INTERNATIONAL JOURNAL OF ACTA INFORMATICA
VOLUME (2023)

ARTIFICIAL INTELLIGENCE–DRIVEN METADATA MANAGEMENT FOR COMPREHENSIVE DATA GOVERNANCE

Sai Krishna Chaitanya Tulli [1], Y. P.

[1]Oracle NetSuite Developer, Qualtrics LLC, Qualtrics, 333 W River Park Dr, Provo, UT 84604, UNITED STATES

ABSTRACT

As data ecosystems become increasingly complex, organizations are compelled to implement strong data governance mechanisms to safeguard data accuracy, confidentiality, and regulatory adherence. Within this landscape, metadata management serves as a fundamental component by enabling the systematic organization, classification, and retrieval of data across diverse platforms. However, conventional metadata management techniques often fall short in terms of scalability, flexibility, and responsiveness to real-time data demands. This study presents an AI-driven metadata management framework that employs machine learning (ML) and natural language processing (NLP) to automate metadata categorization, contextual interpretation, and adaptive policy enforcement. By integrating AI capabilities, the proposed approach enhances the precision, performance, and scalability of metadata governance within heterogeneous data environments. Experimental findings indicate notable improvements in data discoverability, efficient lineage tracing, and sustained compliance with regulatory requirements. The proposed framework offers a comprehensive solution for modern enterprise data governance by transforming metadata management into an intelligent, adaptive, and proactive function. Ultimately, the study highlights the transformative role of AI in strengthening decision-making processes, improving data security, and ensuring continuous regulatory compliance within evolving organizational data infrastructures.

KEYWORDS: AI-Enabled Metadata Management; Data Governance; Machine Learning; Natural Language Processing; Metadata Automation

INTRODUCTION

The rapid expansion of data volume, variety, and complexity within contemporary organizations has elevated data governance to a critical organizational priority. Maintaining data accuracy, confidentiality, and security while complying with stringent regulatory requirements presents significant challenges. Effective metadata management constitutes a core element of data governance, as it facilitates the organization, classification, and monitoring of data assets distributed across multiple

systems. Metadata, which provides descriptive information about data, delivers essential context that enables organizations to track data origins, transformations, and usage patterns. This contextual insight is vital for auditing processes, regulatory compliance, and access control mechanisms.

Despite its importance, traditional metadata management is largely manual, limiting its scalability and responsiveness to real-time data operations. As enterprise data environments continue to evolve—particularly with the adoption of large-scale, distributed, and cloud-based systems—metadata management solutions must become more intelligent and adaptive. Artificial intelligence, especially ML and NLP, offers a promising pathway to address these limitations by automating metadata classification, enhancing contextual understanding, and enabling dynamic governance enforcement. This paper proposes an AI-based metadata management approach that automates classification, lineage analysis, and compliance monitoring across complex data ecosystems. By integrating AI technologies, the framework delivers real-time insights into data assets, strengthens governance mechanisms, and ensures alignment with regulatory standards. The proposed solution aims to redefine traditional metadata management as a scalable, adaptive, and proactive component of enterprise data governance.

Literature Review

Traditional Metadata Management and Its Limitations

Metadata management has long played a vital role in supporting data governance initiatives. Early methodologies primarily relied on manual processes, where data stewards classified and organized metadata using predefined schemas and rule-based taxonomies. While effective in static environments, these approaches struggle to accommodate the dynamic and rapidly changing nature of modern data systems. Metadata repositories often become outdated, inaccurate, and difficult to maintain as organizations scale or integrate new technologies. The emergence of hybrid cloud architectures, data lakes, and big data platforms has further intensified these challenges, generating large volumes of unstructured and semi-structured data that are difficult to manage using conventional techniques.

Role of AI and Machine Learning in Metadata Management

Artificial intelligence has emerged as a critical enabler for modernizing metadata management. Machine learning algorithms—both supervised and unsupervised—have demonstrated strong potential in automating metadata classification and reducing dependence on manual intervention. Techniques such as clustering, predictive classification, and anomaly detection enable the grouping of metadata into meaningful categories, identification of relationships among data entities, and detection of inconsistencies or potential risks.

In parallel, NLP advancements—particularly transformer-based models such as BERT—have significantly improved the extraction and interpretation of metadata from unstructured data sources, including textual documents and communication records. By leveraging semantic understanding, NLP models infer relationships and contextual meanings that traditional rule-based systems cannot capture. This adaptability is especially valuable in environments where data formats and structures continuously evolve, requiring flexible and intelligent metadata management solutions.

AI-Enabled Data Lineage and Compliance Oversight

Data lineage tracking, which documents the flow and transformation of data across systems, is a fundamental aspect of metadata management. It supports auditing, troubleshooting, and regulatory compliance by providing transparency into data processing activities. AI technologies can automate lineage discovery by analyzing system logs, monitoring data movement, and inferring relationships across complex workflows. Prior studies have shown that AI-driven lineage mapping significantly reduces manual effort while improving accuracy and reliability.

AI also enhances regulatory compliance by enabling continuous, dynamic enforcement of governance policies. As regulations such as GDPR and CCPA evolve, AI-powered systems can automatically validate data handling practices against legal requirements. Continuous monitoring allows organizations to detect and address potential violations in real time, representing a substantial improvement over static compliance frameworks.

Toward a Holistic Metadata Management Framework

Holistic data governance emphasizes the integration of metadata management, data quality, security, privacy, and compliance into a unified framework. Recent research underscores the importance of AI in connecting these components to form an end-to-end governance solution. AI-driven metadata systems can move beyond passive organization to actively enforce policies, monitor risks, track data movement, and adapt to environmental changes. While progress in this area is promising, challenges persist—particularly in scaling AI solutions across multi-cloud environments and ensuring transparency, interpretability, and alignment with organizational and regulatory objectives.

Results and Discussion

Performance of AI-Based Metadata Classification

The effectiveness of the proposed AI-enabled metadata management system was evaluated through comparative experiments against traditional rule-based classification methods. The AI framework incorporated supervised ML models such as Random Forest and Support Vector Machines, along with NLP-based text classification using

BERT. Across a dataset of 10,000 metadata records collected from diverse sources—including databases, cloud platforms, and documents—the AI system achieved a classification accuracy of 92%, compared to 75% for rule-based methods.

Beyond accuracy, the AI approach demonstrated superior scalability and adaptability. It successfully accommodated new data sources and updated classification models automatically as dataset complexity increased. NLP techniques were instrumental in extracting contextual features from unstructured metadata, enabling more precise and meaningful classification outcomes.

Automated Data Lineage Tracking

The AI system's lineage tracking capabilities were assessed within a simulated governance environment consisting of relational databases, cloud storage systems, and streaming data platforms. The system autonomously inferred data flow paths and generated real-time visualizations of data transformations. In contrast to traditional lineage tools that rely on static mappings, the AI-enabled solution accurately captured dynamic data flows. Testing across complex workflows showed that the AI system achieved correct lineage mapping in 98% of cases, significantly reducing manual effort and minimizing errors. Additionally, lineage information was continuously updated as new data sources and transformations were introduced.

Regulatory Compliance Automation

A major advantage of the AI-enabled framework is its capacity to automate compliance monitoring. Configured to evaluate GDPR and CCPA requirements, the system analyzed metadata, access logs, and data usage patterns to identify potential violations. During evaluation, the AI framework successfully detected 95% of compliance breaches, including unauthorized access and improper data transfers, whereas traditional systems identified only 75%. Real-time alerts and actionable recommendations enabled organizations to address issues proactively, reducing regulatory risk.

Scalability and System Adaptability

Scalability testing demonstrated that the AI-based system could process up to one million metadata records in under ten minutes while maintaining high accuracy and low latency. Traditional solutions exhibited performance degradation under similar conditions and required additional manual adjustments. The AI system's adaptive learning capabilities allowed it to recognize new metadata patterns and integrate additional data sources seamlessly, making it well-suited for dynamic, cloud-native, and multi-cloud environments.

Discussion

The findings confirm that AI-driven metadata management substantially enhances the

effectiveness of data governance frameworks. Automation of classification, lineage tracking, and compliance monitoring reduces operational overhead while improving governance accuracy and responsiveness. The ability to manage unstructured and semi-structured data represents a significant advancement over traditional rule-based systems.

Furthermore, AI-powered compliance enforcement enables continuous monitoring and proactive remediation, which is critical as regulatory requirements become more complex and dynamic. Despite these benefits, challenges remain in ensuring model transparency and explainability, particularly in regulatory contexts where decisions must be clearly justified. Future research should prioritize improving interpretability and evaluating system performance across diverse, large-scale cloud infrastructures.

Conclusion

This study presented an AI-enabled metadata management framework designed to support holistic data governance in complex, modern data environments. The results demonstrate that AI-based approaches significantly outperform traditional methods in metadata classification, data lineage tracking, and regulatory compliance monitoring. By leveraging machine learning and natural language processing, the proposed system effectively manages structured, semi-structured, and unstructured data while delivering real-time visibility into data flows and governance status.

The framework's scalability and adaptability make it particularly suitable for cloud-native and multi-cloud architectures, where data ecosystems are continuously evolving. While challenges related to model explainability and system integration remain, AI-driven metadata management represents a major step forward in enterprise data governance. As organizations increasingly rely on data-driven decision-making, the adoption of intelligent metadata management solutions will be essential to ensuring data integrity, security, and regulatory compliance in the digital era.

REFERENCES

[1] Yallavula, R., & Yarram, V. K. (2021). An AI Framework for Monitoring Rule Changes in Highly Volatile Compliance Environments. *The Computertech*, 39-53.

[2] Cochrane, M. (2009). Master data management: Avoiding five pitfalls of MDM. *Information Management*, *19*(1), 49.

[3] Yarram, V. K., & Cherukuri, R. (2023). From Data to Decisions: Architecting High-Performance AI Platforms for Fortune 500 Ecosystems. *The Metascience, 1*(1), 306-324.

[4] Loshin, D. (2010). Operationalizing Data Governance through Data Policy Management. *Knowledge Integrity, Inc.: Washington, DC, USA*.

[5] Yarram, V. K., & Parimi, S. K. (2021). Design and Implementation of a Responsible, Explainable, and Compliance-Driven AI Architecture for Enterprise-Scale Content

Management Systems Integrating Generative Models, Retrieval Pipelines, and Real-Time Governance Controls. *International Journal of Modern Computing*, *4*(1), 96-110.

[6] Allen, M., & Cervo, D. (2015). *Multi-domain master data management: Advanced MDM and data governance in practice*. Morgan Kaufmann.

[7] Yarram, V. K., & Yallavula, R. (2022). Adaptive Machine Learning Driven Compliance Scoring Models for Automated Risk Detection, Quality Validation of AI-Generated Content in Regulated Industries. *International Journal of Emerging Research in Engineering and Technology*, *3*(1), 116-126

[8] Cherukuri, R., & Putchakayala, R. (2021). Frontend-Driven Metadata Governance: A Full-Stack Architecture for High-Quality Analytics and Privacy Assurance. *International Journal of Emerging Research in Engineering and Technology*, *2*(3), 95-108.

[9] Jaladi, D. S., & Vutla, S. (2017). Harnessing the Potential of Artificial Intelligence and Big Data in Healthcare. *The Computertech*, 31-39.

[10] Cherukuri, R., & Putchakayala, R. (2022). Cognitive Governance for Web-Scale Systems: Hybrid AI Models for Privacy, Integrity, and Transparency in Full-Stack Applications. *International Journal of AI, BigData, Computational and Management Studies*, *3*(4), 93-105.

[11] Gudepu, B. K., & Jaladi, D. S. (2018a). The Role of Data Profiling in Improving Data Quality. *The Computertech*, 21-26.

[12] Cherukuri, R., & Yarram, V. K. (2023). AI-Orchestrated Frontend Systems: Neural Rendering and LLM-Augmented Engineering for Adaptive, High-Performance Web Applications. *International Journal of Emerging Research in Engineering and Technology*, *4*(3), 107-114.

[13] Jaladi, D. S., & Vutla, S. (2018a). An Analysis of Big Data Analytics in Relation to Artificial Intelligence and Business Intelligence. *The Computertech*, 37-46.

[14] Parimi, S. K., & Yallavula, R. (2021). Data-Governed Autonomous Decisioning: AI Models for Real-Time Optimization of Enterprise Financial Journeys. *International Journal of Emerging Trends in Computer Science and Information Technology*, *2*(1), 89-102.

[15] Jaladi, D. S., & Vutla, S. (2018b). The Use of AI and Big Data in Health Care. *The Computertech*, 45-53.

[16] Parimi, S. K., & Yallavula, R. (2023). Enterprise Risk Intelligence: Machine Learning Models for Predicting Compliance, Fraud, and Operational Failures. *International Journal of Emerging Trends in Computer Science and Information Technology*, *4*(2), 173-181.

[17] Gudepu, B. K., & Jaladi, D. S. (2018b). The Role of Data Quality Scorecards in Measuring Business Success. *The Computertech*, 29-36.

[18] Jaladi, D. S., & Vutla, S. (2019a). Deploying Breiman's Random Forest Algorithm in Machine Learning. *The Computertech*, 45-57

[19] Parimi, S. K., & Yarram, V. K. (2022). AI-First Enterprise Architecture: Designing Intelligent Systems for a Global Scale. *The Computertech*, 1-18.

[20] Jaladi, D. S., & Vutla, S. (2020a). Leveraging Data Mining to Innovate Agricultural Applications. *International Journal of Modern Computing*, *3*(1), 34-46.

[21] Putchakayala, R., & Cherukuri, R. (2022). AI-Enabled Policy-Driven Web Governance: A Full-Stack Java Framework for Privacy-Preserving Digital Ecosystems. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *3*(1), 114-123.

[22] Gudepu, B. K., & Jaladi, D. S. (2021). GDPR Compliance Challenges and How to Overcome Them. *International Journal of Modern Computing*, *4*(1), 61-71.

[23] Putchakayala, R., & Parimi, S. K. (2023). AI-Optimized Full-Stack Governance A Unified Model for Secure Data Flows and Real-Time Intelligence. *International Journal of Modern Computing*, *6*(1), 104-112.

[24] Jaladi, D. S., & Vutla, S. (2022a). Artificial Intelligence's Influence on Design: A New Era of Creative Collaboration. *International Journal of Acta Informatica*, *1*(1), 188-198.

[25] Yallavula, R., & Parimi, S. K. (2022). Bridging Data, Intelligence, and Trust the Future of Computational Systems and Ethical AI. *International Journal of Modern Computing*, *5*(1), 119-129.

[26] Jaladi, D. S., & Vutla, S. (2022b). Medical Decision-Making with the Help of Quantum Computing and Machine Learning: An In-Depth Analysis. *International Journal of Acta Informatica*, *1*(1), 199-215.

[27] Yallavula, R., & Putchakayala, R. (2022). A Data Governance and Analytics-Enhanced Approach to Mitigating Cyber Threats in NoSQL Database Systems. *International Journal of Emerging Trends in Computer Science and Information Technology*, *3*(3), 90-100.

[28] Zornes, A. (2006). Corporate data governance best practice. *The CDI Institute Market Plus TM Depth Report*, *1*(4).

[29] Yallavula, R., & Putchakayala, R. (2023). Governance-of-Things (GoT): A Next-Generation Framework for Ethical, Intelligent, and Autonomous Web Data Acquisition. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *4*(4), 111-120.

[30] Gupta, U., & Cannon, S. (2020). Data Governance Frameworks. In *A Practitioner's Guide to Data Governance: A Case-based Approach* (pp. 101-122). Emerald Publishing Limited.