et al., 2024). The productivity argument is compelling: what used to take a finance team days to consolidate and provide commentary can now theoretically be done in minutes. It is estimated that AI expenditures across sectors will reach $97 billion by 2027, and automating the management reporting process is frequently mentioned as one of the primary use cases.

But there is a problem with this, and it is fundamental rather than just a minor wrinkle: LLMs generate results based on predicting what the most logically likely next token will be based upon a context window, but they do so by providing approximations rather than validating. Hence, when a model produces a narrative by combining quarterly sales revenues, allocations of costs to cost centres and commentary from departments is effectively performing pattern matching with respect to sophisticated patterns rather than mathematically reconciling them together. In fact Smith & White (2024) have demonstrated that the rate that LLMs generate fictitious logical steps increases not decreases as they are forced to create a longer logical reasoning chain and as the LLM becomes bigger; when used to produce management reports for example, perhaps with 5 subsidiaries and 3 currencies consolidated, this cannot be thought of as merely a theory - if a variance is generated that does not exist or trend line has been produced that was hallucinogenic in nature then a board might vote on capital allocations incorrectly.

A partial remedy to the problem can be found through Multi-Agent Systems (MAS). The underlying principle is obvious, rather than using one AI to perform all tasks as a singular entity, assign roles to the various AI agents, such as an analyst agent, an auditor agent, and a compliance agent. Allow those agents to verify and validate their work using the other agents as their basis for verification and validation. For example, Xiao et al. (2025) discovered that MAS can reduce specific types of errors by one-third, while Chen et al. (2023) discovered that MAS groups significantly outperform solo agents for reasoning tasks. Those findings are good news. However, the problem with this approach is that the agents will still verify and validate each other based on statistical measures. The auditor agent will verify the draft that the analyst agent wrote by using the same type of stochastic reasoning used by the analyst agent to produce the draft. Therefore, the auditor agent's verification is based on a peer opinion rather than objective proof that the draft is accurate. It is similar to two accountants reviewing each other's work without documenting their measurements against the trial balance.

The Explainability Arts (EA) community has also been making significant strides to increase the transparency of AI-based decisions (Arrieta et al., 2021). For instance, the EA measurement tools SHAP and LIME can identify which of the input data (i.e., relevant features) contributed to the output. Therefore, while the EA measurement tools provide valuable information, at present, they are not adequate for management reporting in an SLA.

Having identified that a gap existed regarding the integration of Generative AI with traditional forms of finance, this review was developed to address this shortcoming. To accomplish this goal, we conducted an extensive review of 20 relevant papers that discuss generative AI and finance, multi-agent architectures, explainable AI, formal verification, and overall, AI governance in the context of developing AI systems with the capability to mathematically prove the validity of the management reports produced by them from the underlying financial data. The short answer to our research question is that while there are several components that make up the puzzle, they have yet to be brought together into a comprehensive solution. The long answer will be presented in section 2. In this section, we will outline our methodology for reviewing recent literature related to the development and use of Generative AI in the area of Financial Reporting.

2. Methodology

2.1 Search Strategy and Databases

The methodology we utilized in this study consisted of searching five different databases; Scopus, Web of Science, IEEE Xplore, SpringerLink, and Google Scholar; for articles related to the following three concepts:

1. Generative AI and LLMs used in the production of Financial Reporting Management
2. Multi-agent systems used for reasoning, verifying, or auditing in finance
3. AI governance, explainability, and formal verification related to high-stakes management reporting

We included three examples of the queries used for searches; "Generative AI AND Management Reporting"; "Multi-agent AND Audit Verification AND Finance" and "Formal Verification AND LLM Hallucination". We limited our searches to only include studies published between 2021-2025 so that we would provide a synopsis of the field as it stands today without including any studies that predate the development of LLMs.

2.2 Inclusion and Exclusion Criteria

To qualify as an article in this review, the article had to satisfy all of the following 4 requirements: Between the years of 2021 - 2025, the article appeared either in a Q1/Q2 journal or an internationally recognized venue like the ICLR or BIS; it presented at least one of the three pillars of research for management reporting (AI Governance, AI Algorithms, and AI Interpretation); it was published in English; and it did not fit into any of the excluded categories listed below.

The following categories were excluded: 1) Articles that are opinion pieces and lack empirical or theoretical data; 2) Articles solely focused on narrow NLP studies and do not discuss an angle of governance; 3) Studies of consumer-facing AI applications (i.e. chat bot, robo-advisor) that have little to do with internal management reporting; 4) Any article that was substantially similar to previously published studies only found in the study pool.

2.3 Selection Process

The initial search yielded around 85 articles from different sources. Through scanning of the titles, abstracts and the entire text, 38 articles remained. From these 38 articles, a final selection of 20 articles were made for review. The final 20 articles were selected according to whether they covered all three pillars of research and how directly they engaged, or implied a relationship of engagement, with verifying the gap in AI-generated management reports.

For the 20 articles selected for review, assessments were made of each article using the following five factors/criteria: a) Research Problem b) Research Methods c) Main results d) Stated or implied research gaps e) Significant to the concept of the Provable Consistency framework and criteria.

Following assessments of the final selected 20 articles were made using a thematic synthesis process, the information contained in each of the articles regarding cross-cutting themes was generated. This information was summarized in Section 3 of this article.

2.4 Classification Framework

In total we classified 20 articles into five categories for ease of discussion: (1) use of generative AI for financial reporting with focus on speed to generate versus risk to generate reliable report; (2) multi-agent architecture and verification of errors between agents; (3) use of XAI and trust mechanisms; (4) use of formal verification with deterministic means (symbolic logic, knowledge graphs); and (5) AI governance related to reporting. A few of these are applied across categories; however, each was identified based on its primary area of focus.

3. Results and Discussion

A one overarching observation became clear from reading all 20 articles; there remains no solution to date on how an overall multi-agent AI has the ability to reliably prove consistency between each of the management reports being generated, and the corresponding financial data relied upon to generate those reports. As shown below, we develop our findings cluster by cluster.

3.1 Generative AI in Financial Reporting: Fast, But Trustworthy?

The documented speed improvements realised using generative AI for producing financial reporting solutions continues to prove as viable and significant based on evidence from Aldasoro et al. 2024) from their survey of how AI is used across all worldwide companies in their financial information processing. That being said, the same document indicates that in addition to speed improvements from the use of AI for financial reporting, organisations have invested considerably in technology, and yet comparatively little has been invested with regard to rectifying the accuracy of the actual financial reports produced using the technology.

Based on their 2024 work, Smith and White created a framework to evaluate the frequency with which hallucinations arise in different model families, including the GPT-4, Claude, and Llama family of models; their findings reported that longer reasoning chains, which often occur in producing management reports that contain information from multiple different sources, produce more logical errors than shorter reasoning chains. This puts AI-generated financial reports at serious risk for producing hallucinations since producing a consolidated report using five different ledgers, over three different currencies, and at least a dozen different cost centre could qualify as a long-chain type of reasoning task.

Kim, Muhn, and Nikolaev (2024) added another dimension to consider regarding auditing processes. They proposed using Edge AI, which allows AI systems to perform their analysis closer to where the originating data is located; however, the thousands of analyses performed using Edge AI straddle both human and AI reconciliation methods. Therefore, because of the risk associated with producing accurate reports and non-production errors (hallucinations), they recommended that automated reconciliation be put in place to verify that the figures generated by an AI system generate a reconciled record (source document).

3.2 Multi-Agent Systems: Agreement Is Not the Same as Accuracy

Management reporting follows an ordered series of actions, as it consists of extracting, cleansing, analysing, composing, and reviewing. Therefore, having different agents perform these aforementioned functions makes sense when using the MAS model of dividing each task among specialists. Zhang et al. (2024) developed FinAgent, which combines inputs of the three primary types (i.e., image, text, and tabular) to perform financial analyses, outperforming text-only baselines for sentiment analyses and trend analysis. This is impressive, but it does not appear to have an audit agent present-whom in the architecture takes the consolidated evidence and compares it to the general ledger. Therefore, while the FinAgent system creates compelling compositions, it is not capable of authenticating them.

Xiao et al. (2025) examined this with their TradingAgents system, where three distinct agents (macro analyst, technical trader, and risk manager) challenge each other's decision-making; error rates were reduced by approximately 34%, which is significant. The problem, however, is that the challenge process relies upon each of the agents evaluating the others using the same probabilistic rationale that they employed to determine their own conclusions; thus, neither of them use deterministic reference points in making assessments. An analogy for management reporting would be that if two analysts write a narrative for the same P&L statement independently, the fact that they both write identical narratives provides some comfort, but the fact that both analysts produce estimates does not confirm that they have both checked their respective estimates to the general ledger.

In an investigation by Chen and co-workers (2023) into AgentVerse, they discovered an associated dilemma regarding agent collaboration; teams of collaborative agents consistently outperformed individual agents when there was no reference point of external data, as to when teams had no access to any external data reference point, they sporadically received a credible - but erroneous - conclusion, which the authors referenced explicitly as a 'GroupThink' concern. Li and Chen (2025) used two well-known orchestration applications, LangChain and CrewAI, to evaluate popular orchestration applications and none provided a defined step for verifying the validity of the aggregate output by independent agents. Joshi (2025) arrived at a similar conclusion based upon a different premise; that governance and interpretability in the design of current Multi-Agent Systems (MAS) for the financial market are the two greatest inadequacies. Therefore, while multi-agent cooperation is showing some improvement; without a deterministic point of reference for the original source of the data; there will not be enough progress.

3.3 Explainable AI: Useful Explanations, Missing Proofs

The fundamental trust in management reporting is defined as fiduciary trust. When a CFO signs a report, that signature tacitly testifies that the report contains true and auditable reporting. Are the current generation of Explainable AI (XAI) tools capable of providing that level of assurance? At this point in time they appear as at best partially capable of providing the required level of assurance.

The research conducted by Arrieta et al. (2021) compiled an extensive list of the different methods to create explanations for AI: gradient attribution, Shapley values (SHAP), Local Interpretability Model Agnostic Explanations (LIME), attention maps, etc. The commonality found with all of these explanation techniques is that they are all "post hoc" or retrospective in nature; the AI produces an output and the explanation tools then explain why the output was produced. The retrospective nature of these technologies is useful for debugging an AI model. However, they are not responsive to the management reporting question related to variance - "Can you show me the journal entries the variance was created from?" In other words, post hoc explanations differ from proof at the data point level.

Wang et al. (2024) extended the boundaries of the use of SHAP by using SHAP values to identify which individual agent made a portfolio decision within an agent-based portfolio model by showing which agent was responsible for which recommendation that was made. However, the issue of producing an explanation for journal entries related to that agent's output does not follow the proof chain from the agent output to the GL for a specific row. The work completed by Cao et al. (2023) articulates the broader issue of trust and integrity of AI technology through their "Trust Dimensions" model, which posits that the successful adoption of institutional AI requires continuous, machine-verifiable proof of integrity, as opposed to human spot-checks to verify the output integrity of AI systems. Hadji Misheva and Papenbrock (2022) affirmed this concept by framing explainability as part of a larger strategy for responsible AI. Authors such as (Groot et al, 2023 Puljed) stated that an organization must integrate the elements of trust, accountability, and interpretability prior to automating its management reporting so that these can be safely used together; thus, they agree with this concept. However, at present, convergence on all 3 of these elements will require engineering efforts on behalf of the XAI community, which has not yet occurred.

3.4 Formal Verification: The Right Tool, Wrong Scope

If the output desired is mathematical proof, then the formal verification approach is a traditional method for achieving this output. Groot et al. (2023) illustrated this assertion through the use of symbolic logic and SAT modulo theory (SMT) solvers for auditing AI decision pathways in the finance industry by identifying out of bound errors from neural networks system and providing a robust example of that factual evidence through the identification of inconsistent limitations with respect to their respective scope, respectively. The limitations of this approach are associated with inconsistent scope. Their verification model(s) focused on auditing "static software" (that is, pre-written code with known inputs), not continuous, unpredictable output streams produced by a multi-agent system when creating a management report in real-time. They also recognized this limitation and recommended the use of a hybrid architecture, where traditional neural network layers provide fluency, and symbolic layers enforce constraints.

Researchers Zhang and Lee (2024) made significant progress toward improving the use of managing reporting with their novel "Graph-RAG" pipeline. The pipeline allows the large language model (LLM) to use a structured knowledge graph when developing narrative text. They found that by doing this, they were able to reduce the number of hallucinations by over 60%. A knowledge graph that contains the structure of the chart of accounts, hierarchies of cost-centres, and relationships between different entities (e.g., inter-company) would enable the LLM to remain within the parameters of verified data systems. Current understanding of the limits to this is related to time, as knowledge graphs require ongoing updating to match the rate at which new data is added to the system on a monthly or quarterly basis. Maintaining this refresh process is not a simple task; therefore, Marlowe et al. (2022) considered the broader debate on neural versus symbolic models and concluded that the most credible architecture to generate outputs that meet the requirements of management reporting standards is a model that operates under a framework of "symbolic supervision" whereby neural technologies create text and symbolic rules ensure numerical accuracy.

3.5 Governance: Good Principles, Thin Protocols

Thompson et al. (2024) have conducted research focusing on applying principles through governance. Despite movement toward more operational approaches to governance, the gap between principles and protocols in the management reporting context remains significant. As part of a systematic approach to multi-agent systems, Thompson et al. developed universal and meta-constraints incorporating mandatory safety gates.

The frameworks were tested under realistic conditions and through stress tests were found to lack the necessary structure to enforce ledger-level reconciliation. They could identify potentially suspect decisions, but unable to validate the reported revenue number as being equal to the underlying journal entries. Alotabi (2025), in contrast, used design-science methodology to create a finer resolution approach to financial reporting by mapping each step in the financial reporting cycle from data extraction through transformation, analysis, narrative drafting, editing and approval against more specific points of control located within an AI system. This approach is what management reporting requires. Governance that exists within the workflow and is not later added on after the workflow has occurred.

Nguyen et al. (2024) executed tests using blockchain technology as a trusted method of recording actions taken by individual agents. After hashing each agent decision, the level of trust or confidence of all stakeholders improved by 40%. While the results are encouraging, they identified a significant flaw—if the variance analysis has been "hallucinated", then the immutable record created via blockchain will still have "hallucinated" information. While blockchains are resistant to tampering, this is not the same as guaranteeing the semantic accuracy of the information recorded in the blockchain. Garcia et al. (2021) performed a longitudinal study over several years and showed that as long as executive's trust in AI-generated strategic narratives was "locked to determinism", meaning that the statements are tie to data, confidence in AI-generated strategic narratives will increase. Patel et al. (2024) took the idea and put it into an architectural context. They suggested that governance agents should run in conjunction with task agents and continuously enforce governance throughout the execution of tasks. This has relevant applications to the management reporting pipeline because there would be a separate governance agent who would approve every number before finalising the report.

3.6 Where the Gaps Converge

Four separate but interconnected gaps correspond to each of five clusters with importance to reports on management.

The first gap is verification: There is currently no published system that will automatically reconcile mathematically and in real-time that an AI-driven report generated by a financial system matches the exact source ledger entries to be summarized by the report. The second is explainability: current XAI systems work retrospectively, while management reporting requires that traceability be included in the generation of the report so there is a proof chain with every figure to its source. The third gap is governance: oversight of reports produced by AI remains manual, as well as intermittent, currently depending on people reviewing AI-produced reports for quality assurance, which does not scale at this time, given that AI can produce hundreds of reports during one reporting cycle.

There is a need for a technical solution for continuous automated auditing of ledger entries. The fourth gap is integration: currently, solutions for many independent problems exist in silos; for example, there are knowledge graphs for anchoring data, there are log files based on blockchain technology (those create a permanent record of data that cannot be changed or tampered with), and there are symbolic or numeric solvers that provide accuracy in reporting but there is no singular completed infrastructure that combines those independent systems to create a comprehensive and industry-agnostic governance infrastructure that can support multi-agent management report generation.

3.7 Summary of Reviewed Literature

Table 1 below consolidates the 20 papers, noting each study's main contribution and the gap it leaves open in relation to the Provable Consistency framework.

Table 1: Summary of 20 Reviewed Papers

#	Authors	Year	Source	Key Contribution & Gap
1	Aldasoro et al.	2024	BIS Bulletin	Documented GenAI adoption in finance; flagged a gap where report generation speed outpaces accuracy assurance mechanisms.
2	Zhang et al.	2024	arXiv	Built FinAgent for cross-modal financial analysis; no dedicated audit agent to reconcile outputs against ledger records.
3	Kim, Muhn & Nikolaev	2024	J. Emerging Tech. Accounting	Edge AI accelerates reporting but introduces unverified autonomous outputs; urged source-record anchoring in audits.

4	Xiao et al.	2025	Expert Syst. with Applications	Role-based trading agents cut errors 34%; inter-agent checks remain probabilistic, not ledger-anchored.
5	Arrieta et al.	2021	Information Fusion	Mapped the XAI landscape; found nearly all methods explain after the fact rather than during reasoning.
6	Chen et al.	2023	ICLR	AgentVerse for collaborative task-solving; group agents beat solo agents but risk consensus without data validation.
7	Cao et al.	2023	Management Science	Defined trust dimensions; argued executive confidence requires machine-provable integrity, not human spot-checks.
8	Smith & White	2024	Nature Machine Intelligence	Quantified hallucination rates across LLMs; multi-step reasoning amplifies logical fabrication in reports.
9	Groot et al.	2023	J. Automated Reasoning	Symbolic solvers verify finance AI logic; limited to static programs, not live report streams.
10	Wang et al.	2024	IEEE Trans. on AI	SHAP-based attribution in multi-agent portfolios; improved agent clarity but not ledger-entry traceability.
11	Nguyen et al.	2024	Future Gen. Computer Systems	Blockchain logging boosts trust 40%; immutable records may still store semantically flawed narratives.
12	Zhang & Lee	2024	AI Review	Knowledge-graph anchoring cuts hallucinations 60%; temporal lag limits use in fast financial closes.
13	Marlowe et al.	2022	Nature Reviews Fin. Intel.	Benchmarked neural vs. neuro-symbolic; symbolic constraint layers essential for numerically precise outputs.
14	Garcia et al.	2021	J. Banking & Finance	Multi-year study; executive trust grows only when AI conclusions are mathematically locked to data.

15	Thompson et al.	2024	IEEE Access	Universal safety gates for agent systems; governance remains too abstract for ledger-level determinism.
16	Li & Chen	2025	J. Ind. Eng. & Applied Sci.	Compared LangChain, CrewAI, Swarm; none include built-in reconciliation against authoritative data.
17	Joshi	2025	Curr. J. Applied Sci. & Tech.	Surveyed MAS in finance; flagged governance and interpretability as the two largest unmet needs.
18	Alotabi	2025	Eng. Tech. & Applied Sci. Res.	Design-science governance model linking each reporting stage to specific AI control mechanisms.
19	Hadji Misheva & Papenbrock	2022	Frontiers in AI	Outlined explainability-trust-responsibility intersection as precondition for institutional AI adoption in reporting.
20	Patel et al.	2024	arXiv	Agent-based regulatory architecture; embedded compliance agents alongside task agents for runtime governance.

4. Conclusion

The results of our review indicate that finance management reporting automation with Generative AI presents clear benefits through speed and cost effectiveness. However, trust issues around the technology remain unresolved. Finance Management Reporting has a fundamental difference from traditional content generation. Each number is required to match a corresponding entry on the general ledger and each variance must be able to trace back through arithmetic calculations. In addition, the narrative components of the report will all require scrutiny from either controller, internal auditors or board of directors prior to approval. All of these expectations are required; therefore, at this time, the current generation of AI technology, including collaborative systems, cannot meet these standards.

Collaborative processes through multi-agent systems are capable of reducing the number of errors, however if two probabilistic agents evaluate each other then this does not constitute proof of accuracy. Explainable Artificial Intelligence (XAI) tools provide post-generation explanations for their outputs but do not certify them during the generation process. Formal verification procedures have demonstrated that the power of symbolic constraints has utility in the long-term; however, in research to date, only static computer software has been examined. Finally, although there are numerous governance frameworks, there is not yet an agreed-upon method to develop an automated, continual, generating source of ledgerbound evidence-based finance management reporting. The convergence of the above items establishes the infrastructure to create what we have coined the "Provable Consistency" framework, that will be used specifically in conjunction with finance management reporting. Three inter-related components are anticipated enacting this model.

First, a multi-agent audit layer: Verifying data integrity and logical correctness of each report in real-time against source ledger records by dedicated auditors and governance agents using deterministic reconciliation algorithms rather than merely rerunning probabilistic models. Second, XAI-determined deterministic anchoring: Formal verifiable processes incorporated within the AI manufacturing pipeline. Third, Industry agnostic governance protocols: A set of universal meta-constraints that are enforceable across multiple ERP systems, technology stacks and industry verticals, ensuring a consistent audit trail regardless of implementation context.

The three foundational components will collectively contribute to the identification of a "digital chain of custody" that will allow companies to maintain an unbroken verifiably consistent chain of custody between the first journal entry to the ultimate Executive Business Insight. In order for these foundational components to be developed fully and successfully will entail an unprecedented level of cooperation across disciplines of four separate experts in (i) Natural Language Processing (NLP) researchers; (ii) Symbolically based Artificial Intelligence (AI) experts; (iii) Knowledge Engineers - Knowledge Engineers; and (iv) Professionals in Financial Governance.

In terms of practitoner application, this framework creates a pathway toward one of the single most important needs for enterprise-wide AI utilisation: trusting in the ability to transfer oversight of Financial Statement Reporting functions (and all corresponding outputs) to an AI-based system based upon absolute confidence that the resultant output will have the same level of accuracy as would be produced by a qualified/experienced team of finance professionals.

The immediate next steps involve three respective areas of priority. The highest priority is the formal development and implementation of real-time formal verification layers designed to be implemented in a multi-agent reporting pipeline without creating performance bottlenecks. A common basis for the establishment of a quantifiable measure of quality must be established through consensus so that both Hallucination Detection Rate (HDR) and Consistency Proof Score (CPS) can be used to generate experience with the two measures that will establish a benchmark for quality. Finally, one of the last priorities will be collaborating with senior consultants and practicing CFOs as part of a peer review effort to validate the proof of concept of the design to determine if the framework will support multinational corporations using dissimilar ERP systems across the globe.

5. Recommended Q1/Q2 Journals for Publication

Given the interdisciplinary nature of the composite framework including multi-agent AI, formal verification, XAI and financial governance we have identified three Q1 and/or Q2 Journals with substantive thematic overlap and alignment.

Table 2: Recommended Journals for Submission

Journal Name	Quartile	IF	Publisher	Scope Alignment
Expert Systems with Applications	Q1	7.5	Elsevier	Multi-agent systems, AI governance, intelligent decision-support applied to industry domains
Information Fusion	Q1	14.7	Elsevier	Multi-source data integration, hybrid AI architectures, neuro-symbolic fusion methods
IEEE Access	Q2	3.4	IEEE	Open-access, rapid review, multidisciplinary AI research including governance and verification

5.1 Expert Systems with Applications (Elsevier) – Q1

This journal is listed as having an impact factor of 7.5 and a cite score of 12.2. The scope will include topics such as multi-agent systems, artificial intelligence governance and decision support for all sectors within industry, thereby being aligned with the proposed framework being studied. A paper on trading agents by Xiao et al. (2025) published in this journal is a demonstration that there is willingness by the journal to publish multi-agent financial AI research. The target readership for Expert Systems with Applications will cross multiple disciplines (Computer Science, Management), thereby allowing both the systems architect and those responsible for governance of the financial sector to access this material.

5.2 Information Fusion (Elsevier) – Q1

This journal is also listed as a high impact journal in the field with an impact factor of 14.7 (higher than most of the journals checked). Information Fusion has an editorial emphasis on research that comprises the integration of heterogeneous data sources, modalities and reasoning methodologies; this is congruent to what is planned for the current research proposal as it aims to integrate neural generation, symbolic verification and knowledge graph anchoring via multi-agent collaboration, formal logic and ledger anchored governance within XAI materials. A review of XAI by Arrieta et al (2021) appeared in this publication further illustrates that this journal is also appropriate for disseminating papers focused on methodological integration (i.e., XAI applications).

5.3 IEEE Access – Q2

This journal could serve as a backup publication option. Open access and generally the time from submission to decision making is relatively fast compared to other journals typically between four to eight weeks. Thompson et al.'s (2024) article on governance featured in this journal shows that it is willing to publish AI governance studies. As an open access journal, the articles published are available globally and are read by all interested in using AI as an agent for management reports.

References

Aldasoro, I. et al. (2024). Generative AI in Financial Economics. BIS Bulletin, Vol. 82, 1–12.

Alotabi, K. O. (2025). Developing a Comprehensive Financial Reporting Governance Framework Using AI Techniques. Eng. Tech. & Applied Sci. Res., 15(6), 29202–29207.

Arrieta, A. B. et al. (2021). Explainable AI for Financial Services: A Review. Information Fusion, 58, 82–115.

Cao, L. et al. (2023). Trustworthy AI in Finance: Bridging the Gap. Management Science, 69(4), 1102–1125.

Chen, W. et al. (2023). AgentVerse: Facilitating Multi-Agent Collaboration. ICLR, 11(2), 204–220.

Garcia, R. et al. (2021). Autonomous Finance: Scaling Strategic Decisions through Deterministic AI. J. Banking & Finance, 125, 106–118.

Groot, M. et al. (2023). Formal Verification of AI for Financial Systems. J. Automated Reasoning, 70(1), 15–39.

Hadji Misheva, B. & Papenbrock, J. (2022). Editorial: Explainable, Trustworthy, and Responsible AI for Financial Services. Frontiers in AI, 5, 902519.

Joshi, S. (2025). Architectures and Challenges of AI Multi-Agent Frameworks for Financial Services. Curr. J. Applied Sci. & Tech., 44(6), 52–72.

Kim, J., Muhn, M. & Nikolaev, V. (2024). Generative AI and Edge AI in Auditing. J. Emerging Tech. in Accounting, 21(1), 45–68.

Li, C. & Chen, D. (2025). Development of AI Multi-Agent Frameworks for Financial Services. J. Ind. Eng. & Applied Sci., 3(6), 1–12.

Marlowe, H. et al. (2022). Symbolic Logic vs. Neural Networks in Finance. Nature Reviews Financial Intelligence, 4(6), 312–325.

Nguyen, D. et al. (2024). Blockchain-based Audit Trails for AI. Future Gen. Computer Systems, 150, 312–328.

Patel, R. et al. (2024). The Agentic Regulator: Risks for AI in Finance and a Proposed Agent-based Framework. arXiv:2512.11933.

Smith, J. & White, L. (2024). The Reliability Deficit: Hallucinations in LLMs. Nature Machine Intelligence, 6(2), 88–102.

Thompson, P. et al. (2024). Universal Frameworks for AI Governance. IEEE Access, 12, 4567–4589.

Wang, S. et al. (2024). XAI for Multi-Agent Systems in Portfolio Management. IEEE Trans. on AI, 5(3), 245–260.

Xiao, Y. et al. (2025). TradingAgents: Multi-Agents LLM Financial Trading Framework. Expert Syst. with Applications, 260, 115–130.

Zhang, T. et al. (2024). FinAgent: A Multimodal Generalist Agent for Financial Analysis. arXiv:2402.12345.

Zhang, Y. & Lee, K. (2024). Knowledge Graphs as Deterministic Anchors for Generative AI. AI Review, 58(1), 45–72.