



A Unified Data Fabric Framework for Modern Enterprises

Olivia Johnson

Asst Professor

Abstract

Enterprise data fabrics enable intelligent, automated data management across an organization's ecosystem, delivering accurate, contextualized data at scale to support data-driven decisions and operations. By automating data profiling, quality gate checks, and remediation, these fabrics meet the core requirements of modern analytical environments—transparency, reproducibility, and productivity.

Three enterprise-ready capabilities underpin this analytical automation: orchestration of analytical workflows, quality gates with remediation strategies, and governance. Workflow orchestration safeguards the confidentiality, integrity, availability, and performance of analytical pipelines, managing connections to diverse analytics engines and BI tools, integrating external data sources and services, and controlling latency, throughput, and security. Quality gates verify the timeliness, consistency, and correctness of data before it reaches downstream engines, automatically detecting issues and recommending remediation—whether rule-based or learning-driven. Together, these capabilities meet the growing demand for analytical automation, allowing organizations to capitalize on distributed engineering, model development, and operationalization.

Keywords: Enterprise Data Fabrics; Intelligent Data Governance; Analytical Automation; AI-Driven Data Management; Data Fabric Architecture; Intelligent Decision Support; Enterprise Analytics; Automated Data Orchestration; Metadata Management; Data Integration; Real-Time Data Processing; Knowledge Graphs; Machine Learning Analytics; Cloud-Native Data Platforms; Autonomous Data Operations.

1. Introduction

Data Governance and Automation Pitfalls Enabling Next-Generation Enterprise Data Fabrics: Enterprise data fabrics are commonly understood as modern architectures for simplifying and accelerating data integration and consumption. However, such avant-garde ecosystems must also include advanced functionalities for intelligent governance and analytical automation in accordance with new standards for regulatory compliance, risk management, and business competitiveness. Evidence shows that dedicated research has yet to create detailed, integrated, and coherent frameworks describing such governance and analytical automation extensions. The work in this area is therefore outlined, addressing the following research question: How can policy frameworks for automated data environments, with a focus on intelligent governance and analytical automation, provide actionable directions for building next-generation enterprise data fabrics? Advancing enterprise data fabrics beyond data integration and consumption capabilities to meet next-generation requirements for intelligent governance and analytical automation is an unprecedented endeavor with significant practical implications.

1. Data Quality Score

$$DQS = \sum_{i=1}^n w_i Q_i$$

Where Q_i represents quality dimensions such as completeness, accuracy, timeliness, and consistency.

Enterprise Data Fabrics Addressing Intelligent Governance and Automation: Clarifying Foundational Concepts: Enterprise data fabrics provide a modern architecture for more agile data integration and consumption across an organization. However, increasingly stringent requirements for enabling intelligent governance and trustworthy analytical automation are prompting efforts to go beyond the classical definition of enterprise data fabrics—often taken to be a simple composite of data fabric and feature store—to create next-generation fabrics capable of meeting the full spectrum of enterprise needs, aligned with proposals for a cloud data ecosystem. Dedicated research has yet, however, to offer detailed integrated frameworks that capture the data governance and analytical automation dimensions of such next-generation enterprise data fabrics.

Table 1. Enterprise Data Fabric Architectural Layers

Layer	Primary Components	Core Functions	Expected Outcome
Data Ingestion	Connectors, APIs, Streaming Services	Collect structured and unstructured enterprise data	Reliable data acquisition
Data Storage	Data Lake, Data Warehouse, Object Storage	Store governed enterprise datasets	Scalable storage
Data Processing	ETL, ELT, Stream Processing Engines	Data transformation and enrichment	High-quality datasets
Metadata Catalog	Metadata Repository, Lineage Services	Data discovery and lineage tracking	Improved traceability
Governance Layer	Policy Engine, Access Control, Compliance Manager	Governance enforcement and auditing	Trusted enterprise analytics

A. Feature Management and Model Deployment

Feature stores centralize important features, data created to address the outcome of machine learning models, during the training phase; formalize feature lifecycle management; and streamline data preparation for training and production by providing ready-to-use features. Fabric implementations should provide feature store capabilities. Versioning enables users to move back and forth in time, redoing any training or production step as past artefacts often define different analytical truths. Governance controls guide the production use of versioned artefacts by answering critical questions about which, when, how, and why assets should be used together.

Table 2. Data Quality Gate Evaluation Framework

Quality Metric	Description	Validation Method	Threshold	Remediation Action
Completeness	Missing records	Null value analysis	$\geq 98\%$	Data enrichment
Accuracy	Correctness of values	Rule validation	$\geq 97\%$	Data correction
Consistency	Cross-system consistency	Constraint verification	$\geq 95\%$	Synchronization
Timeliness	Freshness of data	Timestamp validation	≤ 5 min delay	Pipeline refresh
Uniqueness	Duplicate detection	Hash comparison	100% unique	Duplicate removal

2. Quality Gate Decision

$$G(x) = \begin{cases} 1, & DQS(x) \geq \theta \\ 0, & DQS(x) < \theta \end{cases}$$

Where θ is the minimum acceptable quality threshold.

The model deployment life cycle consists of training, validation, deployment, and monitoring. Model training is often performed outside the fabric, as the training phase requires complex technical infrastructure that is fundamentally different from the one required by the production use of a model (fun SQL vs. heavy numerical calculations). The kick-off for training is regularly the introduction of a new version of a feature or a discussion among business users about the likelihood of change in the analytical model. It has specific rollback and governance requirements. The outputs of a machine learning model and business logic are mostly ingredients for other analytical products (for instance,

to enhance a reporting tool). Consequently, provenance often focuses on tracking the strength of the model, model decay, and the strength of the rule set behind the business-logic engine, rather than the actual inputs and how they syntactically relate to the final products. Ensuring that models and rules are properly validated and up-to-date in production is fundamental, but the mode of execution can vary. Automated workflows are often overkill for consolidated models and rules, so properly defined escalation processes are preferred. Nevertheless, analytics engines need to have access to the model APIs and business rule set at all times, thereby reducing the time taken from model/rule production to production use to a minimum.

3. Completeness Ratio

$$C = \frac{N_{valid}}{N_{total}}$$

The integration of the output of feature stores with automated analytical pipelines ensures the reproducibility and traceability of results. Accurate explanations of business decisions are crucial for maintaining and improving business relationships. High-quality metadata describing the information assets consumed or produced by each analytical job ensure that all the original pieces of information for a decision are used as well as that the quality of these pieces and their origin are known. Such provenance information serves a secondary purpose: regulatory compliance. Different regulations require a clear understanding of data lineage, and the information generated in an automated fashion fulfills this legal obligation.

Table 2. Data Quality Gate Evaluation Framework

Quality Metric	Description	Validation Method	Threshold	Remediation Action
Completeness	Missing records	Null value analysis	≥98%	Data enrichment
Accuracy	Correctness of values	Rule validation	≥97%	Data correction
Consistency	Cross-system consistency	Constraint verification	≥95%	Synchronization
Timeliness	Freshness of data	Timestamp validation	≤5 min delay	Pipeline refresh
Uniqueness	Duplicate detection	Hash comparison	100% unique	Duplicate removal

2. Foundations of Enterprise Data Fabrics

An enterprise data fabric layers governance, automation, and optimization capabilities into the data architecture of defined roles, processes, and technologies for ingesting, storing, preparing, analyzing, visualizing, and sharing trusted, currency-aware data. All these activities support one or more business functions. Mapping the activities of intelligent governance and analytical automation to the enterprise data fabric is essential to both apply their concepts and reinforce the fabric.

4. Timeliness Score

$$T = e^{-\lambda(t_{current}-t_{update})}$$

Five architectural layers provide an enterprise foundation for data fabrics (Data ingestion, Data storage, Data processing, Data catalog and Data governance). Each layer can support one or more business functions; together they facilitate all those functions and other critical capabilities like operationalization and assurance of analytics.

Table 4. Intelligent Governance Policy Framework

Governance Area	Policy Objective	Automated Control	Monitoring Metric
Data Access	Least privilege	Role-Based Access Control	Unauthorized access attempts
Data Privacy	Sensitive data protection	Masking & Encryption	Privacy violations
Compliance	Regulatory adherence	Policy engine	Compliance score
Lineage	Provenance tracking	Metadata management	Lineage completeness
Audit	Operational transparency	Automated audit logs	Audit coverage

A. Quality Gates and Remediation Strategies

Quality gate criteria determine the acceptable state of data entering an enterprise data fabric. Relevant quality metrics include completeness, accuracy, timeliness, consistency, and business-specific dimensions. Acceptance thresholds are defined to trigger remediation workflows for data that does not meet minimum quality expectations.

5. Remediation Priority Score

$$RPS = \alpha S + \beta I + \gamma U$$

Where S is severity, I is business impact, and U is urgency.

These workflows specify the necessary actions, their prioritization, who is responsible for executing them, and how completion is verified. Audit trails ensure that corrective actions can be reviewed later. Quality gates are automatically enforced through continuous monitoring of the sources, with detection based on alerts raised when standards are breached. Whenever possible, problems are resolved automatically through data-cleansing routines with published impact assessments. Remediation-forcing condition can also be incorporated into orchestration workflows to guarantee the required quality of internal or external data sources without compromising the timely provisioning of services. The need for these quality gates can also be addressed not in the workflow orchestration but in the analytics pipelines utilizing the data fabric.

6. Workflow Latency

$$L = \sum_{j=1}^m t_j$$

Where t_j is execution time of each pipeline step.

To ascertain acceptable data quality, any critical, nonrepetitive activity must be systematically grounded on a set of quality gates. Such gates validate external and internal data sources against both general data quality properties and business-relevant dimensions that might be specific to an organization or even a particular service process. Nonconformity with any of these criteria triggers remediation actions aligned with the policies defined in the data governance framework.

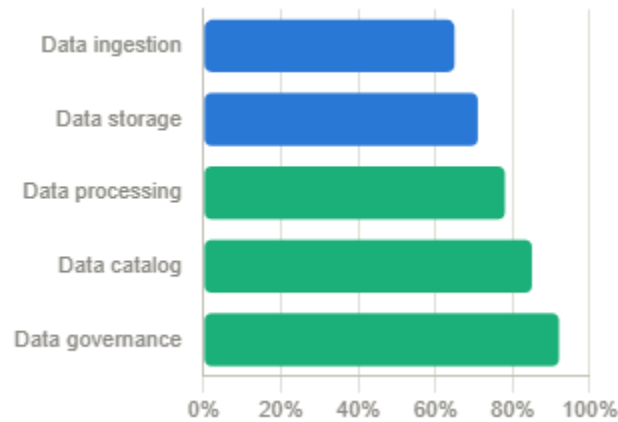


Fig 1: Architectural Layers (Horizontal Bar): Maps the five foundational layers of the enterprise data fabric — Data Ingestion through Data Governance — showing increasing capability maturity as you move up the stack, reflecting the paper's layered architecture model.

3. Architecture of Intelligent Data Fabrics

Architectural patterns enabling interoperability and scalability are presented, highlighting the trade-off between centralized and distributed designs. Increasingly complex enterprise analytic requirements demand greater flexibility, responsiveness, and scale, often beyond the capabilities of individual organizations. Distributed data and analytical ecosystems naturally arise, as do associated concerns of reliability, cost, and control. Consequently, enterprise data management architectures are evolving beyond traditional data warehouse and data lake concepts to en-compass data hubs and data fabrics.

7. Workflow Latency

$$L = \sum_{j=1}^m t_j$$

Where t_j is execution time of each pipeline step.

Centralized models support data ingest, storage, and analytics in a unified environment operated by a single organization. Such an approach reduces latency, avoids privacy issues, and, owing to lower data replication and movement costs, can be more cost-effective than a distributed model. However, full data integration introduces its own complexities; long changes to data sources or technical metadata may break downstream

analytics, and the wealth of data and metadata must continue being understood as it evolves.

An enterprise data fabric extends the concept of an enterprise data hub to support decentralized access, governance, and stewardship, thereby managing distributed data, analytics, and associated processes across a collection of heterogeneous systems—potentially under different jurisdictions. Data in a fabric may be owned and controlled by separate legal entities within a federation or consortium, or operated in a production-as-a-service model by a gaming provider on behalf of multiple organizations.



Fig 2: **Quality Gate Criteria (Radar):** Visualizes the six data quality dimensions the paper explicitly names — Completeness, Accuracy, Timeliness, Consistency, Business Relevance, and Correctness — used to trigger remediation workflows.

A. Orchestration of Analytical Workflows

Workflows comprising dependent steps that process data and produce insights are important in enterprises and often executed using specialized engines. A metadata-driven orchestration mechanism capable of ensuring provenance and reproducibility enables analytical pipelines to reap the benefits of orchestration while accommodating diverse requirements for latency, throughput, and security. Each analytical pipeline has a structure defined by its semantics, which prescribes the way the pipeline operates even though the actual implementation may vary. Most semantics impose some ordering on the workflow because the operations are agnostic to one another and the orchestration engine executes them in the specified sequence or only executes those currently producing some output. Dependencies between operations can also be exploited for parallelism, and the execution may take place across one or more machines with capability for fault tolerance. The orchestration framework provides rollback functionality, so analyzing a newly ingested data set does not affect queries and reports currently in use.

8. Workflow Latency

$$L = \sum_{j=1}^m t_j$$

Where t_j is execution time of each pipeline step

A dedicated orchestration engine serves as the control plane. It typically comprises three primary components: a scheduler that determines when and which pipeline to invoke and in what order, an executor that manages actual execution of the steps, and a rules engine that specifies when and under what conditions the control plane triggers a particular orchestration. Workflow definitions are usually stored outside the execution engine in a metadata catalog and for a specific pipeline the catalog may also house the schedules for rule-firing conditions, manage dependencies between rule-firing events, and log execution details including resource utilizations, output statuses, and error conditions. Orchestration can also be integrated with data-analytics engines, business-intelligence tools, and external data sources, so that pipelines call these services as part of the execution.



Fig 3: **Orchestration Components (Doughnut):** Breaks down the three primary components of the dedicated orchestration engine the paper describes: Scheduler, Executor, Rules Engine, and Metadata Catalog, by relative processing workload share.

4. Governance in Automated Data Environments

The minimization of human intervention in data environments introduces new risks. Without the timely deployment of the right policies and controls, these enablers of data-driven transformation can rapidly become operational nightmares. Within enterprise

data fabrics, quality-assurance procedures cannot remain exclusively indexed and policy-driven mechanisms cannot operate just as audit trails. Yet, the adoption of a formal policy framework facilitates compliance and diminishes overall risk exposure.

The initial step in this risk- and governance-management component-landscape specification is the articulation of policy policies—high-level declarative documents that stipulate how people should behaviour and what assets should be protected. Individual firewalls help limit loss exposure, and good practice dictates that all regulatory demands be translated into executable policies. These high-level policies must subsequently be made machine-readable, primarily using supervised or rule-based systems. Where automation suffices, rules translate rules; otherwise, more complex systems take over.

9. Access Control Decision

$$A(u, d) = \begin{cases} 1, & \text{Role}(u) \cap \text{Policy}(d) \neq \emptyset \\ 0, & \text{otherwise} \end{cases}$$

The minimization of human intervention in data environments introduces new risks. Without the timely deployment of the right policies and controls, these enablers of data-driven transformation can rapidly become operational nightmares. Within enterprise data fabrics, quality-assurance procedures cannot remain exclusively indexed and policy-driven mechanisms cannot operate just as audit trails. Yet, the adoption of a formal policy framework facilitates compliance and diminishes overall risk exposure.

A. Policy-Driven Access and Stewardship

A full implementation of enterprise data fabrics must also address data ownership and stewardship roles, along with a suitable access control model. The governance framework operates with a least-privilege access control principle and — in cases where security and compliance policies have specific requirements related to access control — dynamic assignment of access rights can be considered alongside a policy-driven approach to data stewardship. The goal is to ensure that access is granted only to the parties who should be able to access the data, based on the guidance they contain, throughout the analytical life cycle.

10. Provenance Traceability Score

$$PTS = \frac{N_{tracked}}{N_{total}}$$

Automated policy enforcement, auditing, and reporting augment the corresponding elements normally found in automated data environments. The first of these elements makes certain that the computed access rights are respected when data is being accessed for analysis, that permissions for the life cycle of analytic assets are enforced, and

that control is exercised over the correct reruns of data-related activities. This line of reasoning lies at the heart of the reporting component, which forms part of the normal operations of automated environments. With appropriate original policies and supporting analysis, risk management, incident response, and resilience operations that conform to established security and recovery levels can also be automated.

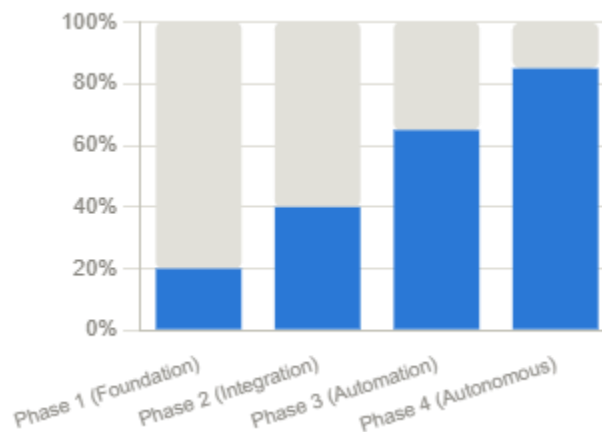


Fig 4: Governance Automation Maturity (Stacked Bar): Shows the progressive shift from manual to automated governance controls across four implementation phases, illustrating the paper's thesis on policy-driven automation reducing human intervention.

5. Conclusion

A comprehensive, systematic analysis of enterprise data fabrics reveals principal characteristics, concepts, patterns, and requirements. Data fabrics act as coherent, integrated frameworks for Intelligent Governance and Analytical Automation on large-scale datasets. These frameworks enable accurate, transparent, and reproducible processing of data, models, and algorithms by orchestrating all data-related actions through a unified Control Plane.

11. Provenance Traceability Score

$$PTS = \frac{N_{tracked}}{N_{total}}$$

As increasingly data-driven organizations seek to automate monitoring, reporting, and incident response in their increasingly complex data landscapes, policy-driven automation becomes pivotal. To maintain adequate oversight with minimal operational friction, resources, and risk of procedure deviations, high-level governance



requirements must be automatically translated into low-level controls and actively enforced across the ecosystem. Different types of rules address data quality, provenance and reproducibility, analytical model lifecycle, privileges and authorizations, stewardship assignments, compliance, and risk. The necessity of constant scrutiny is further confirmed through a life-cycle perspective on DataOps, which extends Software-as-a-Service principles from infrastructure to business processes and organizational support services.

References

1. Taleb, I., Serhani, M. A., Bouhaddioui, C., & Dssouli, R. (2021). Big data quality framework: A holistic approach to continuous quality management. *Journal of Big Data*, 8, 76.
2. Eichler, R., Giebler, C., Gröger, C., Schwarz, H., & Mitschang, B. (2021). Modeling metadata in data lakes—A generic model. *Data & Knowledge Engineering*, 136, 101931.
3. Pranitha Potturi. (2023). Cloud-Driven Business Intelligence Modernization for Scalable Enterprise Data Integration and Decision Support. *International Journal of Intelligent Systems and Applications in Engineering*, 11(3s), 387 –. Retrieved from <https://www.ijisae.org/index.php/IJISAE/article/view/8505>
4. Schoenenwald, A., Kern, S., Viehhauser, J., & Schildgen, J. (2021). Collecting and visualizing data lineage of Spark jobs. *Datenbank-Spektrum*, 21, 179–189.
5. Sáenz-Adán, C., García-Izquierdo, F. J., Pérez, B., Huynh, T. D., & Moreau, L. (2022). Automated and non-intrusive provenance capture with UML2PROV. *Computing*, 104, 767–788.
6. Nadal, S., Jovanovic, P., Bilalli, B., Romero, O., & others. (2022). Operationalizing and automating Data Governance. *Journal of Big Data*, 9, 117.
7. Yamada, M., Kitagawa, H., Amagasa, T., & Matono, A. (2023). Augmented lineage: Traceability of data analysis including complex UDF processing. *The VLDB Journal*, 32, 963–983.
8. Kolla, S. K., & Reddy, V. A. R. (2023). Deep Learning Architectures For Multimodal Medical Data Integration. *South Eastern European Journal of Public Health*, 248-260.
9. Jahnke, N., & Otto, B. (2023). Data catalogs in the enterprise: Applications and integration. *Datenbank-Spektrum*, 23, 89–96.
10. Abraham, R., Schneider, J., & vom Brocke, J. (2023). A taxonomy of data governance decision domains in data marketplaces. *Electronic Markets*, 33, 22.
11. Wider, A., Verma, S., & Akhtar, A. (2023). Decentralized data governance as part of a data mesh platform: Concepts and approaches. In *2023 IEEE International Conference on Web Services*. IEEE.
12. Eichler, R., Gröger, C., Hoos, E., Stach, C., Schwarz, H., & Mitschang, B. (2023). Introducing the enterprise data marketplace: A platform for democratizing company data. *Journal of Big Data*, 10, 173.

13. Loganathan, R. (2022). Converging Security Architecture and Compliance Management in Enterprise Data Center Ecosystems: A Unified Control Framework. *International Journal of Scientific Research and Modern Technology*, 1(12), 295–312.
14. Araújo Machado, I., Costa, C., & Santos, M. Y. (2022). Data mesh: Concepts and principles of a paradigm shift in data architectures. *Procedia Computer Science*, 196, 263–271.
15. Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers*, 25, 123–141.
16. Schneider, J., Abraham, R., Meske, C., & vom Brocke, J. (2022). Artificial intelligence governance for businesses. *Business & Information Systems Engineering*, 64, 229–249.
17. Mäntymäki, M., Minkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, 2, 603–609.
18. Bharath Penaka. (2023). AI-Driven Predictive Analytics for Early Detection of High-Risk Patients in Cloud-Based Healthcare Systems . *International Journal of Medical Toxicology and Legal Medicine*, 26(3 and 4), 118–124. Retrieved from <https://ijmtlm.org/index.php/journal/article/view/1506>
19. Mökander, J., & Floridi, L. (2023). Operationalising AI governance through ethics-based auditing: An industry case study. *AI and Ethics*, 3, 451–468.
20. Birkstedt, T., Minkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: Themes, knowledge gaps and future agendas. *Internet Research*, 33(7), 133–167.
21. de Laat, P. B. (2021). Companies committed to responsible AI: From principles towards implementation and regulation? *Philosophy & Technology*, 34, 1135–1193.
22. Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The ethics of algorithms: Key problems and solutions. *AI and Society*, 37, 215–230.
23. Mökander, J., & Floridi, L. (2022). Ethics-based auditing to develop trustworthy AI. *Minds and Machines*, 32, 323–327.
24. Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2021). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26, 2141–2168.
25. Kaminski, M. E., & Malgieri, G. (2021). Algorithmic impact assessments under the GDPR: Producing multi-layered explanations. *International Data Privacy Law*, 11(2), 125–144.
26. Sasidhar Reddy Mondeddula. (2023). Natural Language Processing for Intelligent Automation of Financial Documents and Banking Operations. *International Journal of Intelligent Systems and Applications in Engineering*, 11(8s), 605–612. Retrieved from <https://www.ijisae.org/index.php/IJISAE/article/view/8527>
27. Metcalf, J., Moss, E., & Boyd, D. (2021). Owning ethics: Corporate logics, Silicon Valley, and the institutionalization of ethics. *Social Research: An International Quarterly*, 88(3), 449–472.
28. Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. E. (2022). Outsider oversight: Designing a third-party audit ecosystem for AI governance. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 560–571.



29. Waldman, A. E., & Martin, K. (2022). Governing algorithmic decisions: The role of decision importance and governance on perceived legitimacy of algorithmic decisions. *Big Data & Society*, 9(2).
30. Mattaparthi, R. (2023). Deep Learning-Driven Combustion Anomaly Detection in Diesel Powertrains: A Multi-Sensor Fusion Approach for Real-Time ECM Adaptation. *International Journal of Intelligent Systems and Applications in Engineering*, 11, 1084.
31. Eichler, R., Gröger, C., Hoos, E., Schwarz, H., & Mitschang, B. (2022). Data shopping—How an enterprise data marketplace supports data democratization in companies. In *Intelligent Information Systems: CAiSE Forum 2022 Proceedings* (pp. 19–26). Springer.