Preprints.org

Article

Not peer-reviewed version

Artificial Intelligence–Driven Financial Reporting: Transforming Customized Financial Statements, Decision Intelligence, and Assurance Practices

Ashif Anwar *

Posted Date: 20 October 2026

Keywords: accounting information systems; explainability; decision usefulness; embedded audit trails; trial balance; human – AI collaboration; design science research

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Artificial Intelligence–Driven Financial Reporting: Transforming Customized Financial Statements, Decision Intelligence, and Assurance Practices

Ashif Anwar

Wolters Kluwer, Coppell, TX, 75019, USA; reachanwar@gmail.com

Abstract

Financial statements are prepared from trial balances that have passed internal-control checks and, for external reporting, have been audited before stakeholders rely on them. Converting this verified ledger into the views needed by investors, creditors, regulators, and management remains largely manual, causing rework, delays, and duplicated reconciliation. This study designs and evaluates an artificial intelligence (AI)-coordinated architecture that generates stakeholder-specific yet reconciled reports from a single governed trial balance. Following a design science approach, the architecture combines centrally configured accounting policies and KPI definitions, lineage metadata, account classification, anomaly detection, sequence forecasting, financial natural language processing (NLP), role-aware customization, and human-in-the-loop explainable AI (XAI). The prototype was evaluated on 15,000 trial balances for 500 entities (12,000 synthetic and 3,000 derived from public SEC filings) against rule-based, robotic process automation (RPA), and commercial generative-AI baselines. It achieved a line-item disclosure fidelity (F1) of 0.965, an AUC of 0.97 for misclassification detection, and a mean absolute percentage error (MAPE) of 4.8% for one- to four-quarter forecasts; in 250 simulated multi-entity scenarios, preparer effort fell from 38.0 to 4.2 person-hours per reporting cycle. Simulated stakeholder and auditor responses illustrate the expected gains in perceived usefulness and trust. Limitations include synthetic and statement-derived data, simulated rather than observed perception measures, and an unevaluated narrative-generation module.

Keywords: accounting information systems; explainability; decision usefulness; embedded audit trails; trial balance; human – AI collaboration; design science research

1. Introduction

Artificial intelligence (AI) in financial reporting and assurance is an emerging paradigm that signifies a revolution in financial information processing, analysis, and reporting. An AI-powered auditing and regulatory compliance system is now being leveraged for enhanced monitoring, lower compliance risk, and to support regulatory innovation without compromising regulatory accountability [1]. AI in banking has been used in personalized financial advice, customer-facing decision support, demonstrating the capability of intelligent systems to enhance the responsiveness of services, customer engagement, and trust in financial institutions [2]. In addition to the adoption of AI, enterprise resource planning (ERP) systems also serve as the structured, transactional backbone for financial reporting, ensuring that all financial, accounting, procurement, sales, and operational data are unified in common financial reporting environments. Based on previous studies, ERP implementation can facilitate transparency, real-time access to financial data, automation of processes related to financial and non-financial reporting, and higher financial and non-financial reporting quality [3]. Leading ERP vendors, such as SAP, Oracle, Microsoft, and Workday, have increasingly integrated machine-learning solutions for essential close functions, such as account reconciliation,

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

journal-entry anomaly detection, and cash forecasting, into their ERP offerings, and are now bringing generative assistants to the ERP close. However, these functions usually work as individual services within a vendor-specific reporting schema and not as a governed layer that is shared by multiple stakeholder groups from one ledger.

The influence of AI and ERP-driven digital infrastructure does not only reach how these reports are generated, but also the quality of the financial reporting. In capital markets, AI can be used to enhance the precision, promptness, and transparency of financial reporting, which can provide investors and regulators with better information to make choices [4]. Similarly, it has been found that the implementation of SAP ERP is related to improvements in financial reporting and accounting information systems [5]. AI disclosure practices also have an impact on the credibility of financial statements, especially in the banking industry, where the quality of disclosure is essential for investors' trust and institutional performance evaluation [6]. In the case of creditor-oriented and banking-sector reporting, Basel III capital requirements are hinged on credit-risk classification, measurement of exposures, assessment of risks, and capital adequacy and liabilities evaluation as part of regulatory assessments and financial-statement interpretation [7].

Furthermore, the use of AI in making decisions for financial statements is a strategic move towards offering personalized reporting and assurance. The use of AI in financial reporting systems can bring about new business value through automated data analysis, predictive modeling and anomaly detection [8]. Strategies for using AI to improve reporting accuracy and regain public trust highlight the importance of strategic planning in AI deployment [9]. The literature on AI in financial statement analysis, corporate finance, and financial technology is expanding to highlight the ability of AI applications to streamline operations and aid sustainable development and decision-making [10–12]. Furthermore, as digital transformation projects grow in emerging markets, AI is plays a crucial role in strengthening accountability and transparency, adding to its significance in the modern financial landscape [13]. Additional opportunities for integration can be seen in emerging natural language processing and information retrieval techniques that can make non-financial disclosure and reporting frameworks fit in with existing accounting processes with ease [14]. These innovations, taken together, underscore the transformative role of AI in enhancing intelligent financial statements, providing personalized insights and assurance mechanisms that reimagine traditional reporting landscapes.

AI can also help with sustainability and stakeholder-based disclosure [15] and large language models can be used to predict earnings changes based on standardized statements [16].

AI in reporting and assurance raises governance issues beyond technology’s accuracy. An algorithm is simply a tool and does not take legal responsibility for these statements; therefore, responsibility for the statements will lie with management and those in charge of governance, and the auditor will retain the responsibility for the opinion; therefore, it is essential to have documentation of who has approved each one that is created automatically. Issues raised as ethical concerns include biased training sets, lack of transparency in model workings and use of model outputs, and reskilling and expected job replacement for accountants [17, 18]. Deep learning models can be trained with new domain patterns and can be used to detect atypical or more complex transactions that cannot be covered by fixed rule sets [19], but their performance drops when the domain patterns change; thus, their knowledge must be supervised and updated by humans.

However, the implementation of AI in core reporting also has its challenges. These include data bias, audit limitations, and susceptibility to cybercrime [20]. In addition, the literature is still scattered and does not in-depth analyze integrated reporting ecosystems [20,21]. These gaps give rise to the critical question of research: How far can an AI-coordinated system be used to move from trial balances to personalized financial statements, thereby providing useful and faithful decisions with respect to the standards of representational verifiability and assurance? This transformation from ERP-based standardized reporting to an AI-orchestrated architecture with outputs for analysts, lenders, regulators and boards, and embedded support, assurance, explanation and security controls is captured in Fig 1.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

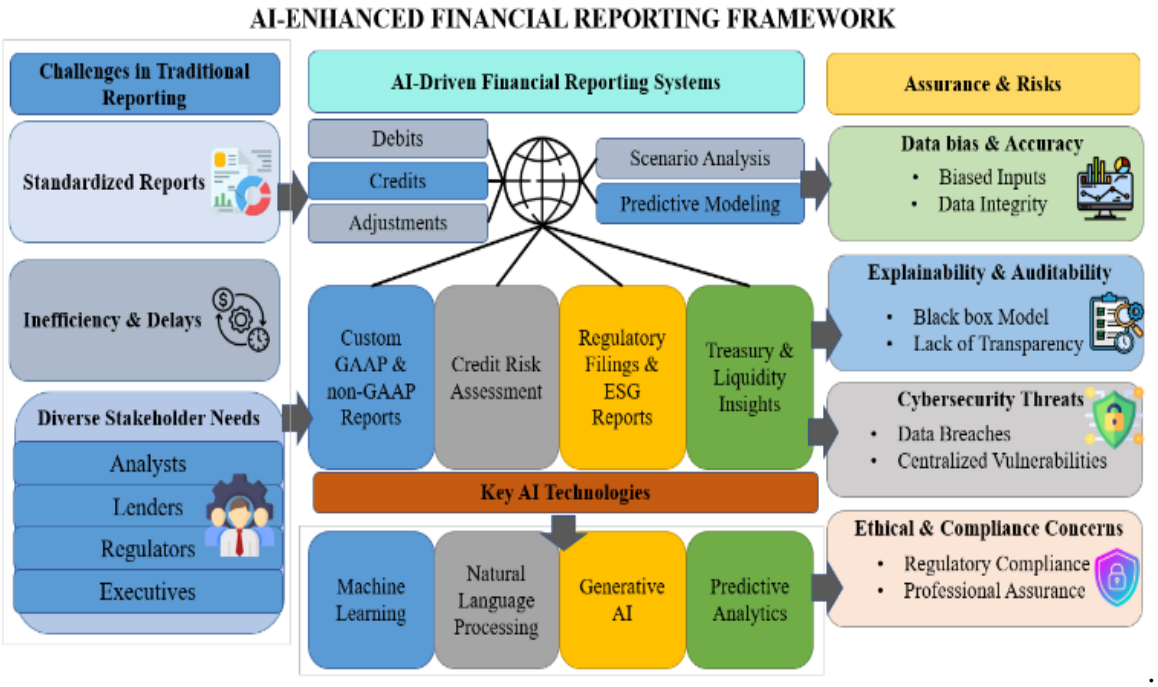


Figure 1. AI-Coordinated Financial Reporting Framework: From Template-Driven Compliance to Stakeholder-Specific Decision Usefulness.

Thus, this study aimed to develop, set up, and test an AI-driven reporting framework that provides stakeholder-specific views from a single, control-checked trial balance that remains reconciled with the statutory financial statements. Its contribution is more than merely a combination of existing technologies. First, customization is limited to the presentation layer that is subject to one set of accounting policies and set of key performance indicators (KPI) definitions, and as a result, each view is by construction reconcilable. Second, the audit trail (lineage hashes, explanation logs, and override records) is built into the reporting process, extending and building on embedded-audit and explainability techniques that have been developed in other accounting contexts [22,23]. Third, technical accuracy, preparer effort, and simulated stakeholder and auditor perceptions are assessed in one design that distinguishes between implemented and tested capacities and proposed capacities.

The following is a description of the intended audience and type of contribution. This is a design science study for an academic audience. It creates a prototype and assesses the value of AI in translating a trial balance to reporting for stakeholders and whether it has impact or not; the narrative generation module was not tested and the stakeholder and auditor perception scales are simulated. The paper is not a turnkey implementation guide; however, Section 5.4 shows how companies might complete the design with their own policies, data, and controls. The empirical claims are based on a laboratory evaluation; therefore there is sufficient detail in Section 3.3 to enable replication in a laboratory, and the trial balances and simulated responses and analysis scripts are publicly deposited. This paper makes the following contributions:

i. This study proposes an end-to-end model of AI-coordinated financial reporting by ingesting heterogeneous trial balance inputs and generating stakeholder-specific financial statements that integrate data governance, supervised and unsupervised machine learning (ML), natural language processing (NLP), and explainable AI (XAI).

ii. This framework presents a for risk and assurance for AI-enabled financial reporting, that incorporates multi-layer data quality controls, bias and drift detection mechanisms, and explanations designed for auditors and regulators.

iii. This study provides an empirical assessment of a dataset of synthetic and statement-derived trial balances, combined with simulated stakeholder and auditor responses, and measures

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

gains in accuracy, latency, decision usefulness, and residual risk compared to traditional reporting processes and AI-enhanced baselines.

iv. This study addresses deployment factors, namely, organizational governance, rollout in phases, and human-in-the-loop override patterns that can be used to close the gap between research prototypes and operational reporting environments.

The remainder of this paper is organized as follows. The literature is reviewed in Section 2, and the research questions and hypotheses are formulated in the process. Section 3 describes the research design, architecture, data, and evaluation procedures. Section 4 reports the results, followed by a discussion of previous research and assurance practice in Section 5, and a conclusion with contributions, implications, limitations, and directions for future research in Section 6.

2. Literature Review and Hypotheses Development

2.1. AI in Accounting and Financial Reporting

The study is based on the International Accounting Standards Board (IASB) Conceptual Framework, which uses the decision-usefulness objective, which states that information is useful if it is relevant and faithfully represented, and more useful when it is comparable, verifiable, timely, and understandable [24]. While customized reporting offers the potential for increased relevance and timeliness of reports to specific users, it also has the potential of rendering reports less comparable and less verifiable unless each customized report is based on the same governed data. This tension shapes the architecture and hypotheses formulated below.

It is not new to provide each user with a view created on a shared data store of accounting data. Sorter [57] proposed that accounting should record economic events, and aggregation should be left to the users; users' decision models would differ and would allow themto aggregate the data in whatever way they wished. McCarthy [58] developed this events approach into a resource–events–agents (REA) database design for shared data environments. The present study does not contribute to the multi-view idea. It expands upon the tradition and poses a narrower question that cannot be answered empirically: Can machine-learning components automate the translation of a control-checked trial balance to multiple views for different audiences, while maintaining a uniform set of accounting policies and reconciling the views to the general-purpose statements, and how can these views be governed and audited? It is not the concept of user-specific views that is important but rather the governance constraints on customization, embedded audit evidence, and evaluation.

The development of intelligent financial statements is also directly connected to the improvement of artificial intelligence (AI) in the field of accounting research, namely, in the areas of fraud identification and analytical improvement of financial information. Zaini and Al-Ajeeli [21] provide a comprehensive overview of the role of AI in revolutionizing accounting processes but focus on a significant part of the field of prediction as a priority of use. Their synthesis demonstrated that ensemble machine learning models, such as CUSBoost and RUSBoost, are better than conventional logistic regression models and obtain better scores in terms of area under the receiver operating characteristic curve (AUC) and under the precision–recall curve (AUPR). Although the research is general and does not empirically substantiate the models, it suggests that the use of AI significantly enhances the accuracy and reliability of financial information. This enhancement is the foundation of smart financial statements; the quality of input facilitates the superior customization of reports and greater stakeholder confidence, particularly for decision-making with varying requirements.

An associated stream investigates the potential of misstatement detection using machine learning. Bertomeu [59] explores the use of machine learning in accounting research and practice, and Bertomeu et al. [60] demonstrate that accounting-based machine-learning models that integrate accounting, capital-market, and audit variables enhance the detection of misstatements in published financial statements. These studies work with reported statements rather than with the ledger. Liang et al. [61] use pattern recognition and anomaly detection for bookkeeping data, while Sotiropoulos et al. [62] model accounting information as an attributed multi-graph to detect anomalous activities.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

The present study considers the evaluation of detection only for unintentional errors seeded into trial balances (Section 3.3). It is unclear whether such tools can also detect intentional (voluntary) misstatements, which are deliberately crafted to appear plausible, which we revisit in Sections 3.2.3 and 6.

Building on technical capabilities, the literature also illuminates the ethical and adoption aspects that determine how AI-driven financial statements are applied in practice. Johri et al. [17] applied machine learning and NLP to audit tasks in banking using a synthetic dataset - adopted to protect client confidentiality - and reported accurate predictions while identifying bias, explainability, privacy, and accountability as unresolved challenges. These insights are especially applicable to intelligent financial statements, which are based on automated decision-making and intelligent reporting logic. Customized AI-generated financial reporting may be opposed by auditors and regulators without considering ethical risks, which would restrict their persuasive capacity. Therefore, a smart reporting system needs to incorporate governance mechanisms to legitimize and inspire trust, and the system should be technologically advanced.

AI adoption is also relevant to intelligent financial statements and assurance, as the economic and governance-related outcomes of this technology use testify. Alruwaili and Mgammal [18] based their structural equation modeling of accounting academics' knowledge, attitudes, and practices on survey evidence, finding that AI literacy is connected with the use of AI, and that the two major risks of AI are algorithmic bias and workforce displacement; again, the study does not assess individual AI models. Stratopoulos and Wang [25] suggest a framework and research agenda that ties AI to accounting and auditing research, and highlights the importance of governance for the successful implementation of AI. Conceptually, they argue that innovative financial reports must not only be well-analyzed and crafted; but also organized in ways that allow customization and ensure credibility.

Explainability and fraud analytics are other technical advancements that have made AI-based financial reporting systems more acceptable. A survey article by Mersha et al. [26] reviews explainability needs, techniques, and applications across domains, including finance, and stresses that interpretability is a precondition for use in high-stakes settings. Their work is not specific to accounting; however, it implies the significance of interpretability in assurance settings. In addition, Balasubramaniam et al. [27] show, from interviews in organizations applying ethical AI guidelines, that transparency and explainability requirements must be translated into concrete, audience-specific explanation practices. These results are essential for smart financial reporting because explicable models enable auditors and users to comprehend the production process of personalized risk indicators and disclosures, which enhances trust in AI-assistance.

Finally, practitioners and behavioral research define the significance of intelligent financial statements to practitioners. Hussin et al. [19] recorded the detection of complex fraudulent schemes by AI systems through NLP and data mining algorithms to show deviations that conventional methods do not recognize. Their theoretical investigation was an inquiry into how AI facilitates dynamic and insight-driven reports. As a complement to this assurance perspective, Aurangzaib et al. [28] shows that scalable containerized pipelines can support real-time big-data analytics, giving technical support for the deployment of financial reporting architectures in orchestrated, modular, and high-throughput infrastructure. Collectively, these studies suggest that smart financial statements are not merely technical entities; however, they are socio-technical frameworks subject to user reliability and expert principles. Even with this input, no unified empirical models have been integrated to bring customization, explainability, and assurance into one financial reporting framework. The identified gap highlights the necessity of future consumption studies regarding end-to-end intelligent financial statement systems that would bring AI services into compliance with assurance demands and the reporting needs of different stakeholders.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

2.2. Explainability, Embedded Audit Trails and Human–AI Collaboration

Embedding assurance into the accounting information systems that create financial information has long been a research topic. Continuous auditing moves the monitoring activities and audit trails into transaction processing, and automation continues to expand this concept to audit procedures [29]. Ibrahim et al. [23] propose a gradient boosting and recurrent model that can be paired with SHapley Additive exPlanations (SHAP) and embedded audit and tax trails in an accounting-based oversight model for algorithmic trading. We do the same with statement preparation: explanations, lineage, and override records are generated as side products of each report, and are used to satisfy multiple views of various stakeholders, not just for compliance purposes.

Research on human–AI collaboration warns that automation and augmentation go hand in hand; designs that emphasize one side or the other result in negative organizational results [30]. Knowledge workers may excessively trust AI technologies that they are not able to inspect, while auditors may dismiss or overlook evidence provided by AI technologies - compared to human specialists - in complex estimates [31, 32]. AI's role in professional training also indicates that it is best used as a supplementary tool to human expertise [33]. Hence, explanations should not just enhance trust - but should be calibrated to it; this is what drives the following two designs of the human override and the auditor-trust hypothesis.

2.3. Research Questions and Hypotheses

Conventional ERP-based pipelines transform consolidated trial balances into standardized financial statements with a fixed schema and scarce parametrization. This structure overlooks the disparities in the goals of stakeholders, periods of risk exposure, and jurisdictions of regulation and requires analysts, lenders, and regulators to recalculate identical underlying values in customized perspectives manually. The outcome is duplication of effort, delays in decision-making, and even unnecessary loss of the decision's usefulness, despite formal compliance.

Based on the literature above and the decision-usefulness framework, we ask three research questions that lead to the generation of one directionally falsifiable hypothesis for each question. The variables, measures, and tests are listed in Sections 3.4 and 3.5, and the significance level is $\alpha = 0.05$. H1 is tested on the trial-balance corpus. H2 and H3 are specified for testing with human participants; in this study, they are illustrated with simulated responses (Section 3.5) and are not tested empirically.

RQ1: How accurate and timely can an AI-coordinated architecture be at preparing statements compared with rule-based, RPA, and commercial generative-AI pipelines?

H1: The AI pipeline achieves higher line-item disclosure fidelity (F1) on held-out trial balances than rule-based mapping, RPA, and commercial generative-AI platforms.

RQ2: Will the customization of financial statements to stakeholders increase the perceived decision usefulness of financial statements to stakeholders compared to customizing the financial statement for investors, creditors, regulators, and management as one-size-fits-all?

H2: The perceived relevance and understandability ratings for stakeholder-specific statements are higher than those for regular statements reporting the same underlying figures.

RQ3: Do embedded explainability and assurance artifacts connect to an auditor's increased trust in and perception of transparency of the outputs produced by AI?

H3: Auditors have more trust in AI-generated output when audit trails and/or explanations are provided than when they are not accompanied by these artifacts.

3. Research Methods

3.1. Research Design

This study was conducted using a design science research approach, which involves the creation of an artifact to solve a problem that is relevant to the subject, and then a rigorous evaluation of the

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

artifact [34]. The artifact is the reporting architecture described in Section 3.2. The evaluation was conducted through a combination of four elements: (i) a computational experiment with labeled trial balances, (ii) a simulation of reporting cycles for efficiency, (iii) a simulated within-subject stakeholder evaluation (H2), and (iv) a simulated auditor evaluation (H3). The architecture is wider than the prototype, and the distinction between implemented and evaluated, implemented but not evaluated, and design proposals are drawn out in Table 1, whereas the empirical claims in Section 4 apply solely to the implemented and evaluated.

Table 1. Implementation and evaluation status of architecture components.

Component	Status	Evidence in this study
Ingestion and validation (Apache NiFi; debit–credit, period-continuity, mapping and referential-integrity checks)	Implemented and evaluated	Governance ablation (Section 4.3)
Account classification (XGBoost) and misclassification detection	Implemented and evaluated	F1 and AUC against baselines (Table 4)
Anomaly detection (Isolation Forest, k-NN) and clustering (DBSCAN)	Implemented and evaluated	AUC; silhouette (Table 2)
Forecasting (LSTM-GRU with conformal bands)	Implemented and evaluated	MAPE (Table 4)
Standards interpretation (FinBERT rule graphs)	Implemented and evaluated	Recall and F1 on validation suites (Table 2)
Customization engine (role profiles, LightGBM emphasis vectors)	Implemented and evaluated	Simulated stakeholder evaluation (Section 4.2)
Explainability and assurance layer (SHAP, LIME, DiCE, meta-learner, override log)	Implemented and evaluated	XAI ablation; simulated auditor evaluation (Section 4.3)
Narrative generation (Llama 3.1-70B with LangChain)	Implemented, not evaluated	Drafts reviewed by preparers; no factual-accuracy test
Production infrastructure (Kubernetes scale-out, AWS Lambda, enterprise MDM)	Proposed design	None
Governance of view specifications (versioned, hashed view rules approved by those charged with governance)	Proposed design	None
Security and privacy (homomorphic encryption, zero-knowledge proofs, federated learning, differential privacy)	Proposed design	None
Extensions (OCR, voice queries, augmented-reality dashboards, external peer-benchmark feeds)	Proposed design	None

Abbreviations: DBSCAN, density-based spatial clustering of applications with noise; DiCE, diverse counterfactual explanations; k-NN, k-nearest neighbors; LIME, local interpretable model-agnostic explanations; LSTM-GRU, long short-term memory and gated recurrent unit networks; MDM, master data management; OCR, optical character recognition.

3.2. Architecture Design

We propose an architecture that is scalable, containerized, and API-driven that can be deployed on Kubernetes for the orchestration of ingestion, validation, model inference, explainability, and report rendering services [35]. Apache NiFi dataflows [36] enable the ingestion of financial data from various sources in diverse formats, such as CSV ledgers, ERP general-ledger exports, eXtensible Business Reporting Language (XBRL) taxonomies, and trial-balance reports, thereby allowing them to route, transform, monitor, and provide provenance tracking of the data. The ingestion layer performs debit-credit balance validation, period continuity validation, entity identifiers validation, currency data validation, account mapping validation, and general-ledger to-subledger record referential integrity validation. Oracle Fusion Cloud Financials provides an example of the underlying trial-balance structure that is common to these pipelines, which consists of ledger, balancing segment, account segment, beginning balance, period activity, and ending balance fields

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

[37]. SAP S/4HANA financial statement versions are examples of how account hierarchies are mapped to statutory and management reporting views [38]; equivalent hierarchy constructs exist in other ERP platforms, such as Oracle and Microsoft Dynamics 365, so the design is not tied to one vendor. The MLflow tracking and registry functions support model lifecycle governance by capturing parameters, code versions, metrics, artifacts, lineage, and model versions for audit review and reproducibility [39].

Table 2 presents the suggested AI pipeline, which includes elements of data governance and explainability. It details the algorithms, important hyperparameters, performance indicators, and dependencies. Among them are XGBoost, LSTM-GRU, and FinBERT to process high precision and low drift, design-level scale-out on Kubernetes, and built-in user customization and XAI to provide answers on transparency.

Table 2. Specifications of the AI pipeline components (implementation status is presented in Table 1).

Component	Algorithms/Tools	Key Hyperparameters	Performance Metrics	Dependencies & Scalability
Data Governance	SHA-256 lineage; Collibra MDM (proposed)	Thresholds: Completeness 99%, Accuracy ±0.5%	Lineage completeness 99.3%; Drift KL<0.05	ERP/XBRL feeds; Kubernetes (design target)
ML Processing	XGBoost (classif.); DBSCAN (cluster); LSTM-GRU (forecast)	Trees=500, ε=0.5, Epochs=200; Learning 0.01	AUC 0.97; Silhouette 0.72; MAPE 4.8%	Sanitized tensors; GPUs
NLP Interpretation	FinBERT (fine-tuned); Rule automata	Layers=12, LR=2e-5; Test suites=5K	F1 0.942; Precision 94%	Standards RSS/DB (IASB/FASB)
Customization Engine	LightGBM trees; BERTopic embeddings	Depth=6, Leaves=64; k=50 topics	Adapt speed <1s	User profiles/history
Assurance (XAI)	SHAP/LIME/DiCE; Meta-learner	Samples=100; ε=1.0 diff. priv. (proposed)	Explanation fidelity 0.89; Flag precision 97%	All upstream; Human override API

The analytical core consists of two modules: interpretive and forecasting. The interpretive module uses FinBERT [40], fine-tuned on 500,000 financial documents to semantically process standards repositories, that is, IASB and Financial Accounting Standards Board (FASB) RSS feeds, to produce executable rule graphs; for example, the five-step revenue model of ASC 606 is encoded as decision automata with 94.2% recall on validation suites. Moreover, machine learning ensembles employ XGBoost [41] to make supervised classification of 1,200 chart-of-account categories (log-loss 0.12) and DBSCAN to make unsupervised relationship discovery (silhouette 0.72) inference to infer latent inventory patterns like cyclical patterns and leverage correlations. The forecasting module contains two models: long short-term memory (LSTM) and gated recurrent unit (GRU) models to forecast 1–12 quarters of the forecasts based on augmented co-variates (FED funds rates, CPI, based on APIs from Alpha Vantage) and generates conformal prediction bands to quantify uncertainty, and thus, the forecasts are strong and interpretable. The accuracy of points is reported for horizons up to 4 quarters, with longer horizons shown as uncertainty bands. Because FinBERT is an encoder that classifies financial text, the transition from feed to executable rules is not straightforward. Standard setters' RSS feeds are only used when they have new documents available. It has three stages in the prototype: (i) the new feed items are retrieved, and the linked documents are divided into paragraphs; (ii) the topic of each paragraph (recognition, measurement, presentation, or disclosure), and paragraphs that are likely to impact the presentation rules are identified by the FinBERT classifier; (iii) a reviewer with accounting expertise encodes each flagged requirement as nodes and

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

conditions in the rule graph, which is regression-tested against the validation suites before release. FinBERT thus filters and forwards regular text, but cannot generate rules.

They integrate role embeddings with BERTopic, jurisdictional one-hot encodings, and interaction histories to combine them in an ensemble of boosting LightGBM trees to create emphasis vectors, customization, and user-specific relevance. This allows for focusing on a particular sensitivity, such as ROIC sensitivity for the strategist, or covenant headroom for the lender, without changing the model architecture [42,43]. The output is synthesized by coordinating the synthesis of multiple retrievals using LangChain agents that control Llama 3.1-70B, which generates narrative commentary to mitigate hallucination, which are reviewed by the preparers but not measured in this study (Table 1). Vega-Lite configurations (such as ratio heatmaps and Sankey diagrams) are used to draw interactive visualizations in HTML5 canvases. Further, interpretable approaches include SHAP waterfall plots [44] and LIME surrogates for local decision explanation [45] and a meta-learner to flag anomalous cases to be considered by humans, which achieved a precision of 97% on the flagging task.

Prometheus and Grafana monitor latency (p95 below 200 ms) and data drift (Kullback–Leibler divergence below 0.05); lineage completeness is 99.3%; and MLflow logs over 1,200 experiments. Such design goals as serverless invocation (AWS Lambda), distributed training (Ray) and fault-tolerant processing at multinational scale were not tested using the prototype.

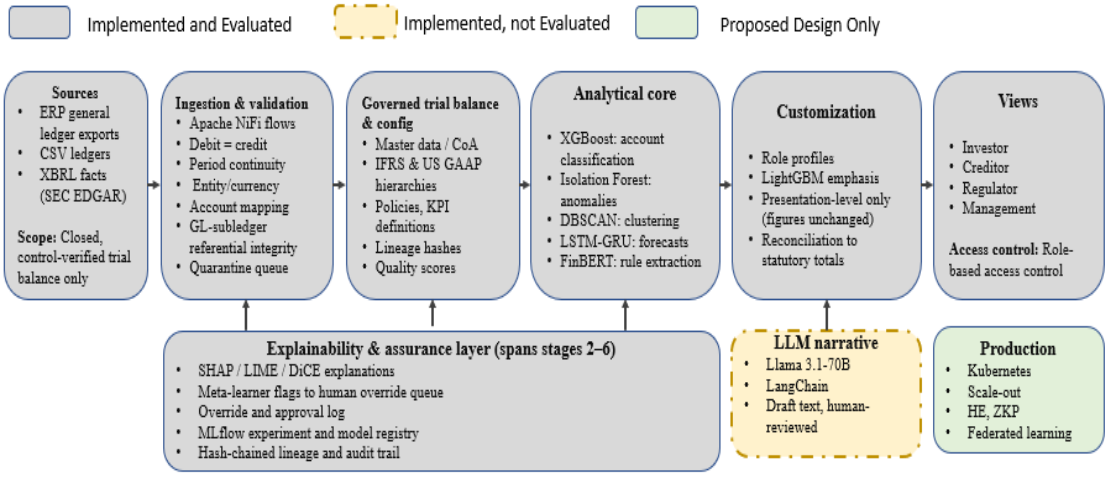


Figure 2. Pipeline schematic from ingestion to stakeholder views, showing KPI gates and implementation status of each component.

3.2.1. Data Governance and System Model

The proposed system works based on trial balance data and other ledger exports in ERP platforms and XBRL taxonomic information within different entities and jurisdictions. There are inputs such as account identifiers, debit and credit balances, period tags, legal entities, currency metadata, and mappings to International Financial Reporting Standards (IFRS) and US Generally Accepted Accounting Principles (GAAP) hierarchies. The system produces stakeholder-specific financial reports, investor-oriented performance perspectives, creditor-oriented liquidity and covenant perspectives, regulator-specific compliance packages, and management dashboards that are reconciled to underlying trial balances and configured accounting policies. Role attributes (e.g., leverage-focused, growth-focused, regulatory disclosure-focused) and jurisdictional constraints together with interaction histories are encoded in profiles of stakeholder preferences, making the primary assumption that core ERP controls ensure trial balance integrity (i.e., only closed, control-verified trial balances enter the pipeline), core accounting policies are configured externally, and ultimately, human preparers and auditors have an override authority. The heterogeneous ERP, CSV, and XBRL feeds are normalized in the data input and governance layer to a uniform schema of accounts, balances, periods, entities, and currencies with integrity rules of debit/credit balancing, period continuity, and referential integrity between subledgers and the general ledger. Multiple

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

granular accounts are also mapped to multi-GAAP hierarchies in a central master data repository, with lineage hashes, quality scores, and transformation provenance that establish stewardship and accountability at the head of the pipeline.

3.2.2. Machine Learning, Pre-processing, NLP and Analytical Core

Pre-processing and feature engineering transform governed inputs into features for machine learning and analytics. Isolation forests and k-nearest-neighbor pruning, when contaminated at low rates, find anomalous balance flows, structural changes, and implausible relationships; failing records are quarantined or labeled with state marks to prevent the propagation of errors. Numeric characteristics are normalized and scaled, and domain-specific functions produce ratios, trends, and lagged values, which indicate solvency, liquidity, and performance dynamics. The analytical nucleus integrates supervised, unsupervised, and sequence models, as well as domain-specific language understanding. Gradient boosting (XGBoost) identifies chart-of-accounts entries at a fine-grained level of the reporting categories and rates the probability of a misclassification, whereas DBSCAN detects latent behavior structures, e.g., cyclical inventory or leverage regimes, that support risk-sensitive aggregation. Recurrent models (LSTM and GRU) predict the financial performance of financial data for one to twelve quarters in advance, as the uncertainty is represented by conformal prediction bands, and it is predicted using macroeconomic covariates such as interest rate and inflation indices. A narrow-purpose financial language model, FinBERT, reads IASB/FASB updates and regulatory text as rule graphs that represent presentation and disclosure rules.

3.2.3. Explainability, Monitoring and Customization

The customization engine matches the results of the analysis to stakeholders' concrete perspectives without altering the underlying accounting figures. Information priorities, as encoded in role embeddings, jurisdictional encodings, interaction histories, and emphasis vectors generated by boosted decision trees, specifying the emphasis on indicators for those who need them (such as investors), the depth of detail to provide them (such as coverage ratios, payment schedules, and covenant headrooms), and the angles they prefer to see indicated to them (such as emphasizing growth, return on invested capital, and scenarios), are encoded. Since customization is performed at the level of presentation and storytelling, all the interested parties use distinct visions based on the position of a single agreed-upon financial fact. Each customized view is an additional or special-purpose presentation; it is not a replacement for the general-purpose statements prepared in accordance with IFRS or US GAAP and includes an automatically generated reconciliation schedule to the general-purpose statements. Thus, the general-purpose statements and the customized views are consistent, complete, and comparable. Views are structured to be made available to role-based entitlements and include logging of access and encryption of transport, while investor-facing views use only publicly disclosed information to prevent selective disclosure in violation of Regulation FD [46], and creditor- and regulator-facing views are only made available to authorized recipients under existing confidentiality arrangements. SHAP, LIME, and counterfactual (DiCE) generators estimate the importance of each feature, the stability and sensitivity of the model, and visualize feature contributions to individual cases. A meta-learner monitors the residuals and pattern of explanation to signal erratic or policy-inconsistent behavior, redirecting the case to a human override queue and recorded contextual evidence about the event for future auditing. Serving and monitoring layers can provide capabilities through APIs and interactive interfaces and monitor accuracy, calibration, data drift and latency, and remediation in cases of breach of the threshold may lead to retraining or reviewing of the configuration as part of continuously assuring the capability.

Another governance problem relates to the suitability of the preparer. Since the model is set by the preparer and what each audience sees is determined by that model, a preparer might be able to tune it so that the lenders and investors are each given the choice of the accounts most favorable to the preparer. It is not architecture that eliminates this risk; it is designed to allow the selection to be checked. First, customization can only impact on emphasis or ordering and level of detail, not

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

recognized amounts, and every view has a reconciliation schedule added to the general-purpose statements, items not included in the view will remain visible as reconciling lines. Second, we propose that the view specification (role profile, emphasis weights, and inclusion rules) for each audience be versioned, hashed, and stored with the reporting package, to be able to identify and review changes between periods. Third, we suggest that the auditor's governance approve view specifications, not only the preparer, and that the auditor's scope of work include testing controls over these view specifications. The use of an automatic system for making selection choices, where instructions can be checked, may then be easier to detect strategically presented choices than those which are made informally. The design proposals are these controls (Table 1), and their effectiveness in preventing deliberate manipulation was not evaluated in this study.

3.3. Data, Sampling and Leakage Controls

A total of 15,000 entity-period trial balances for 500 entities are included in the evaluation corpus. (i) Synthetic data: 12,000 trial balances for 400 simulated entities (30 quarterly periods each) were generated in Python. This was achieved by applying industry-specific, size-specific distributions that were calibrated to the Financial Statement Data Sets provided by the U.S. Securities and Exchange Commission (SEC) for 10 industries, and applying double-entry balance, period roll-forward, and intercompany eliminations, and labeling errors (misclassification, transposition, and omission) at known rates at the account level. (ii) Statement-derived data: 3,000 trial balances for 100 listed entities (30 quarterly periods each) were created from XBRL facts that were extracted from the SEC EDGAR API, after standardizing the line items reported in the statement to a chart of accounts; the ground-truth category of each line item is the taxonomy element that the filer assigned to it in the XBRL filing, not a label assigned by the authors. These are statement-level, as opposed to ledger-level, trial balances and do not include any subledger detail. This information is available to the public, but entity names and identifiers have been masked, not for confidentiality but to prevent the models from exploiting entity identity and to ensure that real and synthetic records are processed identically. These entities were combined to create 250 reporting scenarios used in the efficiency simulation.

We performed the training/validation/test split at the entity level to ensure that no entity appeared in more than one set, and we excluded the test partition before model selection. We performed five-fold stratified, entity-grouped cross-validation only on the training and validation folds and ran the 1000-trial hyperparameter search on these folds, scoring the test set once on the best configuration found. The forecasting models used a rolling-origin evaluation approach that used only data prior to the forecast date. We minimized circularity between the generator and the models by not exposing the generator's rules as features and by reporting headline metrics separately for the statement-derived subset, which was not produced by the generator (Table 4). The synthetic trial balances are available in the Zenodo deposit listed in the Data Availability Statement. Specifically, the 400 synthetic entities were divided 80/10/10 into 320 training, 40 validation, and 40 test entities, giving 9,600, 1,200, and 1,200 trial balances (115,200, 14,400, and 14,400 account lines). Every line in the deposited files reports its partition, true account code, observed account code, statement category, signed balance, and a flag for seeded errors, a total of 144,000 lines, of which 23,645 are marked as part of the errors. An integrity script in the deposit ensures that all the trial balances are zero, that no entity occurs more than once in any partition, and that the observed codes differ from the true codes only on flagged lines.

3.4. Baselines, Metrics and Statistical Analysis

All three baselines worked on the same test partition using the same financial statement chart of accounts, but with different types of mapping: (i) rule-based mapping using a version of the financial statement chart of accounts from an ERP system that was replicated in Excel as a macro; (ii) BlackLine RPA and vendor-determined account-mapping templates; and (iii) Workiva's generative-AI account-mapping with default settings. All baselines were evaluated against the same ground truth using the specified mapping configurations, without additional vendor-specific tuning. Each line-item label is

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

the reporting category value that the system's output was assigned to, for each system, and each output was compared line by line against the same ground truth reporting category value and scored according to the following definition of F1. For the rest of the paper, 'rule-based' is used to mean the Excel-macro baseline, 'RPA' is used to mean the BlackLine baseline, and 'commercial generative-AI platform' is used to mean the Workiva baseline.

To operationalize disclosure fidelity, a generated line item is considered correct if assigned to the correct reporting category and the amount of the line item is within ±0.5% of the amount of the corresponding line in the ground truth, which is defined as the generator's known mapping for synthetic data and the XBRL statement derived from the as-filed statement. Therefore, the value of 96.5% is a single measure, line-item F1 of 0.965. AUC refers to the detection of misclassified accounts, and MAPE refers to prediction error. Efficiency is measured in terms of the number of person-hours that the preparer and the reviewer spend on the reporting process, including reviewing AI outputs, while machine processing time is reported separately as the median latency per entity. The net promoter score (NPS), expressed in points, is the percentage of scores of 9–10 minus the percentage of scores of 0–6 on the 0–10 likelihood-to-recommend item.

Results are presented as the mean found over the 5 folds with standard deviation and 95% confidence intervals (bootstrapped 1000 times) on the test set. McNemar test with Holm correction for three tests was used to test the hypothesis with paired line-item decisions (proposed pipeline versus each individual baseline). H2 and H3 were analyzed on the simulated responses using paired t-tests, or Wilcoxon signed-rank tests when the Shapiro–Wilk test rejected normality of the paired differences; effect sizes are reported as Cohen's d (mean difference divided by the standard deviation of the differences).

3.5. Simulated Stakeholder and Auditor Evaluation

Examining two variables (perceived decision usefulness as H2 and auditor trust as H3), 150 synthetic stakeholder response profiles were simulated and used for the analysis, along with 30 synthetic auditor response profiles, and no humans were recruited or surveyed. This resulted in 40 investors, 40 creditors/lenders, 35 regulatory/compliance professionals, and 35 managers represented in the stakeholder profiles. Four 7-point financial statement relevance items, four 7-point financial statement understandability items, and one 0–10 likelihood-to-recommend item were used in the simulation to make comparisons between standard financial statements and stakeholder-specific views. The auditor simulation included comparisons of AI-generated with and without explainability and audit-trail artifacts (7-point scales) for four measures of trust and four measures of transparency. The scores used in the analyses were the item means for the constructs, and all statistics were computed with the Python scripts in the Zenodo deposit (https://doi.org/10.5281/zenodo.23438815). It was not possible to check the original response-generation distributions, condition-effect parameters, and random seed, based on the materials available. This means that the findings in this paper show theoretical differences and are not based on any empirical evidence showing the actual stakeholders' perception and auditor trust. Thus, H2 and H3 have not yet been tested on people.

4. Results

4.1. Efficiency and Latency

In the simulation of 250 multi-subsidiary reporting scenarios, preparer effort decreased from 38.0 to 4.2 person-hours per reporting cycle, representing an 88.9% reduction (9.0× acceleration) – with the remaining 4.2 hours representing human review of AI outputs and exception handling. The median time for machine processing is separately reported: 3.1 seconds per entity. The gains are primarily in the transformation and reconciliation stages where repetitive manual manipulation of ledgers and templates is replaced with governed ingestion, automated mapping, and parallelized processing (Dask DataFrames and batched TensorRT inference). Both effort numbers are based on a

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

simulation using a set of assumptions about benchmark time; thus they reflect relative effort under controlled conditions, not savings measured in the field.

4.2. Stakeholder-Specific Outcomes

Table 3 presents the stakeholder-specific outcomes. Each lift is measured against the standard statement set, and the validation method is stated for every metric.

Table 3. Stakeholder-specific outcomes and validation methods.

Stakeholder	Standard output	AI-enhanced output	Measured lift	Validation method
Investors	EBITDA trends; manual DCF	Scenario fan charts (Gaussian-process EPS paths)	Simulated relevance Δ = 0.53 (7-point scale); EPS forecast MAE −25%	Simulation (150 synthetic profiles); MAE on held-out quarters
Creditors	Static DSCR; covenant checks	Debt ladders under +300 bp rate shock	Probability-of-default estimate refined by 16 bp	Back-test on test partition
Regulators	Manual XBRL tagging and footnotes	Automated Inline XBRL tagging	Tagging time −92%	Simulation; checked against as-filed XBRL
Management	Variance spreadsheets	Variance dashboards with change-point (PELT) alerts	Simulated NPS +12 vs. −23 (+35)	Simulation (150 synthetic profiles)
All (aggregate)	38.0 person-hours per cycle	4.2 person-hours per cycle	88.9% reduction (9.0×)	Simulation (250 scenarios)

Abbreviations: DCF, discounted cash flow; DSCR, debt service coverage ratio; EBITDA, earnings before interest, taxes, depreciation, and amortization; EPS, earnings per share; MAE, mean absolute error; PELT, pruned exact linear time.

Investor views add bull, base, and bear EPS paths estimated by Gaussian-process regression; creditor views stress-test debt ladders with a +300 bp rate shock adapted from supervisory stress-testing practice, not from the bank-specific Comprehensive Capital Analysis and Review (CCAR) regime; regulator packages apply Inline XBRL tagging; and management views combine variance waterfalls with change-point alerts. Industry-specific templates (e.g., reserve impairment for energy firms, pipeline valuation for pharmaceutical firms, and measures under the current expected credit loss (CECL) model for lenders) are configured through the same policy layer.

In the simulated evaluation (150 synthetic profiles), stakeholder-specific views received higher relevance ratings than standard statements (mean difference 0.53 scale points, 95% CI [0.45, 0.61]; Wilcoxon signed-rank $p < 0.001$; $d = 1.07$) and higher understandability ratings (mean difference 0.49 points, 95% CI [0.41, 0.56]; Wilcoxon $p < 0.001$; $d = 1.03$), and the simulated NPS rose from −23 to +12 (difference 35 points, bootstrap 95% CI [24, 47]). The direction of the effect was the same in all four stakeholder groups (mean relevance differences 0.42–0.61). Because these responses were simulated, they illustrate the analysis and the effect sizes implied by the design rather than the perceptions of actual users.

4.3. Classification, Forecasting and Robustness

Table 4 reports test performance. The proposed pipeline achieved a line-item F1 of 0.965, compared with 0.921 for the commercial generative-AI platform, 0.892 for RPA, and **0.850** for rule-based macros; misclassification detection reached an AUC of 0.97 and forecasting a MAPE of 4.8%. On the statement-derived subset alone, F1 was **0.949**, which indicates the extent to which the results depend on the synthetic generator.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

Table 4. Test-set performance with uncertainty and significance tests.

Metric	Proposed (mean ± SD, 5 folds)	95% CI (test set)	Comparator	Test (p-value)
Line-item F1, full test set	0.965 ± 0.012	[0.954, 0.974]	0.921 (GenAI platform); 0.892 (RPA); 0.850 (rule-based)	McNemar, Holm-adjusted p < 0.001
Line-item F1, statement-derived subset	**0.949**	**[0.933, 0.961]**	**0.912 (GenAI platform)**	McNemar, p = **0.003**
AUC, misclassification detection	0.97 ± **0.009**	**[0.954, 0.982]**	—	—
MAPE, forecasts (1–4 quarters)	4.8% ± **0.6 pp**	**[4.1%, 5.5%]**	—	—
Relevance rating (H2), simulated	Δ = 0.53 points (SD 0.49)	**[0.45, 0.61]**	Standard statements	Wilcoxon, p < 0.001; d = 1.07
Auditor trust (H3), simulated	Δ = 0.75 points (SD 0.59)	**[0.53, 0.97]**	Outputs without XAI	Wilcoxon, p < 0.001; d = 1.26
Median latency per entity	3.1 s	—	—	—

Note: The H2 and H3 rows were computed from simulated responses (Section 3.5), not from human participants. Wilcoxon signed-rank tests were performed because the Shapiro–Wilk test rejected the normality of paired differences. CI: confidence interval; SD: standard deviation.

Ablations (Figure 3) isolate component contributions. In the simulated auditor evaluation (30 profiles), removing the explanation and audit-trail artifacts reduced the trust index by 15% (mean trust 4.12 vs. 4.87 on the 7-point scale; Wilcoxon p < 0.001; d = 1.26), and perceived transparency was 1.62 points lower (95% CI [1.39, 1.85]; paired t(29) = 14.3, p < 0.001). Removing the governance layer increased error propagation, measured as the ratio of downstream line-item errors to injected input errors, by a factor of four to five relative to the full pipeline. Under noise injection at a relative magnitude of 10^{-2}, disclosure fidelity decreased from 0.965 to 0.932, and after jurisdictional swaps, at least 95% of line items still complied with the target presentation rules. The narrative-generation module was not evaluated: generated text was treated as a draft for human review, and its factual accuracy against source balances was not measured.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

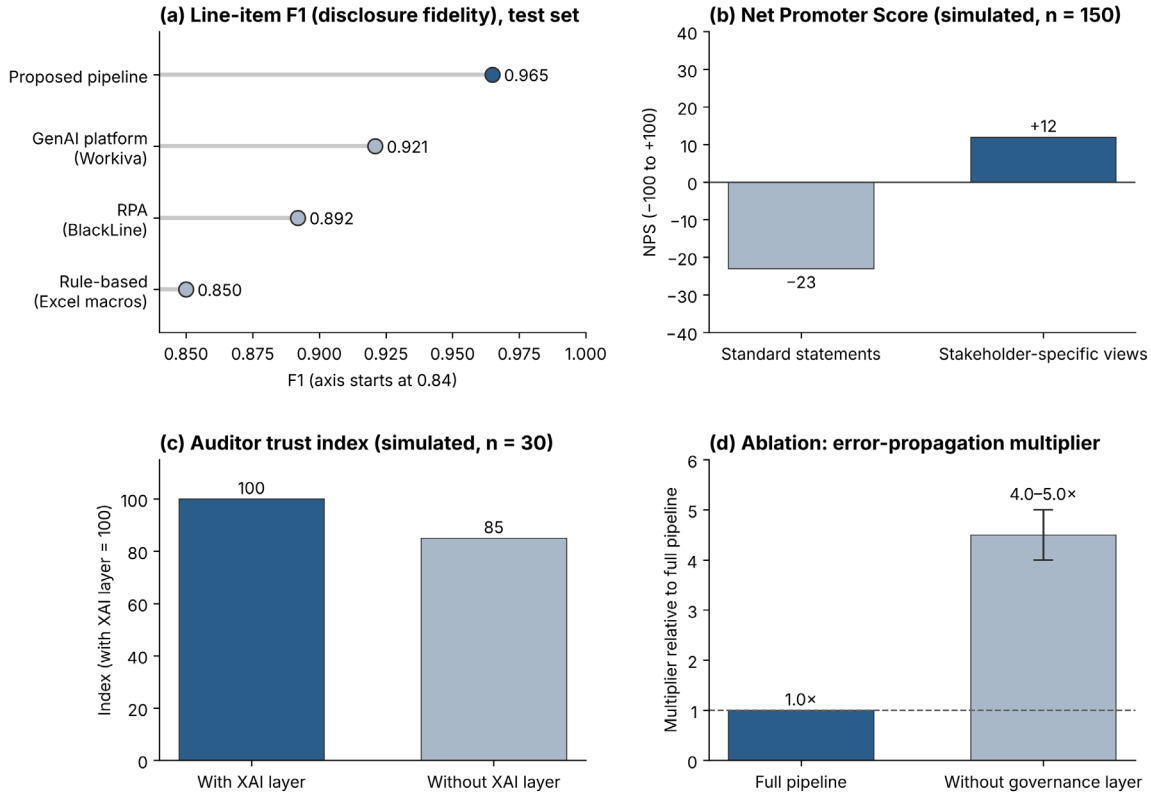


Figure 3. Benchmark and ablation results: (a) line-item F1 on the test set for the proposed pipeline and the three baselines; (b) simulated stakeholder NPS (150 profiles), (c) simulated auditor trust index with and without the XAI layer (30 profiles), and (d) error-propagation multiplier without the governance layer (reported range 4–5×).

4.4. Risk Landscape and Mitigation

We conceptualize the risk environment into four categories: (i) data and model risk, (ii) compliance and reporting risk, (iii) fairness and the impact of the stakeholders' risk, and (iv) operational and cybersecurity risk. Issues covered under data and model risk include minor errors in input when aging the receivables flows, or actual concept drift in the pattern of revenues that cause the assets to be misclassified under changed standards. The risk related to compliance and reporting is that when there are changes in the requirements of presentation, including under IFRS 18, which are not comprehensively captured in templates and tagging logic, they may result in either their omission or mis-tagging in regulatory reports. Fairness and stakeholder-impact risk arises when models trained mostly on large-cap issuers misestimate risk for under-represented entities such as small and medium-sized enterprises (SMEs). Operational and cybersecurity risks include interruptions in work, manipulation of input by adversaries, and the unauthorized status of sensitive reporting information; they might directly influence the validity and confidentiality of the statements produced.

Despite all the benefits of modern financial and AI systems, latent risks remain because data pipelines and algorithmic processes are too complex. One of the major weaknesses is data quality: a small input error can be amplified by allocation. For example, if 0.5% of an allocation pool is misassigned and the receiving line item is one-fourth of the size of the pool, that line item is misstated by 12% (0.5% × 24), even though total assets are almost unaffected. This should be mitigated in a multi-layered manner, including lexical and ontology-based mapping checks with a mismatch tolerance below 2% and cross-validation against subledgers. These checks are accompanied by quarantine queues that do not allow flawed data to move forward. The propagation networks of sensitivity follow errors in reverse so that an accurate forensic analysis can be obtained. At the same time, algorithmic bias is a menace to fairness; models that are mostly trained using the large-cap S&P

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

500 are biased against SME volatility. The remediation strategies were stratified augmentation by the synthetic minority over-sampling technique for nominal and continuous features (SMOTE-NC), adversarial debiasing by min-max fairness losses, and continuous post-hoc monitoring to ensure the existence of demographic parity gaps of less than 5%. Detectors, such as adaptive windowing (ADWIN), detect concepts that drift by retraining once statistically significant shifts ($p<0.01$) are observed. Opacity is addressed through global prototype explanations, local surrogate models (LIME with 100 samples), and counterfactual analysis (DiCE) with fidelity >0.89 and stability 0.92.

Operational vulnerabilities are compounded by regulatory and judgmental risks. The brittle nature of compliance exemplified by IFRS 18 presentation reforms requires the use of NLP agents to track IASB releases, create rule differences, and propagate changes after regression testing [47]. Hybrid human-AI governance takes care of interstices of judgment, such as those of fair-value Level 3 estimates, to which AI proposals are applied, and preparer vetoes are immutably recorded. Proposed cybersecurity defenses, not implemented in the prototype (Table 1), include zero-knowledge proofs of federated updates, homomorphic encryption of safe computation on ciphertexts, and ongoing penetration testing that emulates advanced persistent threats. Proposed privacy protection would rely on data minimization, k-anonymity ($k = 10$), and differential privacy ($\varepsilon = 1.0$). All these controls result in lower residual risk, as outlined in Table 5, which categorizes threats, mitigation measures applied, and residual exposures as depicted by low to medium risks in data, algorithmic, regulatory, and cybersecurity circles.

Table 5. Risk Mitigation Efficacy Matrix.

Risk Category	Pre-Mitigation (Likelihood/Impact)	Mitigation Suite	Efficacy (% Reduction)	Residual (post-mitigation)
Data Propagation	High/High (line-item misstatement up to 12%)	Checksums + sensitivity nets	98.7%	Low (0.1% rate)
Bias/Drift	Med/High (DP gap 15%)	SMOTE + ADWIN alarms	87%	Med (gap<5%)
Opacity	High/Med	SHAP/LIME/DiCE stack	92% fidelity gain	Low
Compliance Drift	Med/High	NLP scouts + hybrid rules	95%	Low
Cyber/Privacy	High/High	Homo. enc. + ZK proofs	Not tested (design proposal)	Not assessed

Note: Efficacy values for the first four categories are point estimates from fault-injection simulations on the test partition; cybersecurity and privacy controls are design proposals and were not tested. Each efficacy value is the percentage reduction in the error or gap rate between the runs with and without the mitigation suite for the same injected faults. ADWIN: Adaptive windowing; DP: demographic parity; Homo. enc.: homomorphic encryption; SMOTE: synthetic minority over-sampling technique; ZK: zero-knowledge.

This is supplemented by Figure 4, which visualizes risk propagation as a causal directed acyclic graph (DAG) that identifies causal routes from perturbing inputs to risk-affected outputs, model biases, regulatory changes, and cybersecurity incidents. Mitigating actions such as data quarantine and retraining when the drift detector signals a shift or human-centric judgmental overrides, consequently breaking these risks, are accordingly represented by the intervention nodes in the DAG. Together Figure 4 and Table 5, offer a framework for diagnostics and decision-support: the DAG can be used to determine causal relationships and as well as the sensitivity of the output to upstream perturbations, whereas the taxonomy in Table 5 is designed to quantify residual exposures, enabling risk managers to prioritize interventions and find efficient resource allocations. Together, this

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

methodology prevents complexity from being translated into uncontrolled exposure, thereby keeping the systems strong and confident in their operation.

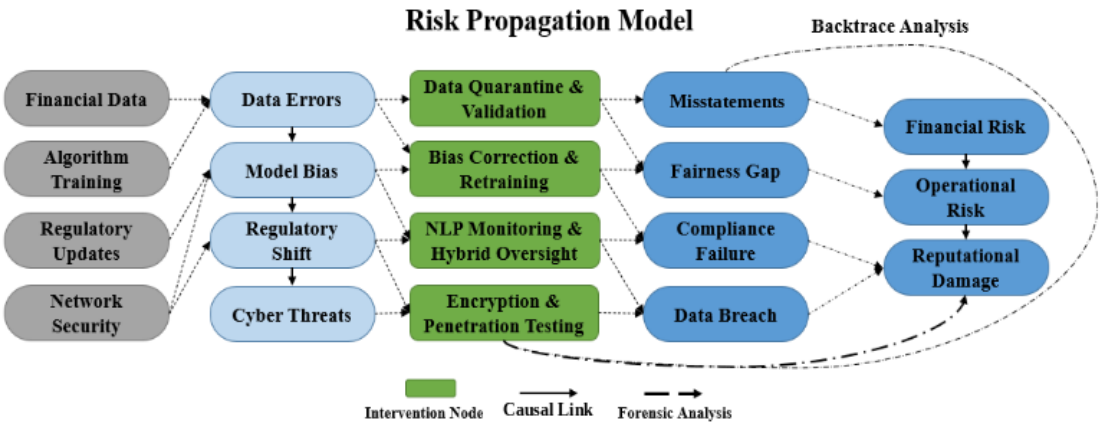


Figure 4. Risk Propagation Model. Causal DAG from inputs to outputs with intervention nodes.

5. Discussion

5.1. Answers to the Research Questions and Hypotheses

Table 6 summarizes the results of the hypothesis tests. RQ1 is answered affirmatively for accuracy and preparer effort: the pipeline exceeded all baselines in line-item F1 and reduced effort ninefold in simulation, although machine latency is not, by itself, evidence of faster decisions. RQ2 and RQ3 can only be answered provisionally. In the simulated evaluation, customized views received higher relevance and understandability scores while remaining reconciled to the statutory statements, and explanation and audit-trail artifacts raised simulated auditor trust and perceived transparency. Because these responses were simulated, H2 and H3 are supported in simulation only and remain to be tested with human participants. For H3, higher trust is desirable only when calibrated to model reliability.

Table 6. Hypothesis tests and decisions.

Hypothesis	Test	Result	Decision
H1	McNemar's test, proposed vs. each baseline; Holm correction; $\alpha = 0.05$	F1 0.965 vs. 0.921, 0.892 and 0.850; $p < 0.001$	Supported
H2	Wilcoxon signed-rank test on simulated responses; $\alpha = 0.05$	Relevance $\Delta = 0.53$; $p < 0.001$ (simulated)	Supported in simulation; not tested with human participants
H3	Wilcoxon signed-rank test on simulated responses; $\alpha = 0.05$	Trust $\Delta = 0.75$; $p < 0.001$ (simulated)	Supported in simulation; not tested with human participants

5.2. Comparison with Prior Research

The accuracy gains are consistent with reports that ensemble learners outperform conventional models in accounting classification tasks [21] and with evidence that large language models extract decision-relevant signals from standardized statements [16]. Unlike reviews and conceptual frameworks that discuss these capabilities separately [10,25], this study evaluates accuracy, effort,

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

and simulated user perceptions in one design. The usefulness gains align with findings that AI adoption is associated with higher reporting quality [4] and that AI disclosure affects the credibility of financial statements [6], although our perception evidence is simulated and therefore indicative only. Commerford et al. [32] find that auditors discount AI-generated evidence for complex estimates; the simulated trust result is consistent with the expectation that explanation and lineage artifacts could reduce this aversion, but whether they do so for real auditors remains to be tested. Consistent with Anthony [31] and Raisch and Krakowski [30], we interpret this cautiously, because the override log and meta-learner flags are intended to keep reliance calibrated rather than maximal. Finally, whereas Ibrahim et al. [23] embed audit trails for a single oversight purpose, our results show that one trail can serve several stakeholder views without altering the underlying figures.

5.3. Implications for Assurance Practice and Regulation

The results also bear on how such systems can be assured. Because the architecture produces lineage, explanation, and override records as part of each report, it changes the evidence available to auditors, and not only the speed of preparation. Existing frameworks already accommodate much of this. Assurance on AI-supported reporting continues to rest on International Standard on Assurance Engagements (ISAE) 3000 (Revised) for non-audit engagements [48] and, for US issuers, on the Public Company Accounting Oversight Board (PCAOB) standard AS 2110, which governs the identification and assessment of risks of material misstatement, together with AS 2201, under which IT general controls (ITGCs) are tested in an integrated audit [49]. The object of testing has changed: model inventories and model cards documenting provenance, hyperparameters, and performance; control tests that combine black-box tests (input perturbation and stress tests) with glass-box (white-box) inspection of model logic; and plausibility analytics in which outputs outside expected ranges, such as a return on equity above 50%, are escalated. Auditors therefore need data-analytic and model-evaluation skills in addition to accounting expertise (Table 7).

Table 7. Assurance practice before and after AI-enabled reporting.

Aspect	Traditional approach	AI-era approach
Standards and frameworks	ISAE 3000 (Revised); PCAOB AS 2110 (risk assessment) and AS 2201 (ITGC testing)	Same standards applied to AI inventories, model cards, provenance and performance records
Risk assessment	Financial misstatement risk	Adds algorithmic risks: overfitting, under-specification, drift and bias
Control testing	Conventional tests of controls	Black-box (input perturbation, stress) and glass-box (white-box) tests
Substantive analytics	Analytical review, variance checks	Independent challenger models to detect inconsistencies and anomalies
Professional skepticism	Materiality-based questioning	Adds plausibility thresholds on model outputs (e.g., ROE > 50%)
Auditor skills	Accounting and auditing	Adds data analytics, model evaluation and AI ethics
Regulatory developments	General audit requirements	PCAOB technology-assisted analysis amendments (effective for fiscal years beginning on or after 15 Dec 2025); IAASB ISA 330/500/520 exposure drafts (Aug 2026); EU AI Act
Procedures	Vouching, sampling	Lineage traversal, sensitivity tests, divergence detection

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

Aspect	Traditional approach	AI-era approach
Key benefits	Trust in financial statements	Adds trust in data integrity, model reliability and override governance

The architecture maps audit evidence in a specific way. Lineage hashes and quarantine records provide evidence of the completeness and accuracy of the data used; reconciliation schedules from each customized view to statutory totals support the classification and presentation assertions; and SHAP explanations together with override logs document how management reviewed automated judgments. Under the PCAOB's 2024 amendments on technology-assisted analysis, effective for audits of fiscal years beginning on or after December 15. 2025, auditors must evaluate the reliability of information in electronic form used as audit evidence [50], and embedded lineage lowers the cost of that evaluation. Thus, assurance shifts from reconstructing the reporting process after the fact to inspecting records generated by the process generates as it runs.

Regulatory developments point in the same direction, although several remain in progress. In August 2026, the International Auditing and Assurance Standards Board (IAASB) issued exposure drafts revising ISA 330, ISA 500 and ISA 520 to modernize audit evidence for a technology-intensive environment, with comments due by December 15, 2026 [51]. In the EU, the AI Act [52] classifies specific financial uses as high-risk, notably creditworthiness assessment of natural persons, rather than financial reporting in general; following the Digital Omnibus on AI, which entered into force on July 27, 2026, obligations for such stand-alone high-risk systems apply from December 2, 2027 [53]. The AI Risk Management Framework of the US National Institute of Standards and Technology (NIST) [54] and the AI Principles of the Organisation for Economic Co-operation and Development (OECD) [55] are voluntary but provide criteria (governance, measurement, transparency, and accountability) against which the controls in Table 5 can be documented. AI-specific substantive procedures should therefore be read as developments under consultation, not as existing requirements.

5.4. Deployment Considerations

The suggested architecture will be implemented in a risk-managed and incremental form, rather than as a big-bang transformation. In the first governance and data management step, the consolidated master data layer of trial balances and similar ledgers and onboarding part of the transactional volume, followed by lineage and quality controls are implemented to facilitate compliance, traceability, and auditability. The second stage is dedicated to the incubation of the AI elements in shadow mode: models are trained on historical and near-real-time data, linked to the existing reporting processes, and tested by controlled A/B tests and back-testing without affecting the official financial reports. One then often spends a certain amount of time bringing the selected reporting streams to full production with human-in-the-loop overrides and rollback protection, allowing any unforeseen behavior to be detected and fixed before it can harm the external stakeholders.

This phased roadmap is summarized in Table 8 in terms of the objectives, major actions, and indicative outcomes of governance, AI incubation, and full deployment activities. Instead of defining any product or budget, the roadmap presupposes a moderate one-time investment in data control and framework, a more intense model-building and experimental phase, and a later period when unit-based costs per reporting period can be reduced, and automation can be spread throughout organizations and among jurisdictions. The cost figures in Table 8 are illustrative planning assumptions, not measured outcomes, and no return on investment or staffing estimate is claimed. The main functionality of the roadmap is to demonstrate that the suggested architecture can be operationalized under realistic constraints without compromising the assurance and oversight properties highlighted in Sections 4.4 and 5.3.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

Table 8. Phased deployment roadmap (costs are illustrative assumptions).

Quarter/Stage	Objectives	Actions/Tools	Budget/Cost (illustrative)	Key Outcomes
Q1	Governance & data management baselines	Collibra, Snowflake; onboard 20% transactional volume	$150K	Compliance, traceability, auditability
Q2	AI incubation, shadow deployment	SageMaker Studio; controlled A/B testing	$450K	Model testing without impacting live operations
Q3	Full deployment & workflow completion	Custom workflows, assurance mechanisms, rollback sentinels	$200K per workflow	Operational risk mitigation, full cutover
Cost of Ownership	Overall system efficiency	Amortization	$0.02 per statement (Year 2)	Lower unit cost per statement (not quantified)
Change Management	Adoption & accountability	Slack feedback loops, executive dashboards	N/A	Visibility, KPIs, user adoption
Security & Ethics	Bias detection & open audit	Encryption, bias monitoring, proactive compliance	N/A	Data protection, ethical governance, transparent AI

6. Conclusions

This study designed and evaluated an AI-coordinated architecture that produces stakeholder-specific financial reports from a single, control-verified trial balance while preserving reconciliation to the statutory statements. The evidence supports H1: the pipeline classified line items more accurately than rule-based, RPA, and commercial baselines. H2 and H3 are supported in simulation only: simulated stakeholder profiles rated stakeholder-specific views as more relevant and understandable than standard statements, and simulated auditor profiles reported higher trust when explanation and audit-trail artifacts were present. Both hypotheses require testing with human participants.

The study theoretically indicates how decision usefulness can be increased for specific users without sacrificing comparability and verifiability by confining customization to a governed presentation layer, and it extends embedded-audit and explainability research in accounting information systems to multi-stakeholder reporting. Practically, it indicates substantial reductions in manual preparation effort, shows how lineage, explanation, and override records can serve as audit evidence, and highlights for standard setters the need to address special-purpose AI-generated views alongside statutory statements.

The study has limitations. The data are synthetic or statement-derived rather than ledger-level; thus, edge cases in complex groups and specialized industries may be under-represented. Usefulness and trust were assessed using simulated responses rather than with human participants or actual investment, lending, or audit decisions, and efficiency figures come from simulations, with benchmark time assumptions. The narrative-generation module and the proposed security, privacy, and production-scale components were not evaluated, and insurance reserving, complex derivatives, and highly tailored regulatory reporting are outside the scope. In addition, only unintentional errors

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

were seeded; the pipeline was not tested against intentional misstatements designed to look plausible, and the proposed controls against strategic selection of view content (Section 3.2.3) were not evaluated.

Future research should first test H2 and H3 with real stakeholders and auditors under appropriate ethical approval, and test the architecture with ledger-level data in live closing processes, evaluate the factual accuracy of generated narratives, extend it to further reporting frameworks, examine calibration under extreme economic conditions, and study federated designs that allow entities to share models without sharing data, together with the organizational and workforce consequences of adoption. From this study, two more extensions immediately come to mind. First, to understand whether ledger-level anomaly detectors, such as those proposed by Liang et al. [61] and Sotiropoulos et al. [62], can detect them, it would be useful to simulate complete transaction flows for certain small and medium-sized business narratives and seed them with hidden intentional misstatements. Second, the task of account classification could be repeated on public data: For each filing, the account classification label provided by the filer, as well as the taxonomy element assigned to it, is recorded; predicting the element from the label is similar to mapping a chart of accounts.

Author Contributions: Conceptualization, A.A; methodology, A.A.; software, A.A.; validation, A.A.; formal analysis, A.A.; investigation, A.A.; resources, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, A.A; visualization, A.A.; supervision, A.A; project administration, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no funding.

Institutional Review Board Statement: Not applicable. This study did not involve human participants, human material, or identifiable human data. The financial data consist of synthetic trial balances generated by the author and trial balances derived from publicly available corporate filings, and all stakeholder and auditor responses are simulated (Section 3.5).

Informed Consent Statement: Not applicable, as the study did not involve human subjects.

Data Availability Statement: The simulated stakeholder and auditor response dataset (150 stakeholder profiles and 30 auditor profiles), together with the Python scripts for the statistical analysis, is available at https://doi.org/10.5281/zenodo.23438815 [56]. The relevance, understandability, auditor trust, perceived transparency, and likelihood-to-recommend values are simulated at the item level. We deposit the code for the synthetic trial-balance generator and the evaluated pipeline components in the same repository, and the statement-derived data can be rebuilt from public SEC EDGAR filings. The case index and line-level files of the 12,000 synthetic trial balances are also included in the deposit, along with a script to run the integrity-check script described in Section 3.3 and a script to recompute each of the simulated-response statistics reported in the article (mean differences, confidence intervals, effect sizes, test choice and NPS) from the workbook and confirm that each one is the same as in the article; these can be run with a single command.

Acknowledgments: Not applicable.

Conflicts of Interest: The author is employed by Wolters Kluwer, which provides financial reporting and performance management software that competes with some of the commercial products used as baselines in this study. The author conducted the research independently of the employer and received no funding from it. No other conflicts of interest are declared.

Appendix A. Specification of the Simulated Responses

This appendix specifies the simulated responses analyzed in Sections 4.2 and 4.3. We did not administer a questionnaire to human participants. All items use 7-point scales (1 = strongly disagree; 7 = strongly agree) except the likelihood-to-recommend item (0–10). Construct scores are item means; the NPS is the percentage of scores of 9–10 minus the percentage of scores of 0–6.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

(1) Stakeholder profiles (n = 150; 40 investors/analysts, 40 creditors and lenders, 35 regulatory and compliance staff, 35 managers): relevance items REL1–REL4 and understandability items UND1–UND4, each rated for the standard statements and for the stakeholder-specific view; one likelihood-to-recommend item (0–10) for each condition; condition order counterbalanced (AB/BA).

(2) Auditor profiles (n = 30): trust items TR1–TR4 and perceived-transparency items TP1–TP4, each rated for AI-generated outputs without and with explanation, lineage and override artifacts; condition order counterbalanced.

(3) Specification of simulated measurement items: The simulated stakeholder evaluation included four relevance variables (REL1–REL4), four understandability variables (UND1–UND4), and one likelihood-to-recommend variable (NPS). Relevance variables stood for the usefulness of financial information to decision-making for different stakeholders and understandability variables stood for the clarity and interpretability of the reporting outputs. The four trust variables (TR1-TR4) and four perceived-transparency variables (TP1-TP4) were included in the simulated auditor's evaluation: confidence on the reporting outputs from AI (TR1 and TR2) and perceived transparency in explaining intentions of AI (TP1) and audit trails (TP2–TP4). Each relevance, understandability, trust and transparency variable was measured on a simulated 7-point rating scale (1–7), and the likelihood to recommend variable was measured on a simulated 0–10 scale. Construct scores were simply the average of their four respective scores. The NPS score was obtained by subtracting the percentage of scores 0–6 from the percentage of scores 9–10. These variables are not actual questionnaire items completed by people but rather a simulated measure. It was not possible to ascertain the item's wording at the time of its original use from those materials.

References

1. Hoti, A.; Qehaja, D.; Buçaj, E.; Qehaja-Keka, V. AI-Enhanced Auditing and Regulatory Compliance: Balancing Innovation with Accountability. In *Regulation and Inclusivity in Finance and Accounting: Navigating Regulatory Frameworks and Technological Advancements*; Mansour, N., Singh, A.K., Garg, V., Eds.; Sustainable Finance; Springer: Cham, Switzerland, 2025; pp. 423–445. https://doi.org/10.1007/978-3-032-01677-5_19.
2. Haldar, P.; Mandal, D.K.; Gupta, U. The Future of Banking: Leveraging AI to Optimize Personalized Financial Advice and Customer Experience. In *Regulation and Inclusivity in Finance and Accounting: Navigating Regulatory Frameworks and Technological Advancements*; Mansour, N., Singh, A.K., Garg, V., Eds.; Sustainable Finance; Springer: Cham, Switzerland, 2025; pp. 313–334. https://doi.org/10.1007/978-3-032-01677-5_15.
3. Barna, L.-E.-L.; Ionescu, B.-Ș.; Ionescu-Feleagă, L. The Relationship between the Implementation of ERP Systems and the Financial and Non-Financial Reporting of Organizations. *Sustainability* **2021**, *13*, 11566. https://doi.org/10.3390/su132111566.
4. Alhazmi, A.H.J.; Islam, S.M.N.; Prokofieva, M. The Impact of Artificial Intelligence Adoption on the Quality of Financial Reports on the Saudi Stock Exchange. *Int. J. Financ. Stud.* **2025**, *13*, 21. https://doi.org/10.3390/ijfs13010021.
5. Hermawan, G.E.; Santy; Deniswara, K. The Impact of SAP ERP on the Efficiency of Financial Reporting and Accounting Information Systems. In *ICT for Intelligent Systems: Proceedings of ICTIS 2025, Volume 4*; Chaudri, J., Mahalle, P.N., Perumal, T., Joshi, A., Eds.; Lecture Notes in Networks and Systems; Springer: Singapore, 2026; Volume 1519, pp. 399–410. https://doi.org/10.1007/978-981-96-8901-9_35.
6. Alzeghoul, A.; Alsharari, N.M. Impact of AI Disclosure on the Financial Reporting and Performance as Evidence from US Banks. *J. Risk Financ. Manag.* **2025**, *18*, 4. https://doi.org/10.3390/jrfm18010004.
7. Anguren, R.; Jiménez, G.; Peydró, J.-L. Bank Capital Requirements and Risk-Taking: Evidence from Basel III. *J. Financ. Stab.* **2024**, *74*, 101292. https://doi.org/10.1016/j.jfs.2024.101292.
8. Artene, A.E.; Domil, A.E.; Ivascu, L. Unlocking Business Value: Integrating AI-Driven Decision-Making in Financial Reporting Systems. *Electronics* **2024**, *13*, 3069. https://doi.org/10.3390/electronics13153069.
9. Ajiga, D.I. Strategic Framework for Leveraging Artificial Intelligence to Improve Financial Reporting Accuracy and Restore Public Trust. *Int. J. Multidiscip. Res. Growth Eval.* **2021**, *2*, 882–892. https://doi.org/10.54660/.IJMRGE.2021.2.1.882-892.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

10. Shi, M. A Literature Review on the Application of Artificial Intelligence in Financial Statement Analysis. *Account. Mark. Organ.* **2025**, *1*, 1000062. https://doi.org/10.71204/dd26he23.
11. Odewuyi, O.M.; Shodimu, O.A.; Kazeem, O.; Phillips, A.P.; Okpo, S.A.; Ogundipe, A.O. Harnessing Artificial Intelligence to Revolutionize Corporate Finance and Financial Decisions in Strategic Consulting for Businesses. *Int. J. Sci. Res. Arch.* **2025**, *14*, 205–218. https://doi.org/10.30574/ijsra.2025.14.2.0353.
12. Elias, O.; et al. Harnessing Artificial Intelligence to Optimize Financial Technologies for Achieving Sustainable Development Goals. *World J. Adv. Res. Rev.* **2024**, *23*, 616–625. https://doi.org/10.30574/wjarr.2024.23.2.2385.
13. Alonge, E.O.; Dudu, O.F.; Alao, O.B. The Impact of Digital Transformation on Financial Reporting and Accountability in Emerging Markets. *Int. J. Sci. Technol. Res. Arch.* **2024**, *7*, 25–49. https://doi.org/10.53771/ijstra.2024.7.2.0061.
14. Faccia, A.; Petratos, P. NLP and IR Applications for Financial Reporting and Non-Financial Disclosure: Framework Implementation and Roadmap for Feasible Integration with the Accounting Process. In Proceedings of the 6th International Conference on Natural Language Processing and Information Retrieval (NLPIR 2022); ACM: New York, NY, USA, 2022. https://doi.org/10.1145/3582768.3582796.
15. Ciccola, R.; Guidi, M.; Chiucchi, M.S.; Giuliani, M. Artificial Intelligence and Sustainability Reporting in Practice: Projections from an Intellectual Capital Perspective. *VINE J. Inf. Knowl. Manag. Syst.* **2025**, 1–23. https://doi.org/10.1108/VJIKMS-08-2025-0360.
16. Kim, A.G.; Muhn, M.; Nikolaev, V.V. Financial Statement Analysis with Large Language Models. *SSRN Electron. J.* **2024**. https://doi.org/10.2139/ssrn.4835311.
17. Johri, A.; et al. Enhancing Audit Quality and Reducing Costs: The Impact of AI in Banking and Financial Services. *Front. Artif. Intell.* **2026**, *8*, 1718854. https://doi.org/10.3389/frai.2025.1718854.
18. Alruwaili, T.F.; Mgammal, M.H. The Impact of Artificial Intelligence on Accounting Practices: An Academic Perspective. *Humanit. Soc. Sci. Commun.* **2025**, *12*, 1197. https://doi.org/10.1057/s41599-025-05004-6.
19. Hussin, N.A.K.M.; Bukhari, N.A.N.M.; Hashim, N.H.A.N.; Bahari, S.N.A.S.; Ali, M.M. The Impact of Artificial Intelligence on the Accounting Profession: A Concept Paper. *Bus. Manag. Strategy* **2024**, *15*, 34. https://doi.org/10.5296/bms.v15i1.21620.
20. Thottoli, M.M. Leveraging Information Communication Technology (ICT) and Artificial Intelligence (AI) to Enhance Auditing Practices. *Account. Res. J.* **2024**, *37*, 134–150. https://doi.org/10.1108/ARJ-09-2023-0269.
21. Zaini, W.H.A.; Al-Ajeeli, E.S.M. Navigating the Future of Finance: How Artificial Intelligence Is Transforming Accounting. *Int. J. Comput. Exp. Sci. Eng.* **2025**, *11*. https://doi.org/10.22399/ijcesen.1648.
22. Vasarhelyi, M.A.; Halper, F.B. The Continuous Audit of Online Systems. *Audit. J. Pract. Theory* **1991**, *10*, 110–125.
23. Ibrahim, Y.; Diaa Moussa, N.; Elmarzouky, M.; Shohaieb, D.; Moubarak, H. Explainable Algorithmic Trading with Embedded Audit and Tax Trails: A Technology-Enabled Behavioral Oversight Model. *Int. J. Account. Inf. Manag.* **2026**, *34*, 550–585. https://doi.org/10.1108/IJAIM-10-2025-0351.
24. International Accounting Standards Board. *Conceptual Framework for Financial Reporting*; IFRS Foundation: London, UK, 2018.
25. Stratopoulos, T.C.; Wang, V.X. Artificial Intelligence and Accounting Research: A Framework and Agenda. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100760. https://doi.org/10.1016/j.accinf.2025.100760.
26. Mersha, M.; Lam, K.; Wood, J.; AlShami, A.K.; Kalita, J. Explainable Artificial Intelligence: A Survey of Needs, Techniques, Applications, and Future Direction. *Neurocomputing* **2024**, *599*, 128111. https://doi.org/10.1016/j.neucom.2024.128111.
27. Balasubramaniam, N.; Kauppinen, M.; Hiekkanen, K.; Kujala, S. Transparency and Explainability of AI Systems: Ethical Guidelines in Practice. In *Requirements Engineering: Foundation for Software Quality*; Gervasi, V., Vogelsang, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13216, pp. 3–18. https://doi.org/10.1007/978-3-030-98464-9_1.
28. Aurangzaib, R.; Iqbal, W.; Abdullah, M.; Bukhari, F.; Ullah, F.; Erradi, A. Scalable Containerized Pipeline for Real-Time Big Data Analytics. In Proceedings of the 2022 IEEE International Conference on Cloud

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

Computing Technology and Science (CloudCom); IEEE: Piscataway, NJ, USA, 2022; pp. 25–32. https://doi.org/10.1109/CloudCom55334.2022.00014.

29. Kokina, J.; Davenport, T.H. The Emergence of Artificial Intelligence: How Automation Is Changing Auditing. *J. Emerg. Technol. Account.* **2017**, *14*, 115–122. https://doi.org/10.2308/jeta-51730.
30. Raisch, S.; Krakowski, S. Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Acad. Manag. Rev.* **2021**, *46*, 192–210. https://doi.org/10.5465/amr.2018.0072.
31. Anthony, C. When Knowledge Work and Analytical Technologies Collide: The Practices and Consequences of Black Boxing Algorithmic Technologies. *Adm. Sci. Q.* **2021**, *66*, 1173–1212. https://doi.org/10.1177/00018392211016755.
32. Commerford, B.P.; Dennis, S.A.; Joe, J.R.; Ulla, J.W. Man versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence. *J. Account. Res.* **2022**, *60*, 171–201. https://doi.org/10.1111/1475-679X.12407.
33. Janowski, A. The Effectiveness of Life Insurance Sales Force Training: Welcome "Me and AI". *Economies* **2025**, *13*, 101. https://doi.org/10.3390/economies13040101.
34. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *MIS Q.* **2004**, *28*, 75–105. https://doi.org/10.2307/25148625.
35. Burns, B.; Grant, B.; Oppenheimer, D.; Brewer, E.; Wilkes, J. Borg, Omega, and Kubernetes. *ACM Queue* **2016**, *14*, 70–93. https://doi.org/10.1145/2898442.2898444.
36. Apache Software Foundation. Apache NiFi User Guide. Available online: https://nifi.apache.org/docs/nifi-docs/html/user-guide.html (accessed on 12 May 2026).
37. Oracle. General Ledger Trial Balance Report. Oracle Fusion Cloud Financials Documentation, Release 25C. Available online: https://docs.oracle.com/en/cloud/saas/financials/25c/ocuar/general-ledger-trial-balance-report.html (accessed on 12 May 2026).
38. SAP SE. Financial Statement Version. SAP S/4HANA Help Portal. Available online: [URL] (accessed on 12 May 2026).
39. MLflow Project. MLflow Tracking and MLflow Model Registry. MLflow Documentation. Available online: https://mlflow.org/docs/latest/(accessed on 12 May 2026).
40. Araci, D. FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models. *arXiv* **2019**, arXiv:1908.10063.
41. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. https://doi.org/10.1145/2939672.2939785.
42. Zhou, J.; San, O.T.; Liu, Y. Design and Implementation of Enterprise Financial Decision Support System Based on Business Intelligence. *Int. J. Prof. Bus. Rev.* **2023**, *8*, e0873. https://doi.org/10.26668/businessreview/2023.v8i4.873.
43. Bhatt, S.; Singh, P. A Comprehensive Review of AI-Enabled Financial Domain: Past, Present & Future Aspects. In Proceedings of the 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT); IEEE: Piscataway, NJ, USA, 2023; pp. 1–5. https://doi.org/10.1109/CISCT57197.2023.10351219.
44. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017; pp. 4765–4774.
45. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.
46. U.S. Securities and Exchange Commission. *Regulation FD (Fair Disclosure), 17 CFR Part 243*; SEC: Washington, DC, USA, 2000.
47. International Accounting Standards Board. *IFRS 18 Presentation and Disclosure in Financial Statements*; IFRS Foundation: London, UK, 2024.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 20 October 2026 https://doi.org/10.20944/preprints202610.1980.v1

48. International Auditing and Assurance Standards Board. *International Standard on Assurance Engagements (ISAE) 3000 (Revised), Assurance Engagements Other than Audits or Reviews of Historical Financial Information*; IFAC: New York, NY, USA, 2013.
49. Public Company Accounting Oversight Board. AS 2110, Identifying and Assessing Risks of Material Misstatement; AS 2201, An Audit of Internal Control over Financial Reporting That Is Integrated with an Audit of Financial Statements. Available online: https://pcaobus.org/oversight/standards/auditing-standards (accessed on 14 October 2026).
50. Public Company Accounting Oversight Board. *Amendments Related to Aspects of Designing and Performing Audit Procedures That Involve Technology-Assisted Analysis of Information in Electronic Form*; PCAOB Release No. 2024-007; PCAOB: Washington, DC, USA, 2024.
51. International Auditing and Assurance Standards Board. Strengthening the Risk-Based Audit Framework: Proposed Revisions to ISA 330, ISA 500, and ISA 520 (Exposure Drafts, August 2026). Available online: https://www.iaasb.org/focus-areas/strengthening-risk-based-audit-framework-proposed-revisions-isa-330-isa-500-and-isa-520 (accessed on 14 October 2026).
52. European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**, L 2024/1689.
53. European Union. Regulation (EU) 2026/1744 Amending Regulation (EU) 2024/1689 (Digital Omnibus on AI). *Off. J. Eur. Union* **2026**, L 2026/1744.
54. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*; NIST AI 100-1; NIST: Gaithersburg, MD, USA, 2023. https://doi.org/10.6028/NIST.AI.100-1.
55. OECD. *Recommendation of the Council on Artificial Intelligence*; OECD/LEGAL/0449; OECD: Paris, France, 2019 (amended 2024).
56. Anwar, A. Code and Data for "Artificial Intelligence–Driven Financial Reporting" [Data set and software]. Zenodo, 2026. https://doi.org/10.5281/zenodo.23438815.
57. Sorter, G.H. An "Events" Approach to Basic Accounting Theory. Account. Rev. 1969, 44, 12–19.
58. McCarthy, W.E. The REA Accounting Model: A Generalized Framework for Accounting Systems in a Shared Data Environment. Account. Rev. 1982, 57, 554–578.
59. Bertomeu, J. Machine Learning Improves Accounting: Discussion, Implementation and Research Opportunities. Rev. Account. Stud. 2020, 25, 1135–1155.
60. Bertomeu, J.; Cheynel, E.; Floyd, E.; Pan, W. Using Machine Learning to Detect Misstatements. Rev. Account. Stud. 2021, 26, 468–519. https://doi.org/10.1007/s11142-020-09563-8.
61. Liang, P.J.; Wang, A.; Akoglu, L.; Faloutsos, C. Pattern Recognition and Anomaly Detection in Bookkeeping Data. Asian Bureau of Finance and Economic Research (ABFER) Working Paper, 2021.
62. Sotiropoulos, K.; Zhao, L.; Liang, P.J.; Akoglu, L. ADAMM: Anomaly Detection of Attributed Multi-graphs with Metadata: A Unified Neural Network Approach. In Proceedings of the 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 15–18 December 2023; pp. 865–874.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.