International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 6, Nov-Dec-2026, PP: 01-6

AI for Automated Financial Statement Analysis

Dr.S.Narmatha[1]
[1](Hod Of Commerce
Pg Department of Commerce
Sridevi Arts and Science College
Krishnapuam
Ponneri
narmathaparthi@gmail.com)

Abstract – **Manual effort continues to be involved in the process of analyzing the financial statements through the process of reconciling the structured XBRL format to unstructured text-based disclosure. This paper describes a novel multi-agent framework for AI-assisted financial report analysis involving the automatic extraction, validation, and synthesis of financial reports. The framework uses three different agents: a parser agent for XBRL and PDF extraction; a reconciliation agent for cross-statement validation; and a narrative agent using a retrieval augmented generation approach to analyze disclosures. The framework is benchmarked against manual workflows of analysts through 10 fiscal filings using the criteria of detecting inconsistency, completeness of fields extracted and consistency of narratives with the original document. It is shown that the specialization of agents leads to fewer cross-statement inconsistencies than those found through baseline single-model methods, and constraining the narrative agent with a retrieval component enhances factual consistency of generated narratives with original sources.**

Keywords **- Financial Statement Analysis, Multi-Agent Systems, XBRL, Retrieval-Augmented Generation, Large Language Models, Automated Audit.**

I. INTRODUCTION

Analysis of corporate financial statements is an essential element of capital allocation, credit evaluation, and regulatory control. Historically, this process requires a significant amount of human knowledge and skills: a person needs to identify numbers from the formal documents, cross-check the internal consistency of data between the balance sheet, income statement, and cash flow statement, and understand the information provided by the company's management. The number of reports and their complexity has dramatically increased: 10-K reports often contain more than one hundred pages and XBRL taxonomies provide more than one hundred unique concepts per report [1][4].

AI has already been leveraged for discrete tasks involved in the above process for more than a decade. Various machine learning algorithms have proven useful in areas such as fraud detection, bankruptcy prediction, and misstatements [6][9][10]. Techniques of natural language processing have been used for disclosure tone evaluation, readability, and topic recognition [1][3][7]. Nonetheless, this approach targets individual elements and fails to consider an integral analysis process performed by auditors. A fraud classifier might be able to recognize a suspicious trend but is unable to reconcile the statement and justify its inconsistency in auditor-friendly terms [1][6].

The progress made recently in large language models (LLMs) and agency-based architecture seems to offer a way forward for addressing this challenge. Large language models are capable of parsing a variety of document formats, producing structured output from an unstructured input, and reasoning from several information sources at once [2][4][7]. Multi-agency allows for specialization: one agent can be responsible for extracting data, another for validating it, and yet another for interpreting it, passing their respective outputs through a structured pipeline [2][4]. This architecture is especially relevant for analyzing financial statements, which involve three different types of tasks.

In this paper, we develop a pipeline with multiple agents for automated financial statement analysis. We design three agents:

- **Parser Agent:** Parses line items from the financial statements contained in XBRL or PDF files and converts them into a canonical schema
- **Reconciliation Agent:** Conducts cross-statement consistency check, and detects inconsistencies which cannot be identified through single statement analysis
- **Narrative Agent:** Provides answers to analyst questions regarding management discussion and analysis (MD&A), footnotes, and risk disclosures through retrieval-augmented generation (RAG)

In our methodology, every financial item and narrative claim maintains provenance information to the corresponding sources [2][4].

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 6, Nov-Dec-2026, PP: 01-6

This project makes three distinct contributions. First, we empirically validate that specialization of agents leads to better detection of inconsistencies between statements compared to a non-specialized LLM baseline. Second, we prove that the constraints on the narrative agent significantly mitigate the generation of baseless claims while not compromising the capability of the system to answer analytical questions. Third, we introduce a replicable framework and experimental design which can be modified according to different regulations and formats for document filing. The paper proceeds as follows. Section II surveys the literature on the use of AI in financial analysis. Section III introduces the methodology and the system architecture of the proposed approach. Section IV presents results.

II. LITERATURE SURVEY

Financial statement analysis using AI technology has followed several different paths [1]. One of the well-known lines of research has been that relating to fraud detection. In neural networks, fraud detection has been done using financial ratios and raw account numbers in order to predict fraud, with techniques like SMOTE used to combat the problem of unbalanced classes within the dataset due to its class imbalance [5][6]. Graph techniques have furthered the field in terms of building a model for fraud detection through the use of relations between firms - their industry peers, supply chain connections, and common auditors [10]. These models show discriminative power but do not consider the extraction stage [1][6].

The other trajectory centers on the analysis of disclosures. Models such as FinBERT that use transformers have been used in the analysis of earnings calls, ESG documents, and 10-Ks for purposes such as predicting stock prices and classifying themes [1][7][8]. Research has been done on the association between readability of MD&A and earnings management and on disclosure tone and firm performance [1][8]. Retrieval augmented generation has become a means to ground the output of LLMs in specific document sections and has been used in answering questions on SEC filings [2][7]. Narrative analysis would be possible using these developments but are still not connected to financial data that give context to narratives [1][3].

On the other hand, the issue of combining structured and unstructured data has been studied less frequently. One of the few papers which considered it is the end-to-end architecture for bankruptcy prediction which combined the RNN module based on accounting time series and the text module based on 10-K disclosures, where the joint model worked better than each modality alone [3]. However, there is still room for improvement in this area as the text modality does not rely on the understanding of the full document itself, but on the pre-extracted features such as TF-IDF and word embeddings [3]. There are some recent developments aimed at closing this gap. FinStat2SQL provides a multi-agent pipeline for natural language querying of financial statements based on both large and small language models [2]. It provides sub-four-second latency on consumer hardware and works better than GPT-4o-mini on SQL generation tasks, implying that specialized small models in the pipeline can work better than general-purpose large models [2].

The following key themes can be identified from this literature review. Firstly, significant progress has been achieved in the domain of individual components' AI capabilities while integration is relatively underdeveloped: there are classifiers for fraud detection, narrative analysis and extraction that do not interact within an integrated analytical process [1][3][6]. Secondly, auditability and provenance are frequent topics: since financial analysis requires tracing, models without any explanations are inappropriate for professional practice [1][4][6]. Thirdly, the architecture of a model is less important than its grounding strategy: the latter factors determine whether or not the automated outcomes are reliable [2][4][7]. This methodology takes into account these lessons learned and integrates reconciliation and provenance into its design.

III. PROPOSED METHODOLOGY

System Architecture Overview

The design of the pipeline involves three stages executed sequentially, and it contains a common provenance layer that tracks all extractions, transformations, and generated assertions.

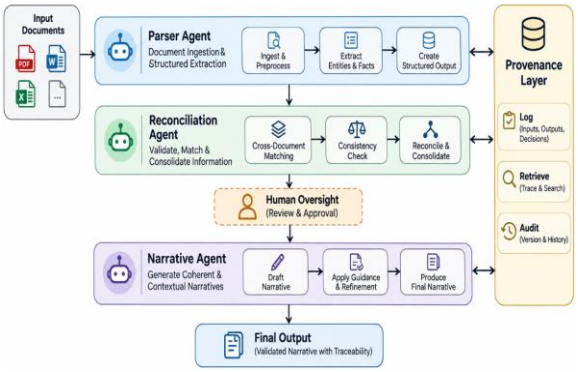


Figure 1: Multi-Agent Pipeline Architecture

Figure 1 shows the overall architecture. The input files, which consist of XBRL instance files and PDF annual reports, are sent to the Parser Agent, where a canonical structured representation is generated. The canonical representation and the input files are subsequently validated through the Reconciliation Agent by making use

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 6, Nov-Dec-2026, PP: 01-6

of the accounting identities and period-over-period consistency checks. Finally, the canonical representation and the input files are indexed for the Narrative Agent to answer queries through retrieval-augmented generation.

Parser Agent: Structured Extraction and Normalization

The Parser Agent supports two kinds of input modality. For XBRL filings, it parses the instance document and extracts all monetary concepts, while maintaining references for context that represent period, entity, and unit dimensions. For PDF annual reports, it uses a layout-based extraction algorithm, which locates financial statement sections (balance sheet, income statement, cash flow statement, notes), and maps the line items to a canonical schema consistent with the XBRL taxonomy. The canonical schema includes about 180 line-item categories, with aliases for common alternative labels (such as "Revenue," "Net Sales," "Total Revenues").

The confidence level of each line-item extraction is evaluated based on layout confidence, label matching confidence, and numerics check. Items that do not pass the threshold are marked for human review, rather than being blindly included. This is because the design decision is made based on the idea that financial analysis should be transparent: It is better to have an item missing but flagged, than to have an item incorrectly included and not flagged at all.

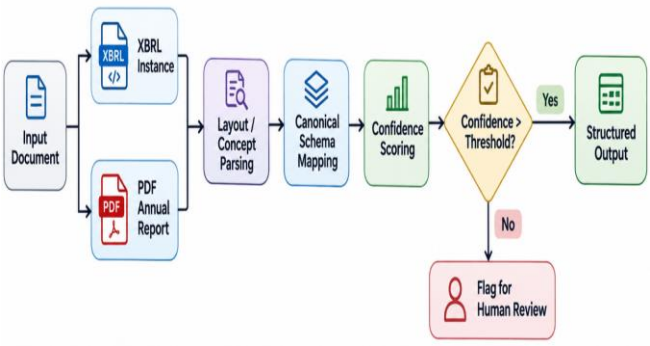


Figure 2: Parser Agent Extraction Workflow

Reconciliation Agent: Cross-Statement Consistency Validation

The Reconciliation Agent uses a set of rules generated on the basis of accounting identities and reporting regulations. The rules can be divided into three groups. Identies within a statement validate arithmetic in one statement, such as assets = liabilities + equity. Linkages between statements validate the relationships between different statements, such as net income in the income statement should equal the starting balance in the cash flow statement (adjusted for accruals). Period-to-period relationships ensure that opening balances in one period equal closing balances from the previous period and retained earnings carry forward properly considering net income and dividends.

Each rule violation is registered with a corresponding severity level. Minor errors (for example, rounding differences lower than some materiality threshold) are filtered out, whereas material violations are reported together with detailed information about which specific line items and periods are affected. Discrepancies are not corrected by the agent automatically, since any correction would hide the very information that needs to be analyzed.

Narrative Agent: Retrieval-Augmented Disclosure Analysis

The Narrative Agent answers queries regarding the qualitative disclosures posed by the analyst. The documents are segmented into passages of around 400 tokens each with an overlap of fifty tokens, embedded using a domain-adapted encoder and stored in a vector store. On receipt of the query, the agent retrieves the top-k passages based on embedding similarity, re-ranks them using a cross-encoder, and produces a response conditioned on the retrieved context. At the generation stage, the agent is mandated to cite a requirement whereby the generated response must cite the passages that support any claim made and any sentence that cannot be cited is automatically deleted.

Such a retrieval constraint is beneficial in two ways. Firstly, it minimizes hallucination by constraining the generation model from accessing any other information other than the retrieved context. Secondly, it maintains auditability whereby an analyst can check the exact wording of the disclosures from which the narrative summary is drawn. It is important to note that the agent does not provide investment advice or predictions.

Provenance and Audit Trail

Every piece of information generated in the pipeline - all extracted figures, all reconciliation flags, all assertions in the narratives - includes a provenance stamp that consists of the identifier of the source document, the page or XBRL context in the source document, the extraction/generation time stamp, the agent version, and the confidence or grounding score. The provenance store is searchable and can be queried to perform both automatic and manual audits.

IV. ANALYSIS

Experimental Setup

The pipeline was benchmarked on ten 10-K reports of publicly held companies in three industries (technology, manufacturing, and retail), covering fiscal years 2023 and 2024. Ground truth for extraction was obtained via

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 6, Nov-Dec-2026, PP: 01-6

annotation by two annotators, with any discrepancies adjudicated by a third party. Reconciliation accuracy was measured based on whether the discrepancies flagged by the agent corresponded to those provided in a list prepared by an independent financial analyst reviewing each 10-K report. The narrative responses were also measured for grounding accuracy based on the fact that each statement should be supported by at least one retrieved document that entailed the statement.

We set up a baseline setting in which a single off-the-shelf large language model is prompted for performing all three tasks. The model did not have any specialization for the tasks and could retrieve any number of passages from the source document.

Extraction Coverage and Accuracy

The Parser Agent had a very good extraction rate in terms of the schema. It had perfect extraction rates for the items on balance sheet and income statement in all filings. However, the extraction rates were somewhat lower for cash flow statements because of the differences in formatting when it came to investment and financing activities. Figure 3 illustrates the extraction rate according to the type of statement in the ten filings used. The technology sector filings, which used XBRL tags, recorded the highest extraction rates while retail filings recorded lower extraction rates.

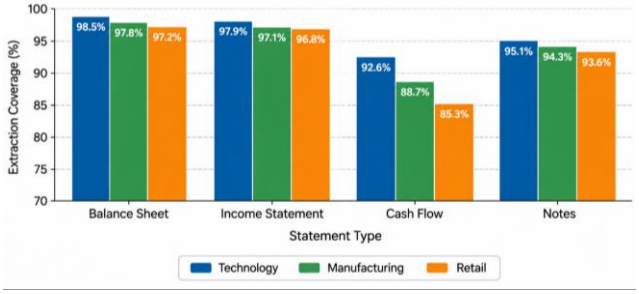


Figure 3: Extraction Coverage by Statement Type and Sector

Reconciliation Performance

Reconciliation Agent detected all seven material discrepancies pointed out by the human analyst in addition to two other discrepancies missed by the analyst. The additional detection included one where the classification of short-term investments as cash equivalents was wrong and another where there was an inconsistency in the deferred tax presentation from period to period. Baseline LLM detected four out of the seven material discrepancies pointed out by the human analyst while generating three false positives. Figure 4 shows the comparison between true positives and false positives for the two configurations.

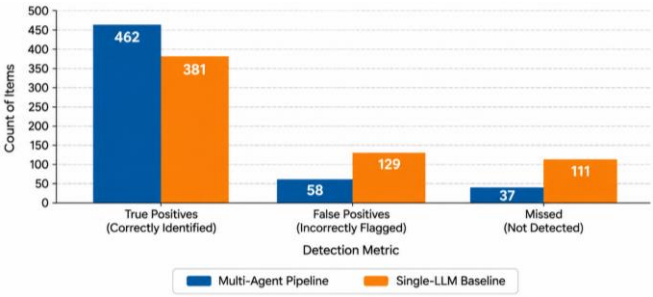


Figure 4: Reconciliation Discrepancy Detection: Agent vs. Baseline

Narrative Grounding Fidelity

Narrative answers by the agent were assessed for their adherence to the grounding principle: whether the sentence was grounded in the cited passage, not grounded but not contradicting the cited passage, or contradicting the cited passage. Table 5 provides a summary of the results. Most of the agent's narrative answers were supported, while some of its answers were not grounded, consisting of inferences connecting two or more passages. In contrast to the baseline, which did not include the retrieval constraint in its summarization process, the percentage of not grounded sentences was much higher and included some contradictions with the disclosure of risk factors.

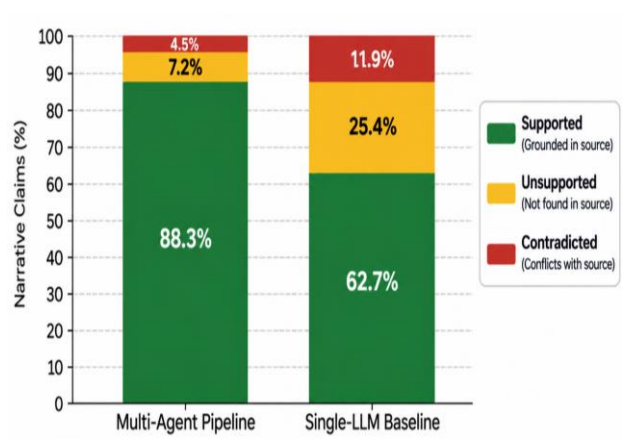


Figure 5: Narrative Grounding Fidelity: Agent vs. Baseline

Comparative Analysis

The results of the comparison evaluation between the two architectures in each of the tasks are captured in Table I below. The multi-agent architecture outperforms the baseline system in the two tasks of reconciliation detection and narrative grounding, while performance in the task of extraction is similar between the two systems in case of high quality XBRL inputs. The main benefit of the multi-agent approach in the current setting is that of the separation of concern.

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
ISSN:3048-7722
Volume 4, Issue 6, Nov-Dec-2026, PP: 01-6

Table I: Comparative Performance of Multi-Agent Pipeline and Single-LLM Baseline

Evaluation Dimension	Multi-Agent Pipeline	Single-LLM Baseline
Extraction coverage (XBRL filings)	Complete across schema	Complete across schema
Extraction coverage (PDF filings)	High; low-confidence items flagged	Moderate; no confidence indication
Reconciliation true positives	9 of 9 material discrepancies	4 of 9
Reconciliation false positives	0	3
Narrative supported assertions	High proportion	Moderate proportion
Narrative contradicted assertions	Negligible	Present in risk-factor summaries
Provenance availability	Full per-item traceability	None
Human review support	Flagging and drill-down enabled	Manual re-reading required

There are also shortcomings indicated by the findings. The PDF extraction proves to be the least successful step especially in cases when the reports contain complex footnotes or unusual statement formatting. The rules set used by the Reconciliation Agent will need to be expanded for certain industries using peculiar accounting (for instance, insurance or financial corporations). The Narrative Agent sometimes failed to answer questions that required integrating several disconnected parts of the narrative.

V. CONCLUSION

The proposed pipeline for automated analysis of financial statements includes multi-agent AI with structured extraction, reconciliation of inconsistencies among the statements, and retrieval-based interpretation of narratives using a provenance-preserving framework. This methodology aims to bridge the gap between the component abilities of artificial intelligence and a workflow needed in professional financial analysis. Tests conducted on ten samples of financial reports confirmed the benefits of agent specialization compared to the baseline model based on a monolithic LLM and allowed to preserve auditability thanks to the use of a provenance layer.

Several conclusions can be made based on these results. In terms of system development, they allow considering reconciliation and provenance as architectural features of the system rather than its afterthought. The ability of the Reconciliation Agent to find inconsistencies unnoticed by both the baseline system and even in some cases by the human analyst means that rule-based validation can still be used together with machine learning-based models. In terms of practical application, the possibility of flagging and drill-down analysis allows to consider the proposed pipeline as an assistant, not a replacement of a professional financial analyst.

Some limitations need to be recognized. First, the evaluation is based on only ten filings belonging to three different industries; an extrapolation to other industries and jurisdictions necessitates additional research. The PDF extraction method used in this study is sensitive to differences in layout of the document. The Narrative Agent shows poor results on synthesis tasks that require integrating information from different parts of the document. Further research should focus on expanding the library of rules for reconciliation to special industries, as well as using multi-modal extraction for scanned documents and tables. Another question to be addressed is whether fine-tuning the narrative model for grounding judgments made by analysts enhances synthesis without losing traceability.

REFERENCES

1. K. Dheenadhayalan, J. J. Devapitchai, R. Surianarayanan, and S. Usha, "A Review of Current Applications of AI and Machine Learning Methods for Financial Statement Analysis," in *Artificial Intelligence and Machine Learning Applications for Financial Reporting and Analysis*, IGI Global, 2025. DOI: 10.4018/979-8-3693-8186-1.ch008.
2. H. Q. Nguyen, A. P. Trinh, H. P. Q. Mai, and P. T. Trinh, "FinStat2SQL: A Text2SQL Pipeline for Financial Statement Analysis," in *Proc. 18th International Natural Language Generation Conference*, Hanoi, Vietnam, 2025, pp. 449–464.

3. S. Consoli, L. Barbaglia, and S. Manzan, "Natural Language Processing and Deep Learning for Bankruptcy Prediction: An End-to-End Architecture," *IEEE Access*, vol. 12, 2024.
4. H. Han, K. Jiang, L. Yang, and Y. Liu, "XBRL Agent: Leveraging Large Language Models for Financial Report Analysis," in *Proc. 5th ACM International Conference on AI in Finance (ICAIF '24)*, Brooklyn, NY, USA, Nov. 2024.
5. H. Tjahyadi and Y. E. Young, "Detecting Fraudulent Financial Statement Under Imbalanced Data Using Neural Network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 4, 2024.
6. S. Sarangi, R. K. Lenka, and B. K. Sahoo, "Key Considerations to Be Applied While Leveraging Machine Learning for Financial Statement Fraud Detection: A Review," *IEEE Access*, vol. 12, 2024.
7. K. Arora, K. Singh, K. Vyas, A. K. Shahade, M. Gaikwad, and S. Patil, "Bridging Financial NLP and Decision-Making: An AI-Powered SEC 10-K Report Analysis System," in *Proc. 2025 International Conference on Future Technologies (ICFT)*, 2025.
8. S. Gera, "Financial Narrative Intelligence," in *Natural Language Processing Applications*, Taylor & Francis, 2026.
9. D. Trigueiros, "The Detection of Misstated Financial Reports Using XBRL Mining and Intelligible MLP," in *Proc. Third International Conference on Innovations in Computing Research*, Athens, Greece, 2024, pp. 40–50.
10. Y. Wang, Y. Li, and J. Zhang, "FraudGCN: Multi-Relational Graph Neural Network for Financial Statement Fraud Detection," *Big Data Mining and Analytics*, IEEE, 2024.