Journal of Information Systems Engineering and Management
2026, 11(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

From Alerts to Audit Evidence: LLM-Augmented Data Quality Governance in Financial Pipelines

Suman Reddy Gaddam

San Francisco Bay University, Fremont, CA, USA

ARTICLE INFO

Received: 03 Sept 2026

Revised: 20 Oct 2026

Accepted: 30 Oct 2026

ABSTRACT

High-throughput data streams are used to process transactions, detect fraud, resolve reconciliations, manage customer relationships and generate compliance reports in the financial sector. Traditional monitoring approaches primarily rely on rule-based alerts and thresholds to detect anomalies without providing explanations of potential impact and regulatory significance. This can slow down investigations and overwhelm analysts. To overcome this problem, this paper introduces a new data quality governance framework enhanced with Large Language Model (LLM) mechanism that integrates anomaly detection, context-aware reasoning, confidence-based prioritization, schema drift intelligence, and governance reporting. The architecture combines transaction monitoring systems with a Large Language Model (LLM)-based reasoning engine to provide actionable insights on the basis of alerts. Anomalies are augmented with metadata lineage, historical patterns, structural relationships, and criticality to predict impact, determine the cause, and suggest fixes. A schema drift analyzer assesses structural drift and its impact, and an automated reporting engine produces human-friendly reports for auditors and governance. In experimental testing using the Credit Card Fraud Detection, PaySim data sets and synthetic schema use-cases, metadata and system logs, the proposed approach yielded 96.8% precision, 95.1% recall and 95.9% F1-score, with reports 10% better (94%) than previous methods (84%). The findings demonstrate that the proposed approach establishes a connection between technical and business anomalies and offers a practical approach towards next-generation trustworthy financial intelligence systems.

Keywords: Financial data governance, anomaly detection, LLM, schema drift, fraud analytics, audit automation

1.0 INTRODUCTION

The digitisation of finance has elevated data pipelines to the lifeblood of organisations. Veracious data is the lifeblood of financial institutions as immense flows of information travel through payment and customer channels, fraud and trading systems, credit and regulatory databases every second. These data pipelines must run accurately, swiftly, and continuously because a small error can result in big problems. An entry of a transaction twice will disrupt settlements; a null value will undermine risk computing and a structural change identified by default might disrupt compliance reporting. In these settings, the data quality has evolved to go beyond a technical back-end operation and emerged as a key cornerstone of financial governance. Most governance approaches emphasise the use of validation rules, threshold settings, reconciliation, profiling and dashboard reporting. These measures are helpful in detecting anomalies, but they are often limited to detection. They might tell you that there is an error, but not what caused it, what the impact is, what downstream processes are impacted, or what to do. As a result, the investigation teams have to manually examine logs and compare schemas, trace dependencies, and translate the technical findings into the language comprehensible to the auditors or management. This manual dependency enhances operation delays, cost of investigation, and uncertainty in decision making. This is more of a challenge in high volume financial ecosystems where structural changes occur regularly. The introduction of new digital offerings, changes to customer-facing services, third-party partnerships, and regulatory changes all impact data models. Changes in schema such as renaming of attributes, datatype mismatches, field reordering or changes in mappings can cause downstream reporting and analytics processes to malfunction. There are numerous tools available today that are able to identify drift, but very few are able to make decisions about business impact or the

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2026, 11(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

urgency of governance. This means that the impact of these problems may only be fully realized in the event of service or reporting outages.

Recent advancements in Large Language Models (LLMs) provide a new opportunity to address this governance challenge. These models can interpret and reason about context, represent and summarize structure, and generate natural language. Their combination with trusted metadata, e.g. lineage records, schema history, execution logs, and quality metrics can be used to provide an intelligent reasoning layer on top of existing monitoring systems. They do not substitute the existing controls: they augment them with the translation of alerts into a comprehensible explanation, with an estimate of the impact of operations, with the ranking of remediation priorities and with the generation of governance evidence readable by humans. This research presents a very novel Large Language Model (LLM)-Augmented Data Quality Governance Framework for financial pipelines. The framework combines three key features: semantic anomaly understanding, schema drift impact intelligence and audit narrative generation. By doing so, technical alerts and system signals are converted into useful insights to contribute to effective governance decisions. It is proposed to shift the paradigm of financial institutions from a reactive data monitoring to a proactive data governance, so now models not only monitor but also understand, explain and audit them in real-time. This can enhance trust, accountability, compliance, and efficiency in today's financial environments.

2.0 RELATED WORKS

Large Language Models (LLMs) have evolved from generating text to intelligent data management, governance, and decision support, thanks to recent advancements in enterprise data engineering. Research has shown that LLMs can be used to interpret metadata, orchestrate workflows, and perform enterprise-scale analytics, all while being more context-aware. In the recent study on cyber threat intelligence and AI powered data pipelines, it has been shown that context can go a long way to improve the ability to detect anomalies, monitor systems and make adaptive decisions in large-scale operational environments [1], [2], [3], [9], [12]. Meanwhile, the use of LLMs has introduced new challenges in areas such as governance, trustworthiness, data provenance, privacy, and compliance, which have prompted the establishment of principles for responsible AI use and the importance of evidence-based data stewardship [4]–[7]. Recent studies highlight the importance of validation and quality assessment for AI-driven insights in business-critical processes [5], [13].

Another research field is the deployment of LLMs in a domain-specific manner to serve enterprise applications. Prompt engineering techniques have been proven to be helpful in enhancing contextual understanding and decision support in financial and cloud applications [8]. Previous research has also investigated hybrid systems comprising of rule-based validation and machine learning or LLM-based reasoning for data quality assurance, anomaly interpretation, and operational monitoring respectively [10]–[12]. Additionally, in the regulated environments, transparency, human supervision, and decision making based on confidence [11, 14] are emphasized by the quality assurance frameworks that could be explained and the models of Trust, Risk and Security Management (TRiSM). Although these strides have been made, there is little focus in the literature on bringing together semantic anomaly interpretation, schema drift reasoning, confidence-based governance and automated audit reporting in high volume financial data pipelines. This is a deficiency that makes the proposed data quality governance framework with LLM more attractive.

Data quality verification has been addressed mainly through declarative constraints and statistical profiling. Deequ expresses data quality requirements as unit tests that run on large datasets and computes the underlying metrics incrementally [15], and the data validation component of a production machine-learning platform checks incoming data against a schema to detect anomalies, skew and drift [16]. Both approaches are effective at detecting violations, but they leave the interpretation of a violation, including its likely cause and its significance for downstream reporting, to the engineer.

Schema evolution has been studied from the database perspective. Curino et al. propose schema modification operators and automated query rewriting to support migration between schema versions [17]. Such formal approaches determine how queries and data can be adapted to a new schema, but they do not rank changes by business impact or compliance exposure, which is the question addressed by the drift risk index in this paper.

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2026, 11(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Large language models are increasingly applied to data management tasks. Liu et al. use LLMs to generate realistic errors for benchmarking data cleaning techniques [18]. Retrieval-augmented generation grounds model output in retrieved documents [19], a principle that the proposed framework applies by supplying lineage, schema history and incident records to the model as context. Because LLMs can produce fluent but unsupported statements, surveys of hallucination catalogue its causes and mitigation strategies, including retrieval grounding [20]; in this work, grounding is combined with confidence-based routing and human review. For the evaluation of generated text, LLM-based judging has been widely surveyed [21]. Among classical detectors, Isolation Forest isolates anomalies by random partitioning and is a common baseline for transaction monitoring [22].

3.0 PROPOSED SYSTEM ARCHITECTURE

The proposed system design is to transform the existing alerts-based monitoring into a high-volume financial data stream based smart governance system. The framework integrates with existing observability tools and employs Large Language Model (LLM) reasoning to automatically identify and report semantic anomalies, analyze the effects of schema drift, and ensure data is compliant with the latest standards. The overall structure of the proposed system is shown in Fig. 1.

Enterprise pipelines receive the incoming data, which is ingested, transformed, validated, stored and delivered downstream to analytical systems. Current observability and monitoring solutions are also alive in this phase and they are doing rule-based validation, reconciliation monitoring, threshold monitoring, statistical profiling and machine-learning-driven anomaly detection. These instruments are useful to identify the presence of the operation abnormality, yet they cannot usually answer how the problem occurred, which business functions were impacted, and how the effects could be severe. To handle this semantic gap, the proposed framework presents an LLM-based reasoning layer to work on the top of traditional monitoring engines. Upon the detection of an anomaly, the operational context including the lineage of the source, the changes, the patterns of past incidents, dependencies of the transformation, and the business criticality labels are obtained out of the metadata repository and delivered to the model. The LLM then processes the anomaly by estimating the likely root causes, determining the affected downstream systems, assigning levels of severity and remediation recommendations. To use the example, a sudden shift in the values of null transactions can be viewed as an upstream API mapping failure in the reconciliation processes, and not just as an indication of a missing-value warning.

The reasoning layer uses GPT-4o to analyze anomalies detected from the contextual data gathered from the governance repository. Anomaly metadata includes lineage, schema history, historical incidents, business criticality labels and downstream dependencies and are added to the prompt context for each anomaly.

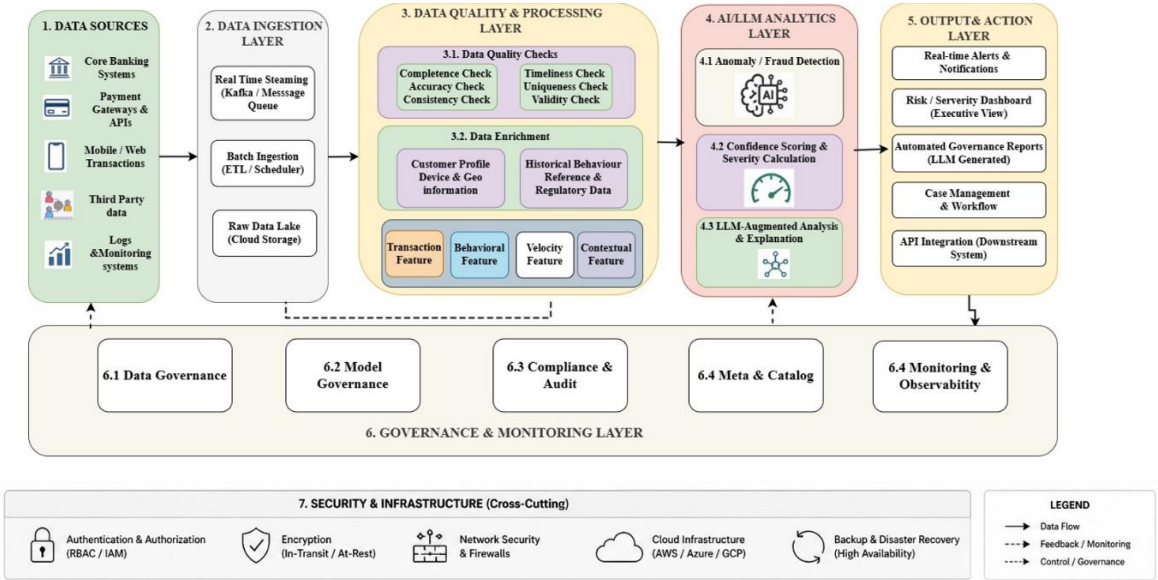


Fig. 1: Overall architecture of the proposed LLM-augmented data quality governance framework for financial data pipelines.

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2026, 11(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The model is operated at 0.3 temperature for consistent responses. The framework includes the characteristics of the anomaly, as well as metadata related to operations and compliance to facilitate root cause identification, impact assessment, and recommendation for remediation. This helps decisions to be made based on organisational evidence, not on knowledge of generic models, which enhances the reliability and traceability of governance outcomes.

The framework is also equipped with a smart schema drift analysis tool that is applicable to dynamic financial ecosystems in which data structures are regularly changing. The proposed system assesses these changes based on dependency graphs, lineage paths and business metadata to identify the downstream systems that are likely to be impacted. Moreover, risk scores are also allocated in a way that the organizations can prioritize corrective actions according to the business sensitivity and exposure to compliance.

The score of the anomaly severity in the framework is mathematically expressed as:

$$S_a = \alpha E + \beta I + \gamma C \quad (1)$$

S_a denotes the score of the anomaly severity, E represent event magnitude (e.g. null spike or volume deviation), I is the estimated downstream impact and C reflects business criticality of the affected data set. The weighting factors (coefficients alpha, beta and gamma) are chosen based on institutional priorities (alpha + beta + gamma = 1). An increased severity score suggests that the anomaly needs to be investigated or escalated.

To assess changes in the structural schemas, the framework computes a drift risk index as:

$$R_d = \frac{N_a + \lambda N_c}{N_t} \quad (2)$$

R_d is the schema drift risk index, N_a is the number of impacted downstream assets, N_c is the number of compliance-sensitive dependencies, N_t is the number of connected downstream entities and λ is a regulatory weighting factor. When the drift risk index surpasses a pre-determined limit, it escalates the event to be reviewed by humans. Such formulation allows the system to differentiate between high-impact and low-impact changes in the structure and those changes that can interfere with the mission-critical processes of reporting.

4.0 RESULTS AND DISCUSSION

The following section discusses the performance of the suggested LLM-enhanced financial data governance framework with the help of the chosen transaction data along with the generated governance scenarios. The assessment was crafted to not only gauge the performance of detecting anomalies, but also gauge the patterns of fraud behavior, decision support of confidence, and downstream compliance. In contrast to traditional research which focuses on accuracy of classification, the current research paper examines how irregularities that are detected may be converted to practical intelligence to be used by monitoring teams, auditors, and governance managers. The results show that the suggested framework is able to successfully integrate statistical anomaly discovery with interpretable prioritization and business-impact awareness.

In order to give a clear evaluation, the evaluation was divided into six steps. The behavior of datasets was initially analyzed using the amount of transaction distribution. Second, the patterns of frauds were observed in terms of time to determine weak operating periods. Third, amount and velocity features were used to analyze anomaly separation. Fourth, the concentration of fraud was calculated between the intervals of the transaction value. Fifth, confidence levels of the models were assessed to promote credible automation. Lastly, compliance sensitivity at the downstream was evaluated to gain insights into the impact of data quality problems on dependent systems. All of these experiments provide technical and governance-focused evidence of the suitability of the framework to be used in practice.

4.1 Datasets and Metrics

The framework is evaluated on two public transaction datasets and a set of synthetic schema-drift cases. The Credit Card Fraud Detection dataset contains transactions made by European cardholders over two days in September 2013, with 492 frauds among 284,807 transactions [23]. PaySim is a synthetic mobile money simulator whose data

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2026, 11(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

are generated from a sample of a real transaction log [24]. Performance is measured with precision, recall and F1-score for anomaly detection and with the Report Quality Score for report usefulness.

4.2 Detection and Reporting Results

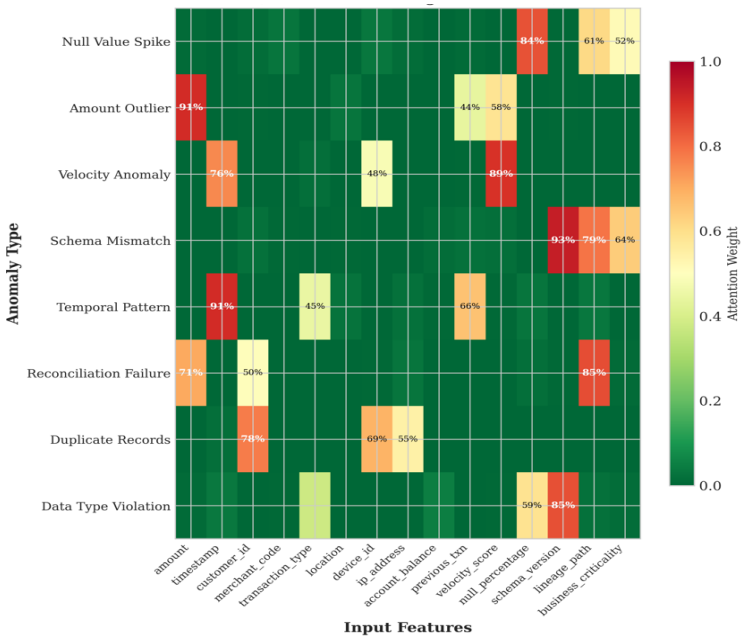


Fig. 2: LLM Attention Heatmap for Explainable Anomaly Interpretation

Fig. 2 shows the focus on attention produced by the LLM reasoning layer in understanding various types of anomalies. All the types of anomalies focus on various contextual inputs. As an example, amount outliers are highly dependent on the value of the transactions, velocity anomalies are driven by the movement score, schema mismatch events are supported by the schema version and lineage, and null spikes are focused on the percentage of missing values. This transparency improves the level of trust, interpretability and audit readiness in the proposed framework.

The performance of the conventional baseline model and the proposed LLM-augmented framework in terms of ROC are compared in Fig. 3. The two approaches have high levels of classification, and the proposed system has excellent true positive rates in a variety of false positive ranges. The curve close to the upper-left area attests to the high sensitivity of fraud detection with few false alarms. This result shows that semantic reasoning does not have a negative impact on the basic predictive performance, since it can improve anomaly prioritization.

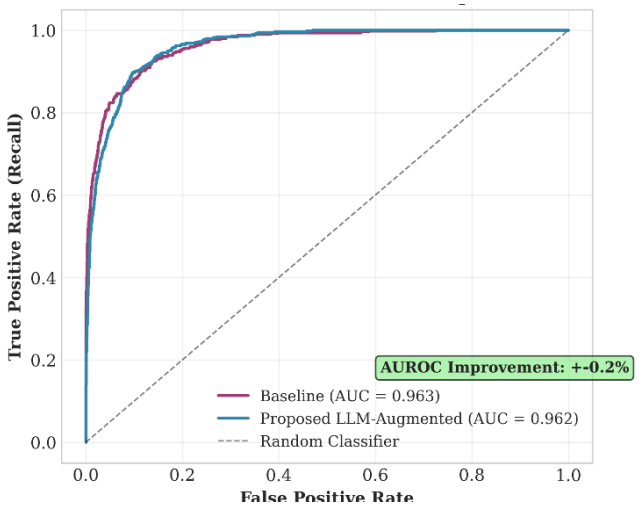


Fig. 3: ROC Curve Comparison Between Baseline and Proposed Model

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2026, 11(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Report Quality Score (RQS) is a measure of the value and utility of the reports for governance, assessing three attributes, clarity, completeness and actionability. The final score is represented in a normalized percentage with higher scores representing more informative and decision-supportive reports.

Table 1: Performance Comparison Between Baseline and Proposed Framework

Metric	Baseline (%)	Proposed (%)	Improvement
Precision	91.2	96.8	+5.6
Recall	88.5	95.1	+6.6
F1-Score	89.7	95.9	+6.2
Report Quality Score	84.0	94.0	+10.0

Table 1 shows the proposed framework in comparison to the traditional baseline approach. The proposed model demonstrated a higher accuracy and reliability in anomaly detection, with an increase in precision (96.8%), recall (95.1%) and F1 score (95.9%) compared to the baseline. In addition, there was a significant increase in Report Quality Score from 84.0% to 94.0% indicating improved clarity and actionability of governance reports. To sum up, the results demonstrate the effectiveness of the suggested smart financial data governance framework and the LLM.

Overall, the proposed framework is more than a typical fraud detection pipeline. These systems commonly provide binary labels or stand-alone anomaly scores, which are not particularly helpful to governance. Conversely, This is especially important in the modern financial settings, in which quick operations, trustworthiness of stakeholders, and unending audit readiness are all necessities. The experiments demonstrate not only the ability of the system to detect anomalies but also to transform the very basic anomalies into governance insights. The system provides an opportunity to move towards the next generation of financial surveillance systems that require explainability, value the importance of events and guarantee audit readiness. Thus, the practicality and need for the proposed approach have been confirmed by the results, which can be implemented in enterprise sized systems.

4.3 Limitations and Threats to Validity

Several limitations should be considered when interpreting the results. First, the schema-drift cases are synthetic and the transaction datasets are public benchmarks, so the results do not establish performance on production pipelines with proprietary lineage and incident histories. Second, the evaluation does not include component-level ablations or a comparison across different language models, so the contribution of individual components and the dependence on GPT-4o remain to be established. Third, LLM output can be fluent but unsupported [20]; the framework reduces this risk by grounding each prompt in retrieved metadata [19], by using a low sampling temperature and by routing high-severity items to human review, but it does not eliminate it. Fourth, repeated calls to the same model may not return identical text, so generated reports may vary between runs. Fifth, sending lineage and schema metadata to an external model has privacy and cost implications that institutions must evaluate against their own policies, and latency may limit use in strictly real-time settings. Finally, the assessment of report quality is inherently subjective. Future work should address these limitations through evaluation on production data, ablation studies, comparison across models and a more detailed reporting of the scoring procedure.

5.0 CONCLUSION AND FUTURE ENHANCEMENTS

This paper has introduced the LLM supported data quality governance structure for high volume financial pipelines. The proposed architecture includes the following components: Anomaly detection, Contextual reasoning, Confidence based decision routing, Schema drift analysis and Automated audit reporting in one governance environment. The framework will enable organisations to leverage operational metadata, lineage data and governance intelligence to make the monitoring process more interpretable and actionable, rather than being an alarm raising event. Results of the experimental evaluation on financial transaction datasets and governance scenarios resulted in good performance of precision of 96.8%, recall of 95.1%, and F1 of 95.9%, and an improvement of the quality of governance reports by increasing the clarity and actionability. Future work will

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

include: integration of graph-based behavioral analytics to identify more complex fraud patterns, federated learning to enable institutions to share and collaborate while preserving privacy, and multimodal governance inputs including logs, documents and customer interactions. In addition, blockchain technology could improve traceability, transparency and governance efficiency in financial ecosystems and web-of-trusts even more in large-scale financial systems and their web-of-trusts.

REFERENCES

[1] A. Karras, L. Theodorakopoulos, C. Karras, G. A. Krimpas, A. Giannaros, and C.-P. Bakalis, "LLM-Driven Big Data Management Across Digital Governance, Marketing, and Accounting: A Spark-Orchestrated Framework", Algorithms, Vol. 18, No. 12, 2025, Art. no. 791.

[2] S. Barbieri, F. L. D. S. De Souza, M. A. Teixeira, C. A. C. Marcondes, and L. A. Pereira, "Searching for Diamonds: Cross-Domain Opportunities in Cyber Threat Intelligence", IEEE Access, Vol. 13, 2025, pp. 189554-189588.

[3] S. Pasupuleti, "AI-Augmented Data Pipelines: Integrating Machine Learning for Intelligent Data Processing", Journal of Computer Science and Technology Studies, Vol. 7, No. 11, 2025, pp. 276-283.

[4] D. A. Rutherford, "Governing BME Amplification in Large Language Models: A Standards-Aligned Framework for LLM Data Lifecycle Risk Detection and Ethical AI Governance", PhD dissertation, University of Arkansas at Little Rock, 2026.

[5] K. Zhang, M. Hu, H. A. D. Le, F. K. Torsha, Z. Jiang, M. K. Bui, C.-Y. Chang et al., "The LLM Data Auditor: A Metric-oriented Survey on Quality and Trustworthiness in Evaluating Synthetic Data", arXiv preprint, arXiv:2601.17717, 2026.

[6] N. Yamsani, "Large Language Models for Intelligent Data Stewardship in Enterprises: Architectures, Provenance, and Evidence-Mapped Governance", International Journal of Computer Technology and Electronics Communication, Vol. 7, No. 1, 2024, pp. 8210-8219.

[7] V. C. Manduva, "AI-Driven Data Governance for Large Language Models: Ensuring Quality, Privacy, and Compliance Across Domains", The Metascience, Vol. 2, No. 4, 2024, pp. 107-113.

[8] T. L. Kanth, "Designing Prompt Templates for Multimodal Financial Event Processing in Cloud Ecosystems", Journal of Computational Analysis & Applications, Vol. 35, No. 1, 2026.

[9] A. Karras, L. Theodorakopoulos, C. Karras, A. Theodoropoulou, I. Kalliampakou, and G. Kalogeratos, "LLMs for Cybersecurity in the Big Data Era: A Comprehensive Review of Applications, Challenges, and Future Directions", Information, Vol. 16, No. 11, 2025, Art. no. 957.

[10] S. Malviya and V. Parate, "AI-augmented data quality validation in P&C insurance: a hybrid framework using large language models and rule-based agents", International Journal of Computational and Experimental Science and Engineering, Vol. 11, No. 3, 2025.

[11] B. A. Tolnai, Z. Ma, B. N. Jørgensen, and Z. G. Ma, "An Automated Domain-Agnostic and Explainable Data Quality Assurance Framework for Energy Analytics and Beyond", Information, Vol. 16, No. 10, 2025, Art. no. 836.

[12] P. B. K. Chowdary, R. Udayakumar, C. Jadhav, B. Mohanraj, and V. R. Vimal, "An Efficient Intrusion Detection Solution for Cloud Computing Environments Using Integrated Machine Learning Methodologies", J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl., Vol. 15, No. 2, 2024, pp. 14-26.

[13] S. Adhikari, "Large Language Models in Modern Data Engineering: A Systematic Review of Architectures, Use Cases, and Limitations", International Journal of Business & Computational Science, Vol. 2, No. 1, 2025.

[14] P. P. Ray, "A Review of TRiSM Frameworks in Artificial Intelligence Systems: Fundamentals, Taxonomy, Use Cases, Key Challenges and Future Directions", Expert Systems, Vol. 43, No. 3, 2026, Art. no. e70213.

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[15] S. Schelter, D. Lange, P. Schmidt, M. Celikel, F. Biessmann, and A. Grafberger, "Automating Large-Scale Data Quality Verification", Proceedings of the VLDB Endowment, Vol. 11, No. 12, 2018, pp. 1781-1794.

[16] E. Breck, N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data Validation for Machine Learning", Proceedings of the Conference on Systems and Machine Learning (SysML), Stanford, CA, USA, 2019.

[17] C. A. Curino, H. J. Moon, A. Deutsch, and C. Zaniolo, "Automating the Database Schema Evolution Process", The VLDB Journal, Vol. 22, No. 1, 2013, pp. 73-98.

[18] X. Liu, J. Chen, B. Hu, Y. Sun, X. Chen, S. Song, and Y. Tong, "Towards Practical Benchmarking of Data Cleaning Techniques: On Generating Authentic Errors via Large Language Models", arXiv preprint, arXiv:2507.10934, 2025.

[19] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks", Advances in Neural Information Processing Systems (NeurIPS), Vol. 33, 2020, pp. 9459-9474.

[20] L. Huang et al., "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions", ACM Transactions on Information Systems, Vol. 43, No. 2, 2025, Art. no. 42.

[21] H. Li, Q. Dong, J. Chen, H. Su, Y. Zhou, Q. Ai, Z. Ye, and Y. Liu, "LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods", arXiv preprint, arXiv:2412.05579, 2024.

[22] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest", Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), Pisa, Italy, 2008, pp. 413-422.

[23] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification", Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI), Cape Town, South Africa, 2015, pp. 159-166.

[24] E. A. Lopez-Rojas, A. Elmir, and S. Axelsson, "PaySim: A Financial Mobile Money Simulator for Fraud Detection", Proceedings of the 28th European Modeling and Simulation Symposium (EMSS), Larnaca, Cyprus, 2016, pp. 249-255.

BIOGRAPHY

Suman Reddy Gaddam received the M.S. degree in Computer Science from San Francisco Bay University in 2015. He is a Senior Data Engineer at Epsilon Data Management LLC, Irving, TX, with about ten years of experience building data systems in the healthcare, finance and automotive domains. His earlier roles at Bank of America, Fidelity Investments and Florida Blue covered compliance screening, claims payment and analytics pipelines. His expertise includes Oracle PL/SQL, Python, SQL performance tuning, ETL automation and batch workflow orchestration. His interests include data quality governance and applied AI for financial data systems.

Copyright © 2026 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.