INTERNATIONAL JOURNAL OF MANAGEMENT AND BUSINESS DEVELOPMENT (IJMBD)

eISSN: 3087-4084

Volume. 02, Issue. 08, pp. 07-11, august 2025"

Integrating Data Quality Governance and Advanced Anomaly Detection for AI-Driven Financial and Enterprise Information Systems

Dr. Alejandro Martínez

Department of Information Systems and Analytics, University of Barcelona, Spain

Article received: 01/07/2025, Article Revised: 14/07/2025, Article Accepted: 2/08/2025

ABSTRACT

The accelerating digitization of enterprises and financial institutions has resulted in unprecedented volumes, velocities, and varieties of data flowing through organizational information systems. While these developments have enabled sophisticated analytics, automation, and artificial intelligence–driven decision-making, they have simultaneously magnified the risks associated with poor data quality, weak governance structures, and undetected anomalies. Anomalies in enterprise and financial data—ranging from benign system glitches to critical indicators of fraud, cyber intrusion, or process failure—pose significant threats to organizational integrity, regulatory compliance, and stakeholder trust. This research article presents a comprehensive and theoretically grounded exploration of how data quality foundations, governance mechanisms, and modern anomaly detection techniques can be systematically integrated to support reliable, AI-driven financial and enterprise information systems.

Drawing strictly on established literature in anomaly detection, novelty detection, data quality, record linkage, data governance, enterprise system modernization, and recent advances in AI-assisted financial reporting, this study synthesizes diverse conceptual streams into a unified analytical framework. Classical perspectives on anomaly detection and novelty detection are revisited to establish foundational definitions and taxonomies, emphasizing their relevance across structured enterprise data, transactional financial records, and complex IT system logs. These perspectives are then connected to data quality theory, highlighting how dimensions such as accuracy, completeness, consistency, timeliness, and validity directly influence the performance and interpretability of machine learning–based anomaly detection models.

The article further examines entity resolution, record linkage, and matching dependencies as critical enablers of trustworthy anomaly detection, particularly in large-scale, heterogeneous enterprise environments. By integrating insights from ERBlox and foundational record linkage literature, the discussion demonstrates how unresolved entity ambiguity can propagate false anomalies or conceal genuine risks. In parallel, the role of data governance is explored as an institutional and organizational scaffold that ensures accountability, standardization, and ethical oversight for AI-driven anomaly detection systems.

Special attention is given to financial and accounting domains, where recent research highlights the growing adoption of deep learning architectures, autoencoders, generative models, and hybrid machine learning frameworks for anomaly detection in transactions, accounting entries, and IT systems. These approaches are critically analyzed not as isolated technical solutions, but as socio-technical systems whose effectiveness depends on data quality controls, governance maturity, and alignment with regulatory requirements such as multi-GAAP reconciliation and financial close processes.

Methodologically, this article adopts a qualitative, integrative research design grounded in extensive theoretical elaboration and cross-domain synthesis. Rather than proposing new algorithms, it provides a detailed interpretive analysis of how existing methods interact with organizational data ecosystems. The results are presented as a set of conceptual findings that articulate causal and reinforcing relationships between governance structures, data quality practices, and anomaly detection performance. The discussion situates these findings within broader debates on explainability, trust, scalability, and ethical AI in enterprise contexts, while also acknowledging limitations related to empirical validation and contextual variability.

The article concludes by arguing that the future of AI-driven financial and enterprise systems depends not on incremental improvements in detection accuracy alone, but on holistic integration of data quality theory, governance design, and anomaly detection methodologies. Such integration is positioned as essential for achieving resilient, transparent, and trustworthy information systems capable of supporting strategic decision-making in increasingly complex digital environments.

Keywords: Data Quality, Anomaly Detection, Financial Reporting, Data Governance, Machine Learning, Enterprise Information Systems

INTRODUCTION

The contemporary enterprise operates within an environment characterized by pervasive digitization, continuous data generation, and increasing reliance on automated decision-support systems. Financial institutions, multinational corporations, healthcare organizations, and public sector entities now depend on complex information infrastructures that process vast quantities of transactional, operational, and analytical data. While this transformation has delivered substantial efficiencies and analytical capabilities, it has also amplified the consequences of data errors, system failures, and anomalous behaviors. An anomaly—defined broadly as a pattern in data that does not conform to expected behavior—may signal anything from a simple data entry mistake to systemic fraud, cyberattack, or catastrophic process breakdown (Chandola, Banerjee, & Kumar, 2009).

Within this context, anomaly detection has emerged as a critical analytical function across domains. Early research framed anomaly detection as a statistical or algorithmic challenge, focusing on identifying outliers in numerical datasets. Over time, the scope of anomaly detection expanded to include complex temporal patterns, high-dimensional data, and dynamic systems, giving rise to related concepts such as novelty detection and change detection (Pimentel, Clifton, Clifton, & Tarassenko, 2014). In parallel, enterprise systems evolved from isolated, on-premise platforms to highly integrated, cloud-based ecosystems encompassing enterprise resource planning, human capital management, financial close systems, and real-time analytics (Padur, 2016; Routhu, 2017).

Despite these advances, a persistent and often underexplored challenge lies in the foundational quality and governance of the data upon which anomaly detection systems operate. Data quality research has long emphasized that analytical outcomes are only as reliable as the data inputs that generate them (Fan & Geerts, 2012; Herzog, Scheuren, & Winkler, 2007). In financial and enterprise contexts, poor data quality can obscure true anomalies, generate excessive false positives, and undermine confidence in automated systems. At the same time, weak data governance structures may leave organizations ill-equipped to manage the ethical, regulatory, and operational implications of AI-driven anomaly detection (Khatri & Brown, 2010).

Recent literature in financial analytics underscores the growing reliance on advanced machine learning and deep learning techniques for anomaly detection in accounting data, financial transactions, and IT systems (Schreyer et al., 2017; Schreyer et al., 2019; Miah, 2025; Tang et al., 2025). These methods promise unprecedented detection capabilities, yet they also introduce new dependencies on data consistency, semantic alignment, and interpretability. Moreover, emerging frameworks for AI-assisted financial reporting and multi-GAAP reconciliation further intensify the need for harmonized, high-quality data across jurisdictions and standards (Farooq, 2025; Kale, 2025).

The problem that motivates this article is not the absence of anomaly detection techniques, but the fragmentation of theoretical perspectives across data quality, governance, and machine learning literatures. Existing studies often address these domains in isolation, leaving practitioners and researchers without a cohesive understanding of how they interact within real-world enterprise systems. This fragmentation creates a literature gap at the intersection of data governance theory, data quality management, and AI-driven anomaly detection in financial and enterprise environments.

The objective of this article is to address this gap by developing a comprehensive, integrative analysis grounded strictly in established references. By synthesizing insights from anomaly detection surveys, data quality foundations, record linkage theory, governance design, enterprise system modernization, and contemporary financial analytics research, the article seeks to articulate a unified conceptual perspective. This perspective emphasizes that effective anomaly detection is not merely a technical endeavor, but a socio-technical process embedded within organizational structures, data lifecycles, and regulatory contexts.

METHODOLOGY

The methodological approach adopted in this research is qualitative, integrative, and theory-driven. Rather than employing empirical experimentation or quantitative modeling, the study relies on systematic conceptual synthesis of established academic and professional literature. This approach is particularly appropriate given the article's objective of developing a unified theoretical understanding across multiple domains that are often studied separately.

The first methodological step involved a close analytical reading of foundational anomaly detection literature. The survey by Chandola et al. (2009) was treated as a primary reference for establishing definitions, categories, and methodological paradigms in anomaly detection. Complementary perspectives from novelty detection research were incorporated to extend the analysis toward dynamic and evolving data environments (Pimentel et al., 2014). These sources provided the conceptual vocabulary and taxonomies necessary to discuss anomalies in enterprise and financial data contexts.

The second step focused on data quality theory and practice. Foundational texts on data quality and record linkage were analyzed to extract core dimensions, principles, and methodological considerations (Fan & Geerts, 2012; Herzog et al., 2007). Particular attention was paid to how data quality issues manifest in large-scale, heterogeneous datasets typical of enterprise systems, including issues of duplication, inconsistency, and semantic ambiguity. The ERBlox framework was examined as an illustrative case of integrating matching dependencies with machine learning for entity resolution (Bahmani, Bertossi, & Vasiloglou, 2016).

The third step involved the analysis of data governance literature. The work of Khatri and Brown (2010) was used to conceptualize governance as an organizational capability encompassing decision rights, accountability structures, and standards. This perspective enabled the integration of technical anomaly detection processes with institutional controls and policies.

The fourth step examined domain-specific literature in enterprise IT modernization and financial systems. Studies on network modernization, online patching, cloud-based HR systems, and enterprise upgrades were analyzed to contextualize anomaly detection within evolving technological infrastructures (Padur, 2016; Routhu, 2017). Financial analytics literature, including deep learning–based anomaly detection and AI-assisted reporting, was then incorporated to ground the discussion in contemporary applications (Schreyer et al., 2017; Schreyer et al., 2019; Farooq, 2025; Antwi, Adelakun, & Eziefule, 2024).

The final methodological step involved integrative synthesis. Concepts from all reviewed domains were systematically compared, contrasted, and connected to identify reinforcing relationships, tensions, and dependencies. The results of this synthesis are presented as descriptive findings rather than statistical outcomes, reflecting the conceptual nature of the research design.

RESULTS

The integrative analysis yielded several interrelated findings that collectively illuminate the role of data quality governance in AI-driven anomaly detection for enterprise and financial systems.

First, the analysis reveals that anomaly detection effectiveness is fundamentally constrained by data quality dimensions. Accuracy, completeness, consistency, timeliness, and validity are not abstract attributes but operational determinants of model performance. Inaccurate or inconsistent data can distort learned patterns, leading anomaly detection algorithms to normalize erroneous behavior while flagging legitimate transactions as suspicious (Fan & Geerts, 2012). This effect is particularly pronounced in unsupervised and semi-supervised learning contexts, where models rely heavily on historical data distributions to define normality (Chandola et al., 2009).

Second, entity resolution and record linkage emerge as critical but often underestimated prerequisites for reliable anomaly detection. In enterprise environments characterized by multiple data sources, duplicated records, and evolving identifiers, unresolved entity ambiguity can create artificial anomalies or mask genuine ones. The ERBlox framework illustrates how combining logical matching dependencies with machine learning can enhance entity resolution, thereby stabilizing the data foundation upon which anomaly detection operates (Bahmani et al., 2016). This finding underscores that anomaly detection accuracy is inseparable from the quality of upstream data integration processes.

Third, the results indicate that advanced machine learning techniques, including deep autoencoders and generative models, offer substantial advantages in capturing complex, non-linear patterns in financial and accounting data. Studies demonstrate that these models can identify subtle irregularities in high-dimensional datasets that traditional rule-based systems might overlook (Schreyer et al., 2017; Tang et al., 2025). However, the analysis also reveals that these advantages are contingent on well-governed data pipelines and transparent feature representations. Without governance oversight, the opacity of deep learning models may hinder interpretability and trust, particularly in regulated financial contexts.

Fourth, data governance structures play a mediating role between technical capabilities and organizational outcomes. Governance mechanisms define who is responsible for data quality, how anomalies are escalated and resolved, and how analytical models are aligned with regulatory requirements. The absence of clear governance can result in fragmented responses to detected anomalies, undermining the potential benefits of advanced analytics (Khatri & Brown, 2010).

Finally, the analysis highlights a convergence between anomaly detection and broader AI-assisted financial reporting initiatives. Frameworks for accelerating financial close processes and enabling multi-GAAP reconciliation increasingly rely on automated anomaly detection to ensure accuracy and compliance (Farooq, 2025; Kale, 2025). This convergence amplifies the strategic importance of integrating data quality management and governance into anomaly detection system design.

DISCUSSION

The findings of this study carry significant theoretical and practical implications. From a theoretical perspective, they challenge the implicit assumption in much of the anomaly detection literature that data quality and governance are external or secondary concerns. Instead, the analysis positions these elements as integral components of anomaly detection systems, shaping both their technical performance and organizational impact.

One important implication concerns the concept of normality itself. Anomaly detection algorithms learn what is normal based on historical data. If historical data reflect systemic errors, biased processes, or incomplete records, the learned notion of normality may be fundamentally flawed. This insight aligns with novelty detection research emphasizing the dynamic and context-dependent nature of normal behavior (Pimentel et al., 2014). Data quality management thus becomes a continuous process of recalibrating the baseline against which anomalies are defined.

Another implication relates to the trade-off between detection power and interpretability. Deep learning–based approaches excel at capturing complex patterns but often lack transparency. In financial and enterprise settings, where auditability and explainability are critical, this trade-off must be carefully managed. Data governance frameworks can provide mechanisms for balancing innovation with accountability by defining standards for model validation, documentation, and human oversight (Khatri & Brown, 2010).

The discussion also highlights limitations. The conceptual nature of this study means that findings are not empirically validated within a specific organizational context. While the integrative synthesis draws on robust literature, real-world implementations may face additional constraints related to organizational culture, legacy systems, and regulatory environments. Future research could address these limitations through case studies, empirical evaluations, and longitudinal analyses.

Future research directions include exploring how data quality metrics can be dynamically integrated into anomaly detection model training and evaluation, as well as examining governance models that support cross-functional collaboration between data scientists, domain experts, and compliance officers. Additionally, the ethical dimensions of anomaly detection, particularly in relation to automated decision-making and potential biases, warrant deeper investigation.

CONCLUSION

This article has presented a comprehensive, theoretically grounded exploration of the integration between data quality governance and advanced anomaly detection in AI-driven financial and enterprise information systems. By synthesizing literature across anomaly detection, data quality, record linkage, governance, enterprise modernization, and financial analytics, the study demonstrates that effective anomaly detection is not solely a matter of algorithmic sophistication. Rather, it is a socio-technical achievement that depends on high-quality data foundations, robust governance structures, and thoughtful alignment with organizational and regulatory contexts.

As enterprises continue to adopt AI-driven analytics and automation, the stakes associated with data anomalies will only increase. The insights offered in this article suggest that organizations seeking to harness the full potential of anomaly detection must invest not only in advanced models, but also in the less visible infrastructures of data quality management and governance. Through such holistic integration, anomaly detection can evolve from a reactive control mechanism into a strategic capability that enhances trust, resilience, and decision-making in complex digital ecosystems.

REFERENCES

1. Antwi, B. O., Adelakun, B. O., & Eziefule, A. O. (2024). Transforming financial reporting with AI: Enhancing accuracy and timeliness. International Journal of Advanced Economics, 6(6), 205–223.

2. Bahmani, Z., Bertossi, L., & Vasiloglou, N. (2016). ERBlox: Combining matching dependencies with machine learning for entity resolution. arXiv preprint arXiv:1602.02334.

3. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882

4. Fan, W., & Geerts, F. (2012). Foundations of data quality. Morgan & Claypool.

5. Farooq, M. (2025). The future of financial close: Leveraging AI and machine learning for faster, more accurate financial reporting. International Journal for Multidisciplinary Research, 7(1), 1–12.

6. Herzog, T. N., Scheuren, F. J., & Winkler, W. E.

(2007). Data quality and record linkage techniques. Springer. https://doi.org/10.1007/0-387-69505-2

7. Kahn, M. G., Raebel, M. A., Glanz, J. M., Riedlinger, K., & Steiner, J. F. (2012). A pragmatic framework for evaluating data quality in electronic health records. eGEMs, 1(1).

8. Kale, A. (2025). AI-assisted multi-GAAP reconciliation frameworks: A paradigm shift in global financial practices. Emerging Frontiers Library for The American Journal of Applied Sciences, 7(7), 67–77.

9. Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152. https://doi.org/10.1145/1629175.1629210

10. Miah, M. K. (2025). Anomaly detection in financial transactions using convolutional neural networks. Journal of Economics, Finance and Applied Statistics, 7(2), 195–210.

11. Padur, S. K. R. (2016). Network modernization in large enterprises: Firewall transformation, subnet re-architecture, and cross-platform virtualization. International Journal of Scientific Research & Engineering Trends, 2(5). https://doi.org/10.5281/zenodo.17291987

12. Pimentel, M. A. F., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. Signal Processing, 99, 215–249. https://doi.org/10.1016/j.sigpro.2013.12.026

13. Routhu, K. K. (2017). The evolution of HR from on-premise to Oracle Cloud HCM: Challenges and opportunities. International Journal of Scientific Research & Engineering Trends, 3(1). https://doi.org/10.5281/zenodo.17669776

14. Schreyer, M., Sattarov, T., Borth, D., & Reimer, B. (2017). Detection of anomalies in large-scale accounting data using deep autoencoder networks. arXiv preprint.

15. Schreyer, M., Sattarov, T., Schulze, C., Reimer, B., & Borth, D. (2019). Detection of accounting anomalies in the latent space using adversarial autoencoder neural networks. arXiv preprint.

16. Tang, T., Yao, J., Wang, Y., Sha, Q., Feng, H., & Xu, Z. (2025). Application of deep generative models for anomaly detection in complex financial transactions. arXiv preprint.

17. Wang, J., Liu, J., Pu, J., Yang, Q., Miao, Z., Gao, J., & Song, Y. (2019). An anomaly prediction framework for financial IT systems using hybrid machine learning methods. arXiv preprint.