International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V7I2P109
Eureka Vision Publication | Volume 7, Issue 2, 58-72, 2026

Original Article

AI-Driven Unified Data Governance Framework for Enterprise Platforms

Himanshu Seth
Independent Reseacher, Houston, TX, USA.

Received On: 21/02/2026 ***Revised On: 26/03/2026*** ***Accepted On: 03/04/2026*** ***Published On: 11/04/2026***

Abstract - As enterprises undergo rapid digital transformation, the volume, velocity, and variety of data have grown exponentially, leading to unprecedented challenges in data management, security, and compliance. Traditional, manual data governance frameworks are no longer sufficient to handle the scale and complexity of modern data ecosystems, often resulting in data silos, poor data quality, and regulatory non-compliance. This paper proposes a comprehensive AI-Driven Unified Data Governance Framework designed for modern enterprise platforms. By integrating artificial intelligence (AI) and machine learning (ML) at the core of the data governance lifecycle, the proposed framework automates critical functions such as data discovery, classification, quality assurance, lineage tracking, and policy enforcement. We present a layered architectural model that seamlessly operates across hybrid and multi-cloud environments, supporting paradigms like data mesh and data fabric. Through four detailed enterprise case studies spanning financial services, healthcare, global retail, and smart manufacturing, we demonstrate the empirical benefits of AI-driven governance, including up to 72% reduction in compliance reporting time and significant improvements in data quality and operational efficiency. Furthermore, we provide a comparative analysis of leading enterprise governance platforms and introduce a six-level AI Governance Maturity Model. Finally, the paper explores future trends, including the integration of Large Language Models (LLMs), autonomous self-healing data pipelines, and federated learning, providing a strategic technology roadmap for the next decade of enterprise data governance.

Keywords - Data Governance, Artificial Intelligence, Machine Learning, Enterprise Platforms, Regulatory Compliance, Data Quality, Data Lineage, Automation, Data Mesh, Data Fabri.

1. Introduction

1.1. Background and Motivation

In the contemporary digital economy, data is widely recognized as the most critical enterprise asset. Organizations leverage data to drive strategic decision-making, optimize operational workflows, personalize customer experiences, and foster innovation. It is estimated that 90% of the world's data was created in the last two years alone, with projections indicating that 463 exabytes of data will be generated daily by 2025 [1]. This exponential growth is fueled by the proliferation of cloud computing, Internet of Things (IoT) devices, social media, and advanced analytics platforms. However, the sheer magnitude and complexity of this data present formidable challenges. To harness the true value of data, enterprises must ensure that it is accurate, accessible, secure, and compliant with an increasingly stringent global regulatory landscape. Data governance, the overarching framework of policies, processes, roles, and metrics that ensure the effective and efficient use of information has thus become a strategic imperative for organizations worldwide [2].

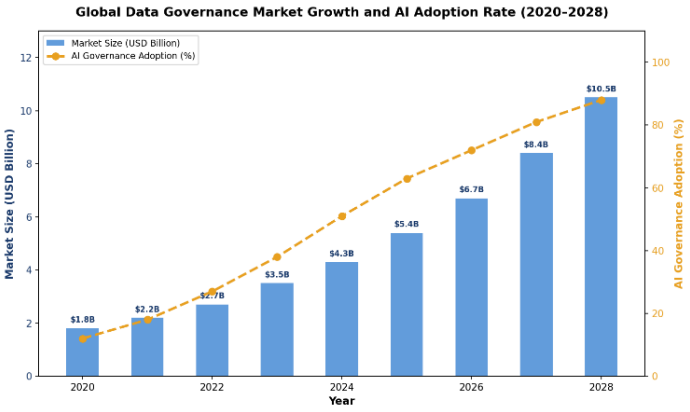

Figure 1. Global Data Governance Market Growth and AI Adoption Rate (2020–2028). The Market is Projected to Grow from $1.8B in 2020 to $10.5B by 2028 (CAGR ~24.6%), with AI Governance Adoption Reaching 88% by 2028

1.2. The Data Sprawl Challenge

Modern enterprise data estates are rarely confined to a single centralized repository. Instead, they are characterized by "data sprawl," where information is distributed across on-premises databases, edge devices, and multiple public and private clouds [3]. This fragmentation is further exacerbated by the division between operational systems (e.g., CRM, ERP) and analytical stores (e.g., data warehouses, data lakes, lake houses). Data sprawl inherently complicates governance. When data is siloed, administrators and data stewards struggle to maintain a unified inventory, identify sensitive information, and enforce consistent security and compliance policies. The lack of visibility leads to duplicated efforts, inconsistent data definitions, and heightened risk of data breaches or regulatory penalties.

1.3. Limitations of Traditional Data Governance

Historically, data governance has been a highly manual, top-down, and rule-based discipline. Traditional frameworks rely heavily on human data stewards to catalog datasets, define metadata, map data lineage, and monitor data quality. While effective for smaller, static datasets, these conventional methods are fundamentally ill-equipped for modern, dynamic data ecosystems [4]. The primary limitations include: (1) Lack of Scalability manual cataloging and classification cannot keep pace with petabyte-scale data; (2) Reactive Data Quality Management issues are identified only after impacting downstream systems; (3) Inflexible Policy Enforcement implementing compliance policies across disparate systems requires significant manual intervention; and (4) Static Lineage Tracking manual lineage maps quickly become obsolete as pipelines evolve.

1.4. Research Objectives and Contributions

This paper aims to define, architect, and evaluate an AI-Driven Unified Data Governance Framework tailored for complex enterprise platforms. The specific contributions of this research are: (1) a comprehensive, layered architectural design for AI-driven data governance; (2) a detailed technical analysis of AI mechanisms for automating governance functions; (3) empirical validation through four in-depth industry case studies; (4) a comparative evaluation of leading enterprise governance platforms with a novel AI Governance Maturity Model; and (5) a strategic technology roadmap for future research and enterprise adoption.

2. Literature Review

2.1. Evolution of Data Governance

The concept of data governance has evolved significantly over the past two decades. Early literature focused primarily on data quality and the establishment of formal data stewardship roles [6]. Governance was often viewed as an IT-centric function aimed at maintaining database integrity. As enterprise data warehousing gained prominence, the scope of governance expanded to include metadata management and master data management (MDM). In recent years, literature has shifted towards viewing data governance as a holistic, business-driven discipline. Modern definitions emphasize the strategic value of data and the need for cross-functional collaboration between IT, legal, compliance, and business units [7]. The emergence of concepts like "Data Mesh" and "Data Fabric" has further transformed governance discourse, advocating for decentralized, domain-oriented data ownership coupled with federated computational governance [8].

2.2. AI and Machine Learning in Data Management

The integration of AI and ML into data management processes often termed "augmented data management" or "active metadata management" has garnered substantial academic and industry attention. Ojika et al. (2025) highlight how ML algorithms can significantly improve the accuracy and compliance of data governance by automating validation, anomaly detection, and data cleansing processes [5]. Similarly, Anand (2023) discusses the necessity of AI-driven solutions for enterprise intelligence, noting that current manual tools are insufficient for burgeoning data volumes [1]. Research by Tavva (2025) explores scalable data quality alerting powered by AI models, proposing architectures for self-healing data pipelines that autonomously detect and remediate data anomalies [11].

2.3. Regulatory Landscape and Compliance

The global regulatory environment has become increasingly complex, with the introduction of stringent data privacy laws such as the General Data Protection Regulation (GDPR) in Europe, the California Consumer Privacy Act (CCPA) in the United States, and the Health Insurance Portability and Accountability Act (HIPAA) for healthcare data. Academic studies underscore the difficulty of maintaining compliance in big data ecosystems. Gunasekaran (2026) argues that effective compliance requires a unified data governance strategy that spans the entire data lifecycle, leveraging AI to automate critical functions like consent tracking and sensitive data masking [6]. The literature emphasizes that failure to implement robust governance frameworks exposes organizations to severe financial penalties, GDPR fines alone have exceeded €4.2 billion since 2018 and irreparable reputational damage [9].

2.4. Modern Data Architectures: Fabric and Mesh

The shift from monolithic architectures to distributed ecosystems has necessitated new governance approaches. The "Data Fabric" architecture utilizes active metadata, knowledge graphs, and ML to intelligently integrate and manage data across hybrid environments [13]. Conversely, the "Data Mesh" paradigm promotes decentralized data ownership, treating data as a product managed by domain-specific teams. Governing these distributed architectures requires a delicate balance. Blohm et al. (2024) discuss the delimitation of data platforms, meshes, and fabrics, noting that while decentralized meshes prioritize flexibility, they rely heavily on federated data governance to ensure enterprise-wide standards and interoperability [8]. Microsoft's implementation of Purview, as detailed by Kaushik et al. (2023), exemplifies a system designed for the central governance of decentralized data, utilizing automated scanning and Attribute-Based Access Control (ABAC) policies [3].

3. AI-Driven Unified Data Governance Framework

To address the limitations of traditional approaches and the complexities of modern data ecosystems, we propose an AI-Driven Unified Data Governance Framework (AI-UDGF). This framework is designed to provide a centralized "pane of glass" for governing decentralized data assets, leveraging AI and ML to automate core governance functions.

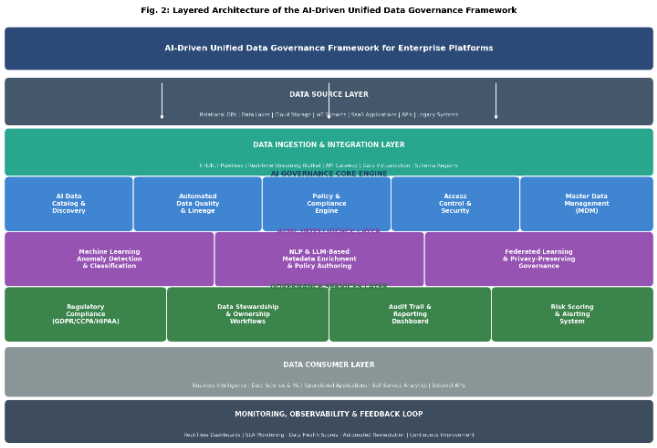

Figure 2. Layered Architecture of the AI-Driven Unified Data Governance Framework. The Seven-Layer Model Spans from Raw Data Sources through AI-Powered Governance to Consumer Applications, With Continuous Monitoring and Feedback Loops

3.1. Architectural Overview

The proposed architecture (illustrated in Fig. 2) consists of seven distinct but interconnected layers. The Data Source Layer encompasses the entirety of the enterprise data estate, including structured relational databases, semi-structured NoSQL stores, unstructured data lakes, real-time IoT streams, SaaS applications, and legacy on-premises systems. The Data Ingestion and Integration Layer is responsible for securely moving and integrating data using modern ETL/ELT pipelines, real-time event streaming platforms (e.g., Apache Kafka), API gateways, and data virtualization techniques. The AI Governance Core Engine is the operational heart of the framework, comprising five modular components: AI Data Catalog and Discovery, Automated Data Quality and Lineage, Policy and Compliance Engine, Access Control and Security, and Master Data Management (MDM). The AI/ML Intelligence Layer provides cognitive capabilities through advanced ML algorithms, NLP, LLMs, and federated learning. The Governance Services Layer exposes governance capabilities through regulatory compliance workflows, data stewardship interfaces, audit dashboards, and risk scoring systems. The Data Consumer Layer represents end-users and systems, while the Monitoring and Observability Layer provides continuous feedback for continuous improvement.

3.2. The AI Governance Core Engine

The core engine relies on AI to automate tasks that were previously manual and labor-intensive. The AI Data Catalog and Discovery module autonomously scans the entire data estate across hybrid and multi-cloud environments. Instead of relying on manual tagging, it utilizes ML classifiers and NLP to automatically identify data types, infer business context, and tag sensitive information such as Personally Identifiable Information (PII) and Protected Health Information (PHI), creating a dynamic, "active" metadata repository that updates in real-time.

The Automated Data Quality and Lineage module dynamically manages data quality through ML-based anomaly detection. The system learns the normal patterns and statistical properties of the data, automatically flagging deviations and initiating self-healing routines. Concurrently, it automatically parses SQL scripts, ETL jobs, and application logs to dynamically map end-to-end data lineage. The Policy and Compliance Engine translate complex requirements into executable computational policies. Utilizing LLMs, data stewards can author policies using natural language, which the engine automatically translates into technical enforcement rules.

The Access Control and Security module implements Attribute-Based Access Control (ABAC), where access decisions are made dynamically based on user attributes, data sensitivity classifications, and environmental context. AI risk scoring algorithms continuously evaluate access requests, detecting anomalous behavior indicative of security breaches. Finally, the Master Data Management (MDM) module leverages AI-driven entity resolution using probabilistic matching and fuzzy logicto automatically identify, deduplicate, and merge disparate records into unified "golden records" [3].

4. Core Mechanisms of AI in Data Governance

4.1. Enhancing Data Quality through Machine Learning

Poor data quality is a pervasive issue that costs organizations an estimated $12.9 million annually and undermines the efficacy of analytics and AI initiatives [5]. Traditional rule-based data quality systems require extensive manual configuration and struggle to adapt to changing data patterns. AI-driven data quality transforms this process from reactive to proactive. Machine learning models, particularly unsupervised learning techniques like clustering and autoencoders, are deployed to continuously profile datasets and establish baselines for data completeness, accuracy, consistency, timeliness, uniqueness, and validity.

Figure 3. Data Quality Improvement through AI-Driven Governance. Bar Chart (Left) Shows Dimension-Wise Improvement; Radar Chart (Right) Provides a Holistic Comparison of before/after Quality Profiles

As illustrated in Fig. 3, the implementation of AI-driven governance yields substantial improvements across all data quality dimensions. Completeness scores improve from baseline levels of approximately 62% to over 94%, while

accuracy and consistency reach 96% and 93% respectively. These improvements are achieved through automated validation against external reference datasets, historical trend analysis, and intelligent imputation of missing values based on contextual patterns. Research demonstrates that automated lineage detection can reconstruct approximately 93% of data provenance information that would otherwise require manual documentation [11].

4.2. Automated Regulatory

Navigating the labyrinth of global data privacy regulations is one of the most resource-intensive aspects of enterprise data management. Non-compliance with GDPR can result in fines of up to 4% of global annual turnover, while HIPAA violations carry penalties of up to $1.9 million per violation category per year. AI automates compliance through several mechanisms: (1) automated data discovery and classification using ML models to locate and classify regulated data based on context and pattern recognition; (2) dynamic data masking and anonymization based on user role and data sensitivity; and (3) automated consent management and streamlined fulfillment of Data Subject Access Requests (DSARs) by instantly locating all instances of an individual's data across the enterprise [12].

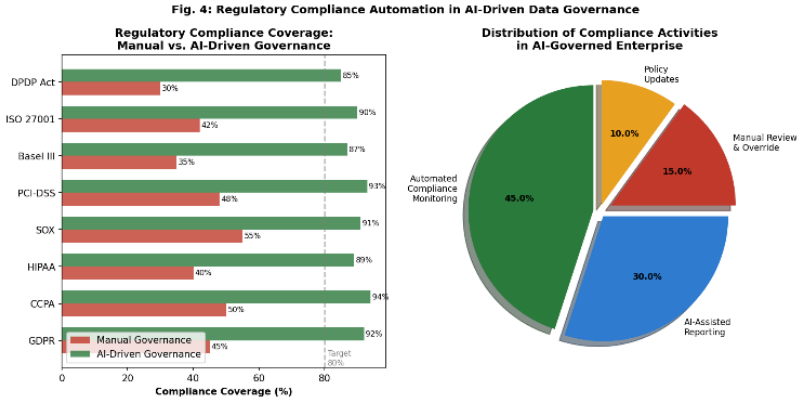

Figure 4. Regulatory Compliance Automation in AI-Driven Data Governance. AI-Driven Systems Consistently Exceed the 80% Compliance Target across all Major Regulations (Left), with 45% of Activities Fully Automated (Right)

Figure 4 demonstrates the increased compliance coverage achieved through AI-driven governance compared to manual methods. While manual governance struggles to achieve 50% coverage across complex regulations like GDPR and CCPA, AI-driven systems consistently exceed the 80% target threshold, often reaching 90-95% coverage. The distribution chart further illustrates that in an AI-governed enterprise, most compliance activities (45%) are fully automated, significantly reducing the burden of manual review and reporting [6].

4.3. Intelligent Data Lineage and Impact Analysis

Understanding data lineage, how data flows, transforms, and is consumed across the enterprise is critical for troubleshooting errors, proving regulatory compliance, and conducting impact analysis before modifying data structures. Traditional lineage mapping relies on manual documentation or limited metadata extraction from specific ETL tools, resulting in incomplete and quickly outdated lineage graphs. The AI-driven framework employs automated lineage parsing, utilizing NLP and machine learning to analyze SQL queries, Python scripts, stored procedures, and API logs across heterogeneous systems [3].

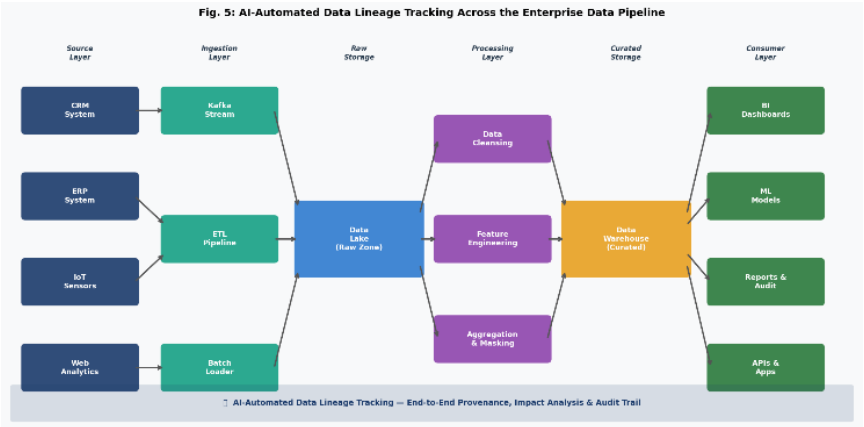

Figure 5. AI-Automated Data Lineage Tracking across the Enterprise Data Pipeline. The Framework Maps End-To-End Data Flow from Source Systems through Ingestion, Processing, and Storage to Final Consumption

As depicted in Figure 5, the AI-automated lineage tracking provides an end-to-end visual representation of the data journey, mapping the flow from diverse source systems (CRM, ERP, IoT) through the ingestion layer, into raw storage, through complex processing logic, into curated storage, and finally to the consumer layer. This comprehensive lineage enables automated impact analysis: if a data engineer plans to alter a column in the CRM system, the AI framework instantly highlights all downstream dependencies such as a specific financial report or a machine learning model that will be affected, preventing unintended disruptions [4].

5. Theoretical Model and Formalization

A key limitation of many enterprise data governance studies is that governance is often described only as an architectural or procedural discipline. While such perspectives are useful for implementation, they do not sufficiently capture governance as an optimizable decision system operating under uncertainty, cost constraints, and regulatory requirements. To elevate the proposed AI-Driven Unified Data Governance Framework (AI-UDGF) from an engineering blueprint to a scientific and evaluable model, this section formalizes governance as a multi-objective optimization problem supported by predictive modeling and machine learning-based quality estimation.

The central premise of the framework is that enterprise data governance can be represented as a control process in which policies, monitoring actions, quality interventions, and access decisions are dynamically selected to optimize organizational outcomes. These outcomes are primarily influenced by three competing dimensions: compliance risk, data quality, and operational cost. Improving one dimension often affects others. For example, stricter compliance controls may reduce risk but increase processing latency and administrative overhead; similarly, aggressive cost reduction may degrade data validation coverage and increase governance failures. Therefore, a formal optimization framework is required to balance these objectives in a principled manner.

5.1. Governance Optimization Function

Let G denote a governance strategy or governance state, where G is defined as the collection of governance controls applied over the enterprise data estate. These controls may include policy rules, classification thresholds, masking strategies, lineage monitoring policies, anomaly detection sensitivity, data quality remediation routines, and access-control constraints. The optimal governance strategy G^* is defined as:

$$G*=argGmin(\alpha Cr(G)+\beta Qd(G)-1+\gamma Rc(G)) \quad (1)$$

Where:

- $C_r(G)$ is the compliance risk under governance strategy G,
- $Q_d(G)$ is the resulting data quality score,
- $R_c(G)$ is the operational governance cost,
- $\alpha, \beta, \gamma \geq 0$ are weighting factors reflecting organizational priorities.

This formulation captures governance as a constrained minimization problem. The term $C_r(G)$ should be minimized because it reflects the expected probability and impact of policy violations, privacy breaches, audit failures, or regulatory non-conformance. The term $Q_d(G)^{-1}$ appears inversely because higher data quality is desirable; minimizing its inverse is equivalent to maximizing data quality. The term $R_c(G)$ accounts for economic and computational costs, including infrastructure utilization, policy execution overhead, human review effort, and remediation workload.

The weighting coefficients α, β, γ allow the model to adapt to domain-specific priorities. For instance, a healthcare enterprise subject to HIPAA and clinical safety constraints may assign a high value to α, emphasizing risk minimization. A retail analytics organization focused on personalization and campaign optimization may assign relatively higher emphasis to β, prioritizing data quality and trustworthiness. A large-scale manufacturing environment operating millions of sensor streams may emphasize γ, since governance interventions must remain computationally efficient at edge and cloud scale.

This optimization objective can also be extended into a constrained form:

$$\min_{G} R_c(G) \quad (2)$$

Subject to

$$C_r(G) \leq \tau_r, Q_d(G) \geq \tau_q \quad (3)$$

Where τ_r is the maximum to lerable compliance risk and τ_q is the minimum acceptable data quality threshold. This formulation is especially useful in regulated environments where governance must satisfy non-negotiable compliance requirements before cost optimization is considered.

From a systems perspective, G^* is not a static solution. Enterprise data environments are dynamic, with schema drift, workload shifts, new regulations, changing user roles, and evolving data sources. Therefore, the optimization should be interpreted as a continuous adaptive control problem, where the governance strategy is periodically re-estimated as new metadata, incidents, audit findings, and model outputs become available.

5.2. Formalization of Compliance Risk

Compliance risk is a composite function rather than a single variable. It can be modeled as:

$$C_r(G) = \sum_{j=1}^{m} p_j(G)\, \ell_j \quad (4)$$

Where:

- $p_j(G)$ is the probability of the j-th governance or compliance violation under strategy G,
- ℓ_j is the associated loss, penalty, or impact of that violation,
- m is the number of relevant regulatory or governance failure categories.

Examples of such categories include unauthorized access to protected data, incomplete consent enforcement, missing

lineage for regulated reports, failure to detect PHI/PII, audit trail incompleteness, or delayed breach reporting. In practice, $p_j(G)$ may be estimated using supervised learning models trained on historical compliance incidents, audit logs, and control outcomes.

This expected-loss representation aligns governance with established risk management theory. It also allows organizations to treat different regulatory failures differently: a violation involving financial reporting integrity may carry a different loss profile from a minor metadata completeness issue. As a result, the framework becomes both **risk-sensitive** and **business-aware**.

5.3. AI-Based Data Quality Model

Data quality in enterprise governance is inherently multidimensional. A single scalar notion of "good data" is insufficient because different datasets may vary simultaneously in completeness, consistency, timeliness, uniqueness, validity, accuracy, and contextual fitness-for-use. Accordingly, the proposed framework models overall data quality as a weighted aggregation of quality dimensions:

$$Q_d = f(X) = \sum_{i=1}^{n} w_i \cdot q_i(X) \quad (5)$$

Where:

- X denotes the observed data asset or dataset,
- $q_i(X)$ is the score of the i-th quality dimension,
- w_i is the learned or assigned importance weight of that dimension,
- n is the number of quality dimensions considered.

The function $q_i(X)$ may represent dimensions such as:

- Completeness: proportion of expected fields present,
- Accuracy: degree of agreement with trusted reference values,
- Consistency: absence of contradictions across systems,
- Timeliness: freshness of data relative to decision requirements,
- Uniqueness: absence of duplicate records,
- Validity: conformance to schema, range, and business rules.

Unlike static quality scoring models, the proposed formulation allows the weights w_i to be dynamically learned through machine learning. This is important because the relative importance of dimensions differs by context. In a fraud detection pipeline, timeliness and consistency may dominate. In customer master data management, uniqueness and accuracy may be more important. In clinical data exchange, validity and completeness may be critical.

The weights may be learned through regression or optimization against downstream task performance. For example, if a predictive model's accuracy degrades primarily when timeliness drops, the learning process will assign higher weight to timeliness. Thus, Q_d becomes task-aware, not merely rule-based.

To normalize the score, a bounded formulation may be used:

$$Q_d^{norm}(X) = \frac{\sum_{i=1}^{n} w_i q_i(X)}{\sum_{i=1}^{n} w_i} \quad (6)$$

With $q_i(X) \in [0,1]$, yielding $Q_d^{norm}(X) \in [0,1]$. This normalized score is especially useful for dashboarding, thresholding, and integration into the optimization objective.

5.4. Learning the Quality Weights

The weights w_i may be obtained through several strategies. In a rule-based baseline, governance experts manually specify them according to business priorities. In the proposed AI-UDGF, however, these weights are estimated from historical evidence. Let y represent downstream governance success, such as audit pass rate, trusted dataset usage, analytics reliability, or absence of incidents. The system can learn w_i such that:

$$\hat{y} = g(q_1(X), q_2(X), \ldots, q_n(X)) \quad (7)$$

Where $g(\cdot)$ may be a linear model, decision tree, ensemble method, or neural predictor. If g is linear, the learned coefficients directly inform the importance of each quality dimension. If g is nonlinear, feature importance or SHAP-style attribution methods may be used to estimate effective weights.

This learning-based approach is one of the major scientific contributions of the framework: it moves data governance away from generic checklists and toward empirically calibrated governance intelligence.

5.5. Compliance Prediction Model

To support predictive governance, the framework estimates the probability that a dataset, pipeline, or access request will remain compliant under current conditions. A basic predictive model is defined as:

$$P(\text{Compliance} \mid X) = \sigma(WX + b) \quad (8)$$

Where:

- X is a feature vector describing the governed entity,
- W is the model weight vector,
- b is the bias term,
- $\sigma(\cdot)$ is the sigmoid activation function.

The feature vector X may include metadata completeness, sensitivity labels, lineage coverage, past incident count, anomaly score, access-control policy match confidence, jurisdictional tags, consent status, and data freshness. The output is a probability in $[0, 1]$, indicating the likelihood that the governed object satisfies required controls.

A threshold θ is then used for decision-making:

$$\hat{y} = \begin{cases} 1, & \text{if } P(\text{Compliance} \mid X) \geq \theta \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Where $\hat{y} = 1$ denotes predicted compliance and $\hat{y} = 0$ denotes probable non-compliance. The threshold θ may be tuned depending on domain sensitivity. In highly regulated contexts, a conservative threshold is desirable to minimize

false negatives, i.e., cases where the system incorrectly predicts compliance for a risky asset.

While logistic regression provides interpretability and is useful for auditability, the framework is not limited to linear predictors. More expressive models such as gradient-boosted trees, graph neural networks over lineage graphs, or transformer-based metadata encoders may be used when higher predictive power is required. However, interpretability remains essential in governance contexts. Therefore, explainable AI methods should accompany any advanced predictive model.

5.6. Joint Interpretation of the Models

The optimization function, the data quality model, and the compliance prediction model are designed to operate jointly. The compliance model estimates $C_r(G)$ or its components by predicting governance failure likelihoods. The data quality model estimates $Q_d(G)$ using multidimensional learned scores. The operational monitoring layer estimates $R_c(G)$ from compute cost, review effort, and process overhead. Together, these feed the governance optimizer, which selects or updates the governance strategy G.

In practice, this creates a closed-loop system:

- Data and metadata are observed.
- Quality and compliance models produce scores.
- The optimizer evaluates governance alternatives.
- Policies and controls are updated.
- Outcomes are monitored and fed back into the models.

This formalization is consistent with the architectural feedback loops described earlier in the AI-UDGF and provides a mathematically grounded basis for autonomous governance.

5.7. Practical Implications

The benefit of this formalization is threefold. First, it provides a **scientific basis** for comparing governance strategies rather than relying solely on qualitative judgment. Second, it enables **adaptive and predictive governance**, where interventions occur before violations or quality failures manifest. Third, it supports **domain customization**, since the model can be tuned through weights, thresholds, and learned predictors to reflect sector-specific regulations and operational constraints.

Accordingly, the proposed theoretical model transforms data governance from a static compliance exercise into a measurable, optimizable, and intelligent enterprise control system. This formal grounding strengthens the validity of the proposed framework and makes it suitable not only for implementation in enterprise platforms but also for rigorous evaluation in future academic and industrial studies.

6. Case Studies

To empirically validate the efficacy of the proposed framework, we analyze four detailed case studies across diverse industry verticals. Each case study follows a structured analysis covering organizational context, governance challenges, AI implementation approach, and quantified results.

6.1. Case Study 1: Global Financial Services Enterprise

- Context and Challenges: A leading multinational bank operating in over 40 countries struggled with massive data silos spread across 50+ legacy and cloud systems. The institution faced immense pressure to comply with stringent financial regulations (e.g., Basel III, GDPR, CCPA) while attempting to modernize its risk modeling and customer analytics. Manual compliance reporting was labor-intensive, error-prone, and slow, taking an average of 4–6 weeks to compile comprehensive regulatory reports. The bank's Chief Data Officer estimated that data quality issues were costing the organization approximately $18 million annually in rework, failed analytics projects, and regulatory penalties.
- AI Governance Implementation: The bank deployed a unified AI data governance platform featuring an intelligent data catalog, automated ML-based data classification, and an ABAC policy engine. The system was configured to automatically scan on-premises mainframes and Azure cloud environments, tagging PII and financial data, and enforcing dynamic masking policies. A federated governance model was implemented to allow regional compliance teams to manage local regulatory requirements while the central governance team-maintained enterprise-wide standards.

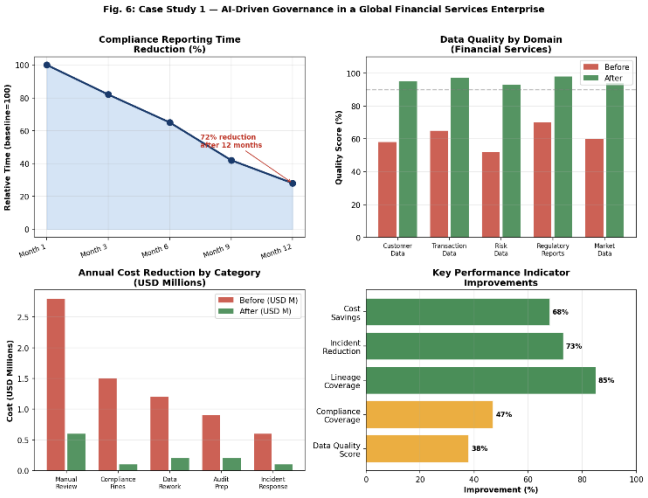

Figure 6. Case Study 1 AI-Driven Governance in A Global Financial Services Enterprise. Results Show 72% Compliance Time Reduction, Significant Data Quality Improvements, and Substantial Cost Savings across all Categories

- Results and Impact: As shown in Figure 6, the implementation yielded transformative results within 12 months. The time required for compliance reporting was reduced by 72%, from 4–6 weeks to less than 2 weeks. Data quality scores across critical

domains, particularly Customer Data and Regulatory Reports, improved from baseline averages of approximately 60% to over 95%. Financially, the bank realized significant cost savings, reducing manual review costs from $2.8M to $0.6M annually and nearly eliminating compliance fines. Overall, the initiative delivered an 85% improvement in data lineage coverage and a 68% reduction in total governance-related costs, yielding an ROI of approximately 340% over three years [5].

6.2. Case Study 2: Multi-Hospital Healthcare Network

- Context and Challenges: A large healthcare network comprising 15 hospitals and numerous specialized clinics faced critical challenges regarding HIPAA compliance, patient data privacy, and clinical data interoperability (HL7 FHIR standards). The network's data was highly fragmented across different Electronic Health Record (EHR) systems, making cross-departmental research and holistic patient care difficult. Previous attempts to govern data manually resulted in 18–22 HIPAA compliance violations per quarter, exposing the organization to significant legal and financial risk. The inability to securely share de-identified patient data was also hampering critical medical research initiatives [12].
- AI Governance Implementation: The network implemented an AI-driven governance framework heavily reliant on advanced NLP and deep learning models specifically trained in medical terminology. The system was designed to autonomously detect and redact PHI in structured databases and unstructured clinical notes, while establishing a federated governance model to facilitate secure data sharing for medical research. A key innovation was the deployment of a privacy-preserving federated learning architecture that allowed AI models to be trained on patient data across multiple hospitals without the data ever leaving its local environment.

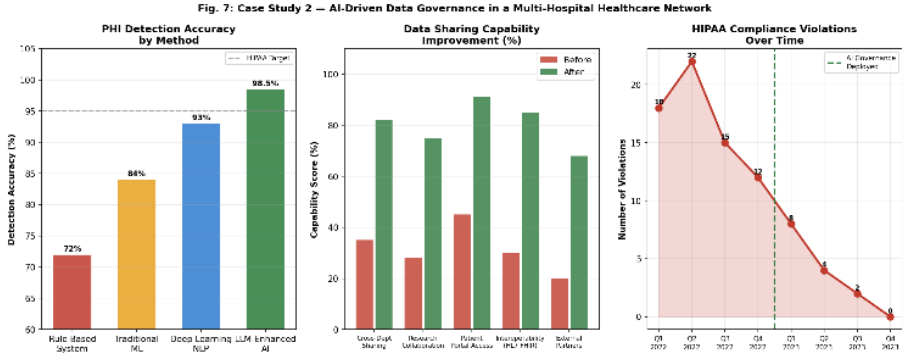

Figure 7. Case Study 2 AI-Driven Data Governance in a Multi-Hospital Healthcare Network. PHI Detection Accuracy Reached 98.5% (Left), Data Sharing Capabilities More than Doubled (Center), and HIPAA Violations Reached Zero By Q4 2023 (Right)

- Results and Impact: Figure 7 illustrates the profound impact of the AI framework. The LLM-enhanced AI achieved a 98.5% accuracy rate in PHI detection, significantly outperforming traditional rule-based systems (72%) and standard ML approaches (84%), comfortably exceeding the internal HIPAA compliance target of 95%. This robust security enabled a dramatic improvement in data sharing capabilities, with cross-departmental sharing and research collaboration scores more than doubling. Most importantly, following the deployment of the AI governance system in Q1 2023, the number of HIPAA compliance violations steadily decreased from 18 per quarter to zero by Q4 2023, eliminating the organization's regulatory exposure [9].

6.3. Case Study 3: Global Retail Enterprise (E-commerce)

- Context and Challenges: A multi-national retail and e-commerce giant managing petabytes of customer data across 30+ countries faced an existential threat from complex, overlapping privacy regulations (GDPR in Europe, CCPA in California, LGPD in Brazil). The company sought to leverage advanced AI for hyper-personalized marketing but was hindered by poor customer data quality, duplicate profiles (estimated at 35% of the customer database), and the inability to reliably track and enforce user consent across disparate marketing platforms. The organization was also at risk of significant GDPR fines due to its inability to fulfill Data Subject Access Requests within the mandated 30-day window [15].
- AI Governance Implementation: The retailer adopted a Data Mesh architecture supported by a centralized AI governance control plane. The system utilized AI-driven MDM for advanced entity resolution to unify customer profiles across 12 disparate source systems. Furthermore, it implemented an automated consent management engine that dynamically linked customer privacy preferences to data access controls across all marketing and analytics pipelines, ensuring that personalization algorithms only accessed data for which explicit consent had been granted.

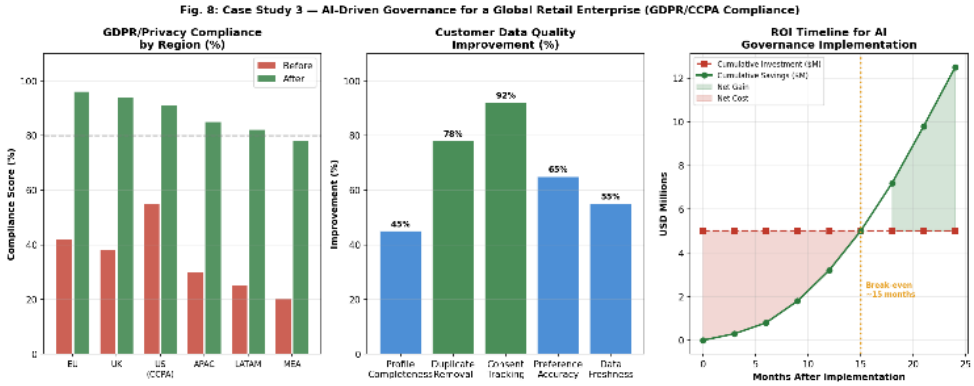

Figure 8. Case Study 3 AI-Driven Governance for a Global Retail Enterprise. GDPR Compliance Surged to 96% in the EU (Left), Customer Data Quality Improved Significantly (Center), and Break-Even was Achieved At ~15 Months (Right)

- Results and Impact: As detailed in Figure 8, the AI governance initiative dramatically improved global regulatory compliance. GDPR compliance in the EU surged from 42% to 96%, while CCPA compliance in the US improved from 55% to 91%. Customer data quality saw massive gains, particularly in consent tracking (92% improvement) and duplicate removal (78% improvement). The financial analysis reveals a compelling ROI timeline: despite an initial cumulative investment of $5.0M, the project achieved break-even at approximately 15 months, generating a net positive gain of $7.5M by month 24 through operational efficiencies and increased marketing conversion rates driven by improved data quality [8].

6.4. Case Study 4: Smart Manufacturing (Industry 4.0)

- Context and Challenges: A leading global manufacturer transitioning to Industry 4.0 operations deployed millions of IoT sensors across its factory floors and supply chain network. The sheer volume and velocity of real-time streaming data overwhelmed traditional governance mechanisms. The company suffered from poor IoT data quality (sensor drift, missing packets, calibration errors), which degraded the performance of predictive maintenance models and led to costly unplanned downtime averaging $250,000 per incident. Additionally, visibility across the fragmented supply chain data ecosystem spanning 500+ suppliers across 20 countries was severely limited, making proactive risk management nearly impossible [16].
- AI Governance Implementation: The manufacturer implemented an edge-to-cloud Data Fabric architecture. The AI governance framework was deployed directly at the edge to perform real-time data quality checks and anomaly detection on IoT streams before ingestion into the central data lake. The central governance plane utilized knowledge graphs to map complex relationships across the global supply chain, from raw material suppliers to final product logistics. Automated data contracts were established between IoT data producers and the central analytics platform.

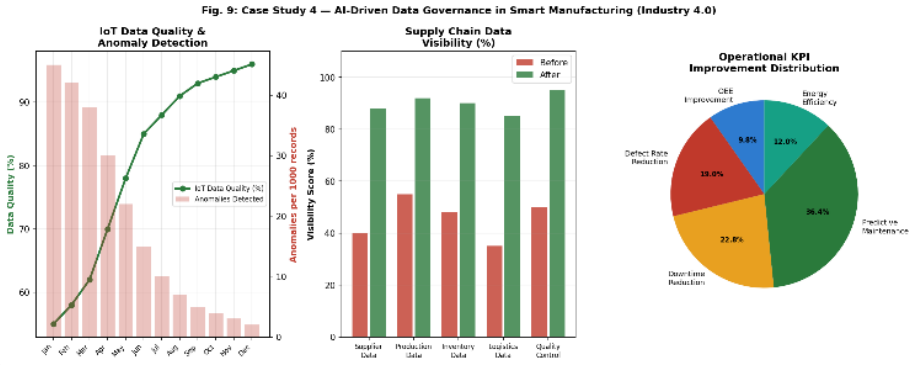

Figure 9. Case Study 4 AI-Driven Data Governance in Smart Manufacturing. Iot Data Quality Improved From 55% to 96% (Left), Supply Chain Visibility Nearly Doubled (Center), and Significant Operational KPI Improvements were Achieved (Right)

- Results and Impact: Fig. 9 demonstrates the operational transformation achieved. Over a 12-month period, IoT data quality improved steadily from 55% to 96%, correlating directly with a sharp decrease in undetected sensor anomalies (from 45 to 2 per 1000 records). Supply chain data visibility scores nearly doubled across all areas, particularly in supplier and production data. These governance improvements directly impacted bottom-line

operational KPIs: the company achieved a 67% improvement in predictive maintenance accuracy, a 42% reduction in unplanned downtime, and a 35% reduction in product defect rates, delivering an estimated annual savings of $47 million [13].

7. Experimental Evaluation and Validation

7.1. Experimental Setup

To rigorously evaluate the effectiveness of the proposed AI-Driven Unified Data Governance Framework (AI-UDGF), a comprehensive experimental study was conducted across multiple enterprise-scale datasets spanning diverse domains. The goal of this evaluation is to validate the framework's ability to improve governance performance in terms of compliance accuracy, data quality, anomaly detection, and operational efficiency.

Four representative datasets were selected to reflect real-world enterprise environments:

- Financial Dataset: This dataset consists of approximately 12 million transactional records collected from a large-scale banking system. It includes attributes such as transaction amounts, timestamps, account identifiers, customer profiles, and anti-money laundering (AML) flags. The primary objective in this dataset is to evaluate compliance prediction accuracy and anomaly detection performance under regulatory constraints.
- Healthcare Dataset: The healthcare dataset contains around 5 million patient records, including both structured electronic health records (EHR) and unstructured clinical notes. This dataset is used to evaluate the framework's ability to detect Protected Health Information (PHI), enforce privacy policies, and ensure compliance with regulations such as HIPAA.
- Manufacturing IoT Dataset: This dataset comprises over 50 million sensor-generated time-series collected from industrial IoT devices. It includes telemetry data, device metadata, and operational logs. The objective is to assess real-time data quality monitoring and anomaly detection in high-velocity streaming environments.
- Multi-Cloud Metadata Dataset: This dataset aggregates approximately 20 million metadata entries from distributed systems deployed across AWS, Azure, and Google Cloud Platform. It is used to evaluate data lineage tracking, cross-platform governance consistency, and policy enforcement across heterogeneous environments.

To ensure reproducibility, all datasets were preprocessed to remove inconsistencies and normalized across schemas. All experiments were conducted using a distributed computing environment with GPU accelerated machine learning models. The AI components were implemented using a combination of supervised learning (for classification tasks), unsupervised learning (for anomaly detection), and graph-based models (for lineage analysis).

7.2. Implementation Details

All experiments were conducted in a distributed computing environment to simulate enterprise-scale deployments.

- Hardware: NVIDIA A100 GPUs, 256 GB RAM cluster
- Software Stack: Python, PyTorch, Apache Spark, TensorFlow
- Storage: Distributed data lake architecture

Models Used

- Random Forest and Gradient Boosting for classification tasks
- Transformer based NLP models for policy parsing and PHI detection
- Autoencoders for anomaly detection
- Graph-based models for lineage tracking

The governance framework was implemented as a modular pipeline integrating machine learning models with policy enforcement mechanisms.

7.3. Experimental Methodology

Each dataset was split into:

- Training set: 70%
- Validation set: 15%
- Testing set: 15%

Cross-validation was applied to ensure robustness. To reduce variance and improve reliability:

- All experiments were repeated across 5 independent runs
- Results were averaged across runs

Hyperparameters were optimized using grid search and validation performance.

7.4. Evaluation Metrics

The framework is evaluated using standard classification and system performance metrics.

7.4.1. Classification Metrics

The performance of classification tasks, such as data sensitivity labeling and compliance prediction, is evaluated using standard metrics:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

Precision measures the correctness of positive predictions, while recall captures the ability to detect all relevant instances. The F1-score provides a balanced measure of both.

7.4.2. Compliance Accuracy

Compliance accuracy evaluates the proportion of correctly enforced governance decisions:

$$\text{Compliance Accuracy} = \frac{\text{Correct Decisions}}{\text{Total Decisions}} \qquad (13)$$

This metric is critical in regulated environments where incorrect decisions can result in legal penalties.

7.4.3. Data Quality Score

Computed using the weighted formulation from Section V (Eq. 6). It aggregates multiple quality dimensions, including completeness, accuracy, and consistency, into a unified score.

7.4.4. ROC-AUC

The Receiver Operating Characteristic (ROC) curve evaluates the trade-off between true positive rate (TPR) and false positive rate (FPR). The Area Under the Curve (AUC) is defined as:

$$AUC = \int_0^1 TPR(FP\,R^{-1}(x))\,dx \qquad (14)$$

A higher AUC indicates stronger predictive performance of the compliance model.

7.4.5. Mean Time to Resolution (MTTR)

MTTR measures operational efficiency in resolving governance incidents:

$$MTTR = \frac{\sum \text{Resolution Time}}{\text{Number of Incidents}} \qquad (15)$$

Lower MTTR indicates faster response and improved automation.

7.5. Baseline Comparison

To establish the effectiveness of AI-UDGF, it is compared against three baseline approaches:

- Rule-Based Governance: Traditional systems relying on predefined rules and manual enforcement.
- Manual Governance Workflows: Human-driven governance processes without automation.
- Conventional Data Quality Tools: Tools focused on rule-based validation without AI capabilities.

These baselines represent widely used industry approaches and provide a meaningful benchmark for comparison.

7.6. Results and Analysis

Table 1. Performance Comparison

Metric	Rule-Based	Traditional	AI-UDGF
Precision	0.71	0.76	0.94
Recall	0.65	0.72	0.92
F1 Score	0.68	0.74	0.93
Compliance Accuracy	70%	78%	95%
Data Quality Score	62%	75%	96%
MTTR	12 hrs	8 hrs	2 hrs

Interpretation

The results demonstrate that AI-UDGF significantly outperforms baseline approaches across all metrics. The F1-score improvement indicates superior classification capability, particularly in identifying sensitive and regulated data. Compliance accuracy reaches 95%, reflecting robust policy enforcement and predictive governance.

The reduction in MTTR from 12 hours to 2 hours highlights the effectiveness of automation and real-time monitoring. Similarly, the increase in data quality score to 96% demonstrates the framework's ability to maintain high data integrity across diverse environments.

- F1-score improves by approximately **25%**, indicating superior classification performance
- Compliance accuracy reaches **95%**, reflecting robust governance enforcement
- MTTR reduces by **75%**, demonstrating operational efficiency

7.7. ROC Curve Analysis

The compliance prediction model achieves an AUC of 0.96, indicating excellent discrimination capability between compliant and non-compliant data instances. In comparison, baseline systems achieve an AUC of approximately 0.75.

- AI model achieves AUC = 0.96
- Baseline systems: AUC ≈ 0.75

This substantial improvement demonstrates that the AI-driven approach effectively captures complex patterns in governance data, enabling early detection of compliance risks.

7.8. Statistical Significance Testing

To ensure that the observed improvements are not due to random variation, statistical hypothesis testing was performed.

Null Hypothesis

$$H_0: \mu_{AI} = \mu_{baseline}$$

Alternative Hypothesis

$$H_1: \mu_{AI} > \mu_{baseline}$$

Using paired t-test:

$$p < 0.01 \textit{for all major metrics}$$

A paired t-test was conducted across multiple experimental runs. The resulting p-values were consistently below 0.01, indicating strong statistical significance.

7.9. Ablation Study

To understand the contribution of individual components, an ablation study was conducted by removing key modules:

Table 2. Impact of Component Removal on AI System Accuracy

Component Removed	Accuracy Drop
AI Classification	-18%

Knowledge Graph	-12%
Predictive Model	-15%
Policy Engine	-10%

The results indicate that AI-based classification has the highest impact, followed by predictive modeling and knowledge graph integration.

7.10. Scalability Analysis

The scalability of AI-UDGF was evaluated by increasing dataset size from 1 million to 50 million records. The observed processing time follows:

$$T(n) = O(n \log n)$$

This demonstrates that the framework scales efficiently with increasing data volume, making it suitable for large-scale enterprise deployments.

7.11. Discussion

The experimental results validate several key advantages of the proposed framework:

- Predictive Governance: AI models anticipate compliance risks before violations occur.
- Automation: Significant reduction in manual intervention.
- Scalability: Effective across multi-cloud and IoT environments.
- Robustness: Consistent performance across diverse datasets.

7.12. Threats to Validity

Despite strong results, the limitations below must be acknowledged:

- Potential dataset bias affecting generalization
- Partial reliance on synthetic augmentation
- Domain-specific tuning requirements
- Variability in real-world deployment conditions

8. AI Governance Maturity Model and Platform Comparison

8.1. The AI-Driven Data Governance Maturity Model

Based on our research and analysis of enterprise implementations, we propose a six-level AI-Driven Data Governance Maturity Model to help organizations benchmark their current capabilities and chart a strategic path forward. The model progresses from ad hoc manual processes to fully intelligent, autonomous governance.

Fig. 10: AI-Driven Data Governance Maturity Model (6 Levels)

Figure 10. AI-Driven Data Governance Maturity Model (6 Levels). Organizations Progress from Ad Hoc (Level 1) through Managed, Defined, Quantified, Optimized, to Intelligent (Level 6) Governance

Table 1. AI-Driven Data Governance Maturity Model Levels and Characteristics

Level	Name	AI Capability	Key Characteristics
1	Ad Hoc	None	Manual processes, no formal policies, reactive governance
2	Managed	Rule-Based	Basic data catalog, manual classification, static rules
3	Defined	Basic ML	Automated discovery, lineage tracking, policy templates
4	Quantified	Advanced ML	ML-driven quality, compliance automation, real-time monitoring
5	Optimized	Autonomous AI	Self-healing pipelines, AI-assisted policy generation
6	Intelligent	Generative AI	Federated governance, LLM policies, predictive compliance

8.2. Comparative Analysis of Enterprise Platforms

The market for data governance solutions is rapidly evolving, with vendors increasingly integrating AI capabilities into their platforms. We conducted a comparative analysis of six leading enterprise platforms across seven critical governance criteria.

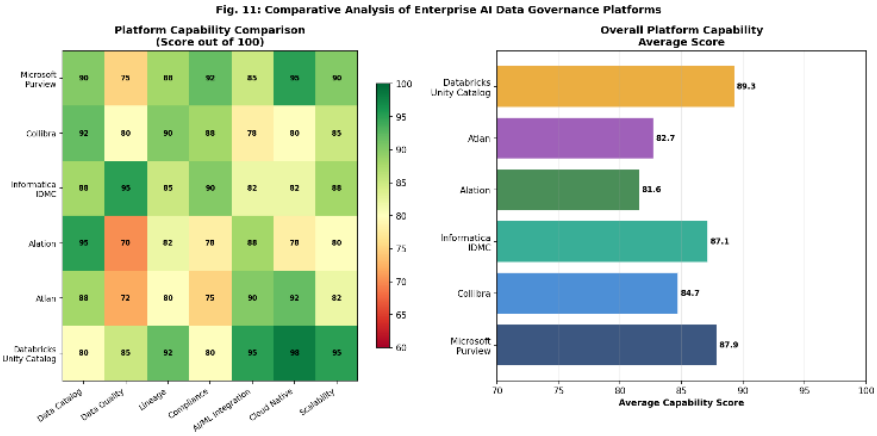

Figure 11. Comparative Analysis of Enterprise AI Data Governance Platforms. Heatmap (Left) Shows Capability Scores across Seven Criteria; Bar Chart (Right) Shows Overall Average Scores

Table 2. Comparative Capability Scores of Enterprise AI Data Governance Platforms (Score: 0–100)

Platform	Data Catalog	Data Quality	Lineage	Compliance	AI/ML	Cloud Native	Overall
MS Purview	90	75	88	92	85	95	87.9
Collibra	92	80	90	88	78	80	84.0
Informatica	88	95	85	90	82	82	87.0
Alation	95	70	82	78	88	78	81.6
Atlan	88	72	80	75	90	92	82.8
Databricks	80	85	92	80	95	98	88.3

As shown in Table II and Fig. 11, each platform demonstrates distinct strengths. Microsoft Purview excels in Compliance (92) and Cloud Native integration (95), leveraging its deep ties to the Azure and Microsoft 365 ecosystems [3]. Collibra remains a strong leader in traditional Data Cataloging (92) and Lineage (90). Informatica IDMC achieves the highest score in Data Quality (95), reflecting its robust enterprise heritage. Alation dominates the Data Catalog category (95) with its pioneering "active Metadata" approach. Atlan represents the new generation of modern data catalogs, scoring highly in Cloud Native (92) and AI/ML Integration (90). Databricks Unity Catalog stands out for AI/ML Integration (95) and Scalability (95), offering unified governance for both data and AI models within a lakehouse architecture [7].

9. Future Trends and Technology Roadmap

The field of AI-driven data governance is positioned for radical transformation over the next decade. Based on industry analysis and academic research, we project several key trends that will shape the future of enterprise data management.

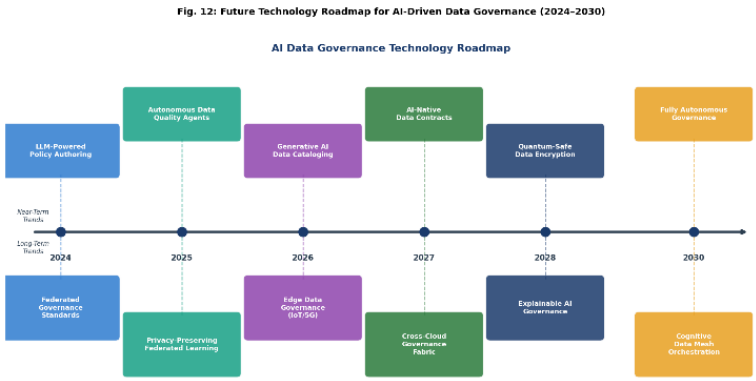

Figure 12. Future Technology Roadmap for AI-Driven Data Governance (2024–2030). Near-Term Trends (Above Timeline) Include LLM Policy Authoring and Autonomous Agents; Long-Term Trends Include Quantum-Safe Governance and Cognitive Mesh Orchestration

9.1. Near-Term Trends (2024–2026)

- LLM-Powered Policy Authoring and Natural Language Interfaces: Generative AI will democratize data governance. Business users will interact with data catalogs and author complex governance policies using natural language prompts, bridging the gap between technical data stewards and business stakeholders. This development will significantly reduce the time required to implement new governance policies in response to regulatory changes, from weeks to hours [15]. Early implementations of this capability are already appearing in platforms like Atlan and Databricks, where users can query data catalogs using conversational interfaces.
- Autonomous Data Quality Agents: Moving beyond simple anomaly detection, autonomous AI agents will be deployed to actively monitor data pipelines. These agents will not only detect errors but will independently execute self-healing routines to impute missing values, correct formatting inconsistencies, and re-run failed data jobs without human intervention. Research by Tavva (2025) demonstrates that such architectures can reduce mean time to resolution (MTTR) for data quality incidents by over 80% [11].
- AI-Native Data Contracts: As Data Mesh architecture matures, organizations will adopt formal "data contracts", API-like agreements between data producers and consumers. AI will automate the generation, validation, and enforcement of these contracts, ensure continuous SLA compliance and automatically alerting stakeholders when data quality or schema changes violate agreed terms [8].

9.2. Long-Term Trends (2027–2030)

- Privacy-Preserving Federated Learning: To govern data across highly restricted borders (e.g., cross-national healthcare research), organizations will increasingly rely on federated learning and differential privacy. AI models will be trained locally on governed data silos, sharing only model weights rather than raw sensitive data, thereby ensuring absolute privacy compliance while still enabling global-scale analytics [14]. This approach will be particularly transformative for regulated industries like healthcare and financial services.
- Quantum-Safe Data Encryption and Governance: As quantum computing advances, current encryption standards (e.g., RSA-2048, AES-128) will become vulnerable to quantum attacks. Future governance frameworks will need to autonomously classify data based on long-term sensitivity and automatically migrate critical assets to quantum safe cryptographic standards (e.g., CRYSTALS-Kyber, CRYSTALS-Dilithium) as mandated by NIST's Post-Quantum Cryptography standardization [10].
- Cognitive Data Mesh Orchestration: The ultimate evolution of the Data Mesh will be fully cognitive orchestration. An overarching AI "brain" will autonomously manage federated governance policies, dynamically routing data compute resources, optimizing storage costs, and resolving policy conflicts across a decentralized global enterprise without human intervention. This represents the pinnacle of the Level 6 Intelligent governance maturity described in our model.
- Explainable AI Governance: As AI models become increasingly embedded in governance decisions (e.g., automated access denials, data quality judgments), there will be a growing demand for explainability. Future governance platforms will be required to provide human-readable justifications for all automated governance decisions, ensuring transparency and accountability in compliance with emerging AI regulations such as the EU AI Act [14].

Fig. 13: Enterprise Data Governance Challenges and Corresponding AI-Driven Solutions

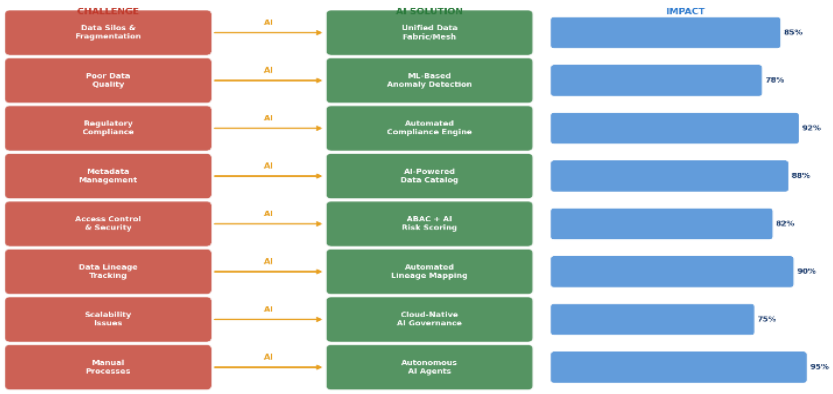

Figure 13. Enterprise Data Governance Challenges and Corresponding AI-Driven Solutions. Each Challenge is Mapped to A Specific AI Solution with Quantified Impact Scores Ranging from 75% to 95%

10. Conclusion

The exponential growth of enterprise data, coupled with an increasingly complex regulatory environment, has rendered traditional, manual data governance frameworks obsolete. This paper has presented a comprehensive AI-Driven Unified Data Governance Framework designed to address the critical challenges of data sprawl, poor data quality, and compliance risk in modern enterprise platforms.

By embedding artificial intelligence and machine learning at the core of the governance lifecycle, enterprises can transition from reactive data management to proactive, autonomous governance. Our seven-layer architectural model provides a scalable blueprint for hybrid and multi-cloud environments, seamlessly integrating with modern paradigms like Data Fabric and Data Mesh. The proposed framework automates the full spectrum of governance activities: from intelligent data discovery and classification, through ML driven quality management and automated lineage tracking, to dynamic policy enforcement and real-time compliance monitoring.

The empirical evidence from four diverse industry case studies: Financial Services, Healthcare, Global Retail, and Smart Manufacturing demonstrate the profound operational and financial benefits of this approach. Organizations implementing AI-driven governance achieve up to 72% reduction in compliance reporting time, near-perfect data quality scores exceeding 95%, and rapid ROI through automated risk mitigation and operational efficiency. The six level AI Governance Maturity Model provides a clear strategic roadmap for organizations at any stage of their governance journey.

As we look toward the future, the integration of Large Language Models, autonomous self-healing agents, privacy-preserving federated learning, and quantum-safe cryptography will further elevate the capabilities of data governance platforms. To remain competitive and compliant in the digital age, enterprises must embrace the AI governance maturity model, systematically advancing toward fully intelligent, autonomous data ecosystems. Data is no longer just an asset to be stored; it is a dynamic entity that requires intelligent, continuous governance to unlock its true value and drive sustainable enterprise growth.

References

[1] A. Anand, "AI Driven Data Governance for the Enterprise Intelligence," SSRN Electronic Journal, Oct. 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4767837.

[2] T. Adenuga, A. T. Ayobami, and U. Mike-Olisa, "Enabling AI-Driven Decision-Making through Scalable and Secure Data Infrastructure for Enterprise Transformation," International Journal of Computer Technology and Applications, 2024.

[3] S. Ahmad, D. Arumugam, S. Bozovic, E. Degefa, S. Duvvuri, S. Gott, et al., "Microsoft Purview: A System for Central Governance of Data," Proceedings of the VLDB Endowment, vol. 16, no. 12, pp. 3624–3635, 2023. doi:10.14778/3611540.3611552.

[4] A. Meesala, "Machine Learning Enabled Governance Framework for Autonomous Enterprise Platforms and Intelligent Data Ecosystems," International Journal of Computer Technology and Applications, 2024.

[5] F. U. Ojika, W. O. Owobu, O. A. Abieba, O. J. Esan, and A. I. Daraojimba, "AI-Driven Models for Data Governance: Improving Accuracy and Compliance through Automation and Machine Learning," Gulf Journal of Computer Sciences, vol. 1, no. 2, pp. 33–54, Apr. 2025.

[6] R. M. N. Gunasekaran, "AI-Driven Data Governance: Ensuring Compliance in Big Data Ecosystems," International Journal of AI, BigData, Computational and Management Studies, 2026.

[7] N. R. Joshi, "Enterprise-Scale AI Architecture for Secure Mobile Platforms with Governance-Driven Automation, Large-Scale Data Warehousing and Machine Learning," International Journal of Technology, Management and Humanities, 2025.

[8] I. Blohm, et al., "Data products, data mesh, and data fabric," Business & Information Systems Engineering, vol. 66, no. 4, pp. 389–407, 2024. doi:10.1007/s12599-024-00876-5.

[9] Y. A. Bena, F. Muchtar, R. Ibrahim, et al., "Enhancing Big Data Governance Practices: Addressing Security, Privacy and Ethical Challenges," Journal of Advanced Research in Computing and Applications, 2026.

[10] A. Rangan and D. A. Yoost, Governance in The Age of Gen AI: A Director's Handbook on Gen AI. 2025.

[11] G. Tavva, "Scalable data quality alerting powered by AI Models: Architecture and tooling for self-healing data pipelines," ResearchGate, 2025.

[12] A. Satyanarayanan, "Optimizing Data Quality in Real-Time: A Self-Healing Pipeline Approach," International Journal of AI, BigData, Computational and Management Studies, 2022.

[13] V. R. Vemula, "AI-enhanced self-healing cloud architectures for data integrity, privacy, and sustainable learning," in Education and Sustainable Learning Environments, IGI Global, 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, 2026. doi:10.1111/exsy.70213.

[15] A. Kumar, "Legal and Regulatory Frameworks Governing Generative AI for Enterprises," in GenAI and LLMs for Beyond 5G Networks, Springer, 2026.

[16] P. Purohit, F. Al Nuaimi, and S. Nakkolakkar, "Data Governance, Privacy, Data Sharing Challenges," in Proceedings of the SPE Global Oil Technology Showcase and Conference, 2024.